

Diseño y análisis de experimentos

Diseño y análisis de experimentos

Luz Mery González García



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Bogotá, D. C., Colombia, 2024

Catalogación en la publicación Universidad Nacional de Colombia

González García, Luz Mery, 1975-

Diseño y análisis de experimentos / Luz Mery González García. – Primera edición. – Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias : Coordinación de publicaciones Facultad de Ciencias, 2024.

1 recurso en línea (483 páginas) : ilustraciones (algunas a color), diagramas. – (Colección Notas de clase)

Incluye referencias bibliográficas e índice analítico

ISBN 978-958-505-767-8 (digital)

1. Diseño experimental – Problemas, ejercicios, etc. 2. Análisis de varianza 3. Modelos matemáticos 4. Toma de decisiones (Estadística) – Ejercicios y problemas 5. Modelos lineales (Estadística) 6. Diseños factoriales 7. Muestreo (Estadística) 8. Modelos multinivel (Estadística) 9. Polinomios ortogonales 10. Diseño experimental – Procesamiento de datos 11. Procesos estocásticos 12. Experimentación científica 13. Investigación – Metodología I. Título II. Serie

CDD-23 519.57 / 2024

© Universidad Nacional de Colombia

Facultad de Ciencias

© Luz Mery González García

Primera edición, agosto de 2024

ISBN 978-958-505-767-8 (digital)

Edición

Daniela Guerrero Acosta

Coordinación de publicaciones - Facultad de Ciencias

coopub_fcbog@unal.edu.co

Corrección de estilo:

Gloria Agudelo Molina

Diseño de la colección

Leonardo Fernández Suárez

**Maqueta **

Camilo Cubides

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales

Contenido

Lista de figuras	VII
Lista de tablas	XIII
Prefacio	XXIII
Capítulo 1	
Principios del diseño de experimentos	1
1. Unidades experimentales y unidades muestrales	5
2. Fuentes de variación	6
2.1. Factores y niveles	7
3. Tres aspectos del diseño experimental	8
Capítulo 2	
Introducción a los modelos lineales	19
1. Método de estimación por mínimos cuadrados	23
2. Estimabilidad	25
3. Mejor estimador linealmente insesgado	31
4. Matriz de proyección (P_x)	35
5. Modelo lineal en dos partes	36

Contenido

6. Distribución de formas lineales y cuadráticas	38
6.1. Momentos de L y Q	39
7. Ejercicios	43

Capítulo 3

Diseño completamente aleatorizado	45
1. Aleatorización	47
2. Diseño a una vía de clasificación	50
2.1. El modelo	52
2.2. Intervalos de confianza para la media muestral de un tratamiento e intervalos de confianza para la diferencia de medias entre dos tratamientos	62
2.3. Análisis de varianza	66
2.4. Tamaño de muestra	68
3. Submuestreo en un DCA	70
3.1. El modelo	72
3.2. Análisis de varianza	76
3.3. Tamaño de muestra en submuestreo	79
4. Modelo de efectos aleatorios	79
4.1. El modelo	80
4.2. Tamaño de muestra [21, p. 628]	83
5. Verificación de supuestos en un DCA	88
5.1. Gráfico Q-Q plot para los datos $x_1, x_2, \dots, x_n$	91
5.2. Test para normalidad y homogeneidad de varianza	93
5.3. Algunas transformaciones	95
6. Contrastes	96
6.1. Distribución y esperanza de las suma de cuadrados de $\lambda^\top \hat{\beta}$	98
6.2. Contrastes ortogonales	102
6.3. Pruebas de comparación múltiple	104

7. Polinomios ortogonales	123
8. Análisis de covarianza	141
8.1. El modelo	143
8.2. Análisis de varianza	148
9. Prueba Kruskal-Wallis	152
10. Ejercicios	155

Capítulo 4

Diseños con una o más fuentes de bloqueo 159

1. Diseño en bloques completos aleatorizados	161
1.1. El modelo	163
1.2. Análisis de varianza	168
1.3. Datos faltantes	173
1.4. Eficiencia relativa de un DBCA frente a DCA	177
2. Prueba de Friedman para k muestras	178
3. Ideas sobre bloques incompletos balanceados	182
3.1. Características	182
4. Diseño en cuadro latino	187
4.1. El modelo	189
4.2. Análisis de varianza	191
5. Diseño en cuadro grecolatino	196
5.1. El modelo	198
5.2. Análisis de varianza	201
6. Ejercicios	205

Capítulo 5

Diseños con arreglos factoriales 209

1. Principios fundamentales para efectos de diseños con arreglos factoriales	215
--	-----

Contenido

2. El modelo para un diseño con arreglos factoriales $A \times B$	216
2.1. Análisis de varianza	222
2.2. Verificación de supuestos	229
3. Diseño con arreglos factoriales $A \times B$ en bloques completos . . .	230
3.1. Análisis de varianza	231
4. Diseño con arreglos factoriales $A \times B \times C$	237
4.1. Análisis de varianza	246
5. Diseño con arreglos factoriales 2^k	256
5.1. Diseño 2^2	260
5.2. Diseños con arreglos factoriales 2^k con una sola réplica o no replicado	263
6. Diseños con arreglos factoriales 3^k	271
6.1. Contrastes ortogonales en diseños con arreglos factoriales 3^k	283
7. Ejercicios	291

Capítulo 6

Confusión y diseños con arreglos factoriales fraccionados 293

1. Confusión en diseños con arreglos factoriales 2^k	295
1.1. Confusión de un diseño con arreglos factoriales 2^k en dos bloques	296
1.2. Confusión de un diseño con arreglos factoriales 2^k en cuatro bloques	305
1.3. Réplicas de diseños 2^k confundidos en 2^p bloques	309
2. Diseño con arreglos factoriales fraccionado 2^{k-1}	312
3. Confusión en diseños con arreglos factoriales 3^k	323
4. Diseños con arreglos factoriales fraccionados 3^{k-1}	331

5. Pseudofactores	339
6. Diseño central compuesto	341
7. Ejercicios	349

Capítulo 7

Diseño anidado en dos etapas 353

1. El modelo	356
2. Análisis de varianza	361
2.1. Casos a considerar para la interpretación del análisis de varianza	363

Apéndice A

Inversa generalizada 369

1. Inversa de Moore-Penrose	371
2. Inversa generalizada G	375

Apéndice B

Deducción del modelo DCA 379

1. V's A's involucradas en el DCA	381
1.1. Respuestas y observaciones conceptuales	382
1.2. Considerando otros errores en el modelo	385
2. Análisis de varianza - DCA	387
2.1. Distr. y esper. de las SC	388
2.2. Cuadrados medios (CM)	392
3. Análisis de varianza con número desigual de replicaciones	397
4. Tamaño de muestra	399
4.1. Programa básico para tablas sobre tamaño de muestra ..	401
5. Análisis de varianza - submuestreo	402
5.1. Distribución S.C.	403
5.2. Esperanza de los cuadrados medios	408
5.3. Cociente de cuadrados medios	409
5.4. Ejercicios	410

Contenido

6.	Análisis de varianza - modelo de efectos aleatorios	411
6.1.	Distr. y esperanza de la S.C.	412
6.2.	Cuadrados medios	416
7.	Construcción envelope	418
8.	Análisis de varianza - polinomios ortogonales	420
8.1.	Sumas de cuadrados de $\gamma_l, l = 1, 2, \dots, a - 1$	420
8.2.	Suma de cuadrados del error (SCE)	421
9.	Análisis de varianza - modelo con una covariable	422
9.1.	Distribución y esperanza de las S.C.	424
9.2.	Ejercicios	431

Apéndice C

Análisis de varianza - diseño anidado **433**

1.	Análisis de varianza - anidado - caso 1	435
1.1.	Hipótesis lineal general	435
1.2.	Distribución y esperanza de las S.C.	436
1.3.	Independencia entre las sumas de cuadrados consideradas	443

Bibliografía **445**

Índice **453**

Lista de figuras

1.1. Pasos en el diseño de investigaciones	11
1.2. Situación 1	11
1.3. Situación 2	12
1.4. Situación 3	14
1.5. Situación 4	15
3.1. Gráfico de puntos para el ejemplo sobre la altura de las plan- tas de colza	53
3.2. Boxplot para el ejemplo sobre rendimiento en frijol	71
3.3. Gráfico de puntos para el ejemplo sobre rendimiento en fri- jol por parcela	72
3.4. Envelope para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 3.3	91
3.5. Graficos para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 3.3	92
3.6. Agrupaciones y rango (mínimo y máximo) obtenidas en el ejemplo 3.23	114
3.7. Gráfico de puntos para el ejemplo sobre los almidones y la velocidad de flujo (g/seg)	115

Lista de figuras

3.8. Agrupaciones y rango (mínimo y máximo) obtenidas en el ejemplo 3.24	117
3.9. Intervalos de confianza. Método Dunnett - ejemplo 3.27	122
3.10. Gráfico de puntos para el ejemplo sobre los concentrados con probiótico y peso promedio de los peces por acuario a los 45 días	130
3.11. Gráfico de dispersión, nivel de probiótico en unidades z_i versus peso promedio de los acuarios a los 45 días, y línea ajustada bajo el modelo de regresión lineal $\hat{y} = 26.380 + 3.834z$ para los datos del ejemplo 3.29	135
3.12. Gráfico de puntos para el ejemplo sobre los concentrados con probiótico y peso de los peces a los 45 días	137
3.13. Gráfico de dispersión para el ejemplo sobre los niveles de probióticos teniendo en cuenta el peso promedio inicial y el peso promedio a los 45 días por acuario	143
4.1. Gráfico de dispersión para el ejemplo sobre el rendimiento de 5 variedades de cebada	164
4.2. Envelope para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 4.1	180
4.3. Diseño tipo I	184
4.4. Diseño tipo I	184
4.5. Diseño tipo II	184
4.6. Diseño tipo III	185
4.7. Diseño tipo III	185
4.8. Ejemplo de un diseño en cuadro latino para $a = 4$	188
4.9. Ejemplo de un diseño en cuadro latino replicado	194
4.10. Ejemplo de un diseño en rectángulo latino	195
4.11. Ejemplo de un diseño en cuadro grecolatino para $a = 4$	197
5.1. Promedios de contenido de fenoles por temperatura y precisión - ejemplo 5.3	213

Lista de figuras

5.2. Promedios de contenido de fenoles por temperatura y presión . Datos modificados para efecto ilustrativo	223
5.3. Promedios de altura por dosis de silicio y niveles de riego - ejemplo 5.6	234
5.4. Envelope para evaluar el cumplimiento de los supuestos del modelo ajustado a los datos del ejemplo 5.6	235
5.5. Promedios de porcentaje de recuperación de compuestos volátiles. Efectos principales	239
5.6. Promedios de porcentaje de recuperación de compuestos volátiles. Interacciones dobles	241
5.7. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple	243
5.8. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple	244
5.9. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple	245
5.10. Envelope para el modelo ajustado a los datos del ejemplo 5.7	252
5.11. Valores ajustados versus residuales estudentizados para el modelo ajustado a los datos del ejemplo 5.7	253
5.12. Envelope para el modelo ajustado a los datos del ejemplo 5.7 retirando la observación 19	254
5.13. Promedios por niveles de factor - ejemplo 5.9	258
5.14. Promedios por niveles de factor cruzados dos a dos - ejemplo 5.9	258
5.15. Gráfico de probabilidad normal de los efectos estimados para los datos del ejemplo 5.11	266
5.16. Half-normal plot de los efectos estimados para los datos del ejemplo 5.11	269
5.17. Boxplot según velocidad de corte y según ángulo de filo - ejemplo 5.12	273
5.18. Promedios por niveles de factor - ejemplo 5.12	273
5.19. Promedios por niveles de factor cruzados - ejemplo 5.12	274

Lista de figuras

5.20. Gráfico de dispersión (puntos) y valores ajustados (triángulos) de acuerdo al modelo presentado en la tabla 5.31 - ejemplo 5.12	282
5.21. Half-normal plot de los efectos estimados para los datos del ejemplo 5.16	289
6.1. Esquema de asignación de un diseño con 2^2 tratamientos a dos bloques	296
6.2. Esquema de asignación de un diseño con 2^3 tratamientos a dos bloques	298
6.3. Gráfico de Pareto de los efectos estimados para los datos del ejemplo 6.1. Margen de error (ME) y margen de error simultáneo (SME) obtenidos por el método de Lenth	300
6.4. Valores medios estimados por peso de carga y para la interacción peso del proyectil con el propulsor	303
6.5. Esquema de asignación de un diseño con 2^4 tratamientos a cuatro bloques	306
6.6. Asignación de 2^4 tratamientos a cuatro bloques considerando un diseño completamente confundido	310
6.7. Asignación de 2^4 tratamientos a cuatro bloques considerando un diseño parcialmente confundido	310
6.8. Tratamientos efectivamente aplicados en el ejemplo 6.3, según tabla 6.10	314
6.9. Gráfica half-normal (arriba) y Pareto (abajo) de los parámetros estimados - ejemplo 6.4	318
6.10. Valores promedio estimados de acuerdo a la presión del gas-plasma, la distancia entre boquilla y pieza, y el método de cortado - ejemplo 6.4	321
6.11. Esquema de asignación de un diseño con 3^2 tratamientos a tres bloques usando AB^2	326
6.12. Valores ajustados promedio - ejemplo 6.6	329
6.13. Valores ajustados promedio - ejemplo 6.8	337

Lista de figuras

6.14. Perspectivas de la superficie de respuesta ajustada a los datos del ejemplo 6.9	347
---	-----

Lista de tablas

1.1. Jerarquía de los diseños para controlar el error	17
1.2. Jerarquía de los diseños de tratamientos	18
1.3. Jerarquía de los diseños muestrales	18
2.1. Datos ilustrativos. Una vía de clasificación	26
2.2. Análisis de varianza simple	36
2.3. Tabla de análisis de varianza para un modelo lineal ordenado en dos partes	37
3.1. Tabla de respuestas potenciales o conceptuales para 6 individuos y dos tratamientos	48
3.2. Tratamientos considerados para el ejemplo de los fertilizantes	51
3.3. Altura de la planta de colza (<i>Brassica sp.</i>) bajo algunos niveles de fertilizante	52
3.4. Análisis de varianza para un diseño completamente aleatorizado	66
3.5. Análisis de Varianza para los datos del ejemplo 3.3	67
3.6. Ejemplo del nitrógeno y el frijol	71
3.7. Tabla de análisis de varianza para un diseño completamente aleatorizado en presencia de submuestreo	77

Lista de tablas

3.8. Análisis de varianza para el ejemplo de los porcentajes de nitrógeno y el frijol	78
3.9. Algunos escenarios de tamaños de muestra usando submuestreo en el ejemplo de las plantas de colza	80
3.10. Ejemplo de las hojas del árbol y el peso seco	80
3.11. Tabla de análisis de varianza para el ejemplo 3.15	82
3.12. Tabla obtención tamaño de muestra del ejemplo para $r = 11$	85
3.13. Tabla obtención tamaño de muestra del ejemplo para $r = 20$	86
3.14. Tabla obtención tamaño de muestra del ejemplo para $r = 60$	87
3.15. Tabla obtención tamaño de muestra del ejemplo para $r = 5$..	87
3.16. Algunas transformaciones	96
3.17. Análisis de varianza para un diseño completamente aleatorizado y contrastes ortogonales	102
3.18. Análisis de varianza para los datos sobre fertilizantes y altura de las plantas de colza en contrastes ortogonales	103
3.19. Análisis de varianza para un diseño completamente aleatorizado usando polinomios ortogonales	129
3.20. Pesos (g) promedio de las mojaras por acuario a los 45 días ...	129
3.21. Análisis de varianza para los datos del ejemplo 3.29, sin considerar el submuestreo	131
3.22. Promedio de los pesos (g) de mojarra roja a los 45 días (fase de levante), en el experimento de los concentrados con probióticos	131
3.23. Análisis de varianza para los datos del ejemplo 3.29 usando polinomios ortogonales	134
3.24. Análisis de varianza para los datos del ejemplo 3.29 considerando el submuestreo	137
3.25. Análisis de varianza para los datos del ejemplo 3.29 usando polinomios ortogonales y submuestreo	139

Lista de tablas

3.26. Peso promedio inicial y peso promedio a los 45 días de las mojaras, por acuario, según el nivel de probiótico	142
3.27. Tabla de análisis de covarianza para un diseño completamente aleatorizado	149
3.28. Tabla de análisis de covarianza para el ejemplo 3.30	150
3.29. Ejemplo del número de ciegos pilóricos en peces de un caño de aguas negras amazónicas	153
3.30. Ejemplo del número de ciegos pilóricos en peces de un caño de aguas negras amazónicas. Rangos	154
3.31. Tamaños de muestra para el ejemplo 3.9	155
3.32. Radón liberado versus diámetro de los orificios de las regaderas	156
4.1. Esquema general - diseño en bloques	162
4.2. Ejemplo de las variedades de cebada y el rendimiento en gramos por parcela de $7.5m^2$	163
4.3. Análisis de varianza para un diseño en bloques completamente aleatorizados	169
4.4. Análisis de varianza para un diseño en bloques completamente aleatorizado (balanceado)	170
4.5. Análisis de varianza para los datos del ejemplo 4.1	171
4.6. Análisis de varianza para los datos del ejemplo 4.1, considerando la pérdida del dato asociado con la variedad 2 y la réplica III	176
4.7. Ejemplo 4.1 sobre las variedades de cebada y el rendimiento en gramos. Rangos	181
4.8. Ejemplo de las dietas y las llamas en cuadro latino	189
4.9. Análisis de varianza para un diseño en cuadro latino (balanceado)	192
4.10. Análisis de varianza para los datos del ejemplo 4.8	192
4.11. Análisis de varianza para un diseño en cuadro latino replicado	195
4.12. Análisis de varianza para un diseño en rectángulo latino	196

Lista de tablas

4.13. Ejemplo del porcentaje de carbón orgánico extractable del suelo. Cuadro grecolatino	198
4.14. Análisis de varianza para un diseño en cuadro grecolatino (balanceado)	202
4.15. Análisis de varianza para los datos del ejemplo 4.12	204
4.16. Datos sobre 6 dietas y la ganancia de peso en conejos	205
4.17. Datos ilustrativos considerando una variable respuesta y una covariable	206
4.18. Análisis de varianza mostrado por el investigador del ejercicio 6.7	207
5.1. Ejemplo de la actividad citotóxica, contenido de fenoles expresados en mg AG (ácido gálico)/100 g en base seca	213
5.2. Tabla de análisis de varianza para un diseño con arreglos factoriales $A \times B$. Diseño balanceado	223
5.3. Tabla de análisis de varianza para el ejemplo 5.3	227
5.4. Análisis de varianza para un diseño con arreglos factoriales $A \times B$ en bloques completos. Diseño balanceado	233
5.5. Datos del ejemplo sobre el efecto de silicio en la producción de pepino	234
5.6. Análisis de varianza para los datos del ejemplo 5.6 sobre la producción de pepino	235
5.7. Datos ejemplo sobre porcentaje de recuperación de compuestos volátiles	239
5.8. Análisis de varianza para un diseño con arreglos factoriales $A \times B \times C$. Diseño balanceado	249
5.9. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles	250
5.10. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles sin interacción triple	251

Lista de tablas

5.11. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles considerando la interacción tiempo y salinidad	251
5.12. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles considerando solo efectos principales	252
5.13. Modelos ajustados para los datos del ejemplo 5.7 con su respectivo <i>AIC</i> , sin la observación 19	253
5.14. Porcentajes esperados de recuperación de compuestos volátiles, modelo (5.4)	254
5.15. Notación de los tratamientos en un diseño con arreglos factoriales 2^2	257
5.16. Datos ejemplo 5.9 sobre el rendimiento de biodiésel, datos ilustrativos	257
5.17. Análisis de varianza para un diseño 2^2	261
5.18. Porcentaje de degradación de clorpirifós	262
5.19. Análisis de varianza sobre los datos de la degradación de clorpirifós	262
5.20. Datos ejemplo 5.11	265
5.21. Análisis de varianza sobre los datos del ejemplo 5.11, retirando la interacción cuádruple del modelo	267
5.22. Análisis de varianza sobre los datos del ejemplo 5.11 con la interacción triple relación:catalizador:temperatura	267
5.23. Notación de los tratamientos en un diseño con arreglos factoriales 3^2	271
5.24. Energía específica de corte (J/mm^2) bajo 30 % de base húmeda - ejemplo 5.12	272
5.25. Análisis de varianza sobre los datos de energía específica de corte - ejemplo 5.12	277
5.26. Análisis de varianza sobre los datos de energía específica de corte - ejemplo 5.12 sin considerar la interacción	278

Lista de tablas

5.27. Análisis de varianza sobre los datos de energía específica de corte - ejemplo 5.12, considerando en el modelo solo el ángulo de filo	278
5.28. Entradas matriz diseño con arreglos factoriales 3^2	279
5.29. Análisis de varianza para un diseño con arreglos factoriales 3^2 con niveles cuantitativos e igualmente espaciados	280
5.30. Análisis de varianza para los datos del ejemplo 5.12 considerando polinomios ortogonales	281
5.31. Análisis de varianza para los datos del ejemplo 5.12 considerando polinomios ortogonales, solo con el efecto del ángulo de filo	281
5.32. Ejemplo de entradas matriz diseño en un diseño con arreglos factoriales 3^2 con niveles cualitativos	283
5.33. Análisis de varianza para los datos del ejemplo 5.12 considerando contrastes ortogonales	287
5.34. Datos ejemplo ilustrativo 5.16	288
5.35. Análisis de varianza usando contrastes para los datos del ejemplo 5.16 sin considerar los contrastes C_7 y C_8	290
5.36. Análisis de varianza para los datos del ejemplo 5.16 sin considerar los contrastes C_6 , C_7 y C_8	290
6.1. Asignación de 2^2 tratamientos a dos bloques	296
6.2. Asignación de 2^3 tratamientos a dos bloques	297
6.3. Datos ejemplo 6.1	299
6.4. Análisis de varianza datos ejemplo 6.1 considerando las interacciones dobles	301
6.5. Análisis de varianza datos ejemplo 6.1 considerando solo la interacción proyectil:propulsor	301
6.6. Asignación de los 2^4 tratamientos a cuatro bloques confundiendo AB , CD y $ABCD$	306
6.7. Análisis de varianza con causa de variación y grados de libertad de un diseño 2^4 en cuatro bloques considerando un diseño completamente confundido	311

Lista de tablas

6.8. Asignación de 2^4 tratamientos a cuatro bloques confundiendo AC , BD y $ABCD$	311
6.9. Análisis de varianza con causa de variación y grados de libertad de un diseño 2^4 en cuatro bloques considerando un diseño parcialmente confundido	312
6.10. Asignación de los 2^{4-1} tratamientos usando la interacción $ABCD$	313
6.11. Datos sobre el estudio de la exactitud del corte con plasma en el largo de la pieza - ejemplo 6.4	317
6.12. Análisis de varianza sin considerar la interacción velocidad* método en el ejemplo 6.4	319
6.13. Análisis de varianza sin considerar algunas interacciones dobles para el ejemplo 6.4	320
6.14. Análisis de varianza con las interacciones dobles presión y distancia, y presión y método del ejemplo 6.4	320
6.15. Asignación de 3^2 tratamientos a tres bloques usando un grado de libertad del factor A	324
6.16. Asignación de 3^2 tratamientos a tres bloques usando un grado de libertad de la interacción $A \times B$	325
6.17. Esquema de asignación del diseño 3^3 y datos del ejemplo 6.6 en tres bloques	328
6.18. Análisis de varianza para los datos del ejemplo 6.6	328
6.19. Análisis de varianza para los datos del ejemplo 6.6 considerando polinomios ortogonales	329
6.20. Asignación de los 3^{3-1} tratamientos usando la componente ABC^2	332
6.21. Tratamientos seleccionados	332
6.22. Tabla de contrastes ortogonales para el ejemplo 6.7	333
6.23. Análisis de varianza para los datos del ejemplo 6.8	337
6.24. Análisis de varianza para los datos del ejemplo 6.8 considerando solo peso y pendiente longitudinal del tramo	337
6.25. Asignación de los 6^{3-1} tratamientos usando las componentes $A_1B_1C_1$ y $A_2B_2C_2^2$	341

Lista de tablas

6.26. Diseño con arreglos factoriales 2^2 con tres puntos centrales . .	342
6.27. Diseño con arreglos factoriales 2^2 con tres puntos centrales y puntos axiales	343
6.28. Diseño con arreglos factoriales 2^2 con tres puntos centrales y puntos axiales, $\alpha = 1.414$	344
6.29. Datos ejemplo sobre absorción del CO_2 de un biogás	346
6.30. Estimación de parámetros para el modelo de superficie de respuesta ajustado a los datos del ejemplo 6.9	346
6.31. Estimación de parámetros para el modelo de superficie de respuesta ajustado a los datos del ejemplo 6.9	347
7.1. Datos de un diseño anidado en dos etapas - ejemplo	355
7.2. Datos ilustrativos sobre nivel de ozono (O_3 en ppb) - ejem- plo 7.2	356
7.3. Análisis de varianza para un diseño anidado en 2 etapas, ba- lanceado	362
7.4. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i y $\beta_{j(i)}$ asociados a efectos fijos	363
7.5. Análisis de varianza para los datos el ejemplo 7.2 asumiendo que los factores tipo de máquina y empresa son de efectos fijos	364
7.6. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i asociado a efectos fijos y $\beta_{j(i)}$ asociado a efectos aleatorios	365
7.7. Análisis de varianza para los datos el ejemplo 7.2 asumiendo que el factor máquina es un efecto aleatorio y el factor empresa es un efecto fijo	366
7.8. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i y $\beta_{j(i)}$ asociados a efectos aleatorios	367
B.1. Análisis de varianza corregida por la media	388
B.2. Análisis de varianza corregida por la media	392

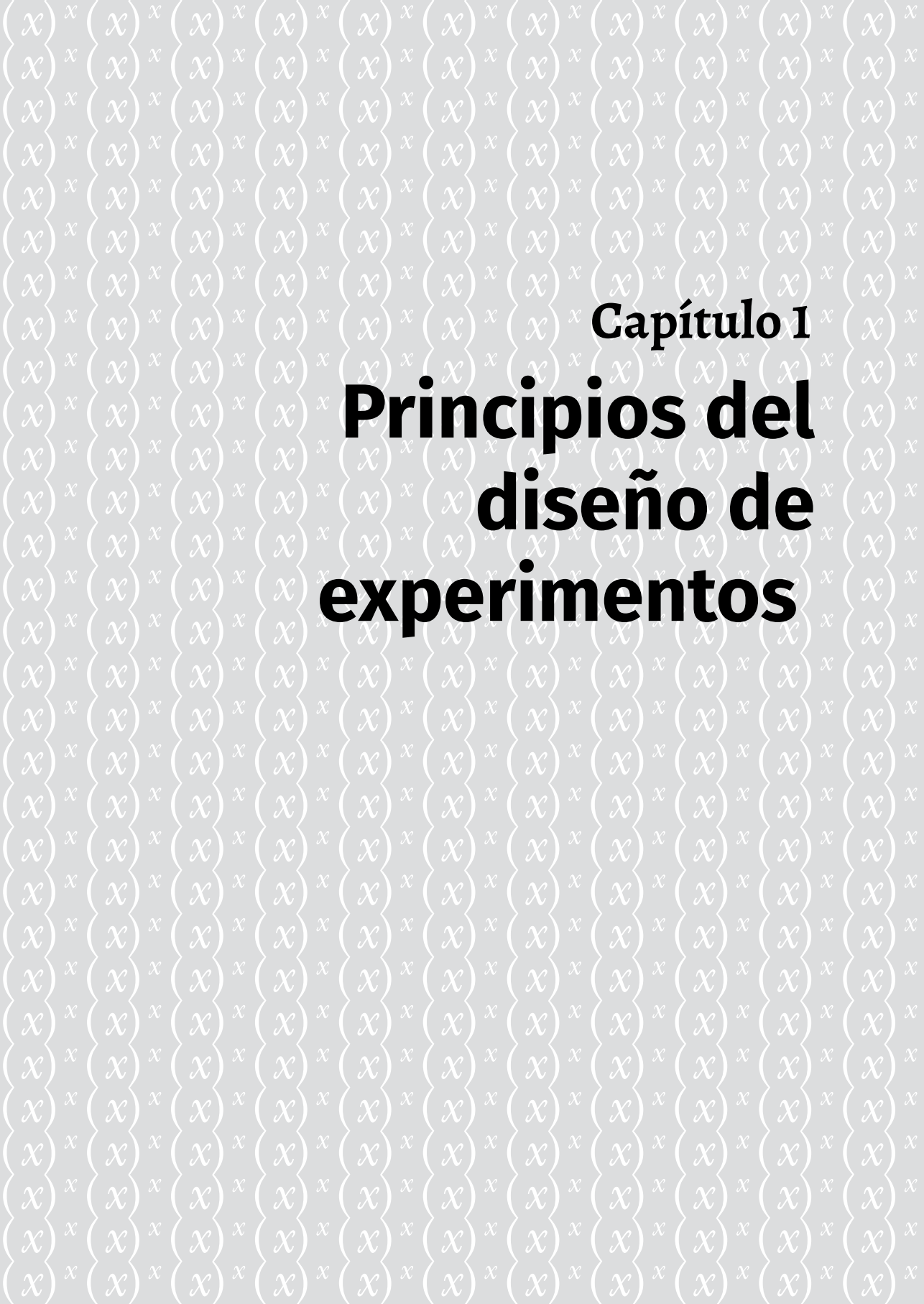
Lista de tablas

B.3. Análisis de varianza incluyendo esperanza de los cuadrados medios	394
B.4. Datos ilustrativos	395
B.5. Análisis de varianza para el ejemplo ilustrativo B.2	396
B.6. Análisis de varianza para un diseño a una vía de clasificación no balanceado	397
B.7. Datos ilustrativos	398
B.8. Análisis de varianza para los datos ilustrativos presentados en la tabla B.7	399
B.9. Análisis de varianza incluyendo esperanza de los cuadrados medios	409
B.10. Análisis de varianza incluyendo esperanza de los cuadrados medios	417
B.11. Organización de los residuales estandarizados de las m simulaciones	419
B.12. Análisis de varianza incluyendo esperanza de los cuadrados medios	430

Prefacio

Las notas de clase de *Diseño y análisis de experimentos* presentan los temas tratados en la asignatura Diseño de experimentos de la carrera de estadística de la Universidad Nacional de Colombia, sede Bogotá, cuyo enfoque parte desde los modelos lineales. Las notas contienen siete capítulos. El primer capítulo hace una introducción al diseño de experimentos. En el segundo se presentan los conocimientos sobre modelos lineales necesarios para el análisis de varianza de los diseños experimentales a tratar en los siguientes capítulos. En el tercer capítulo se desarrollan los diseños completamente aleatorizados y sus variantes, acompañando los temas de ejemplos y códigos en R. Además, se realizan demostraciones algebraicas de los resultados teóricos presentados, las cuales se presentan en los anexos para facilitar la lectura de los estudiantes. En el capítulo cuatro se trata el tema de control local a través de bloqueo. En los capítulos cinco y seis se presentan los diseños factoriales. Finalmente, en el capítulo siete se hace una muy breve introducción a los diseños anidados.

En el transcurso de las notas de clase se usan ejemplos tanto de textos tradicionales como de tesis desarrolladas en la Universidad Nacional de Colombia, haciendo énfasis en estos últimos. Considero que este es uno de los puntos favorables de estas notas de clase.



Capítulo 1

Principios del diseño de experimentos

Según la Real Academia Española (RAE), un *experimento* es la “acción y efecto de experimentar” y *experimentar* es “probar y examinar prácticamente la virtud y propiedades de algo”. Así mismo, un *diseño experimental* es “un arreglo o procedimiento usado en un estudio experimental” en que el término *estudio experimental* indica que el investigador puede deliberadamente influenciar eventos e investigar los efectos de esa influencia [25].

En este capítulo se introducen conceptos generales de los diseños experimentales, que son importantes al momento de planear un experimento desde el punto de vista de modelos lineales, entre ellos: los objetivos al planear el diseño, la toma de datos en cuanto a si habrá o no submuestreo, si hay variables que se usarán como control además de los tratamientos, y la clasificación de los diseños bajo algunos criterios.

En cuanto a los objetivos del problema al planear el experimento, [61, p. 3] clasifican los problemas experimentales en cinco grandes categorías:

1. **Comparación entre tratamientos.** El principal objetivo es comparar varios tratamientos y seleccionar el mejor.

Ejemplo 1.1. [15] presentan un estudio en perros sobre niveles de histamina en sangre. Consideraron dos medicamento (morfina y trimethaphan) y dos suplementos de histamina (histamina empobrecida e histamina intacta). El objetivo era seleccionar el mejor medicamento y el mejor suplemento de histamina.

Ejemplo 1.2. Considere un experimento realizado en la Universidad Nacional de Colombia para determinar los efecto de la viscosidad, la salinidad y el tiempo sobre el porcentaje de recuperación de compuestos volátiles. Se tuvieron en cuenta tres niveles de viscosidad, tres niveles de salinidad y tres niveles de tiempo [6].

2. **Revisión para selección de variables.** Este tipo de estudios busca detectar entre una gran cantidad de variables, cuáles son las más importantes.

Ejemplo 1.3. [3] realizaron un estudio en la Universidad Nacional de Colombia sobre pruebas de extrusión. Consideraron dos tipos de tornillos (tornillo con diámetro variable y un tornillo con tres zonas), dos tipos de resinas (poliestireno 421 y polietileno 641), tres velocidades de giro del tornillo y 7 niveles de temperatura. Se midió la velocidad másica, la presión y el torque. Aunque el interés en el

trabajo era evaluar un modelo matemático, del experimento se puede determinar cuáles de las condiciones experimentales consideradas pueden ser más importantes frente a las variables medidas.

3. **Exploración de superficies de respuesta.** En este tipo de estudios, el objetivo usualmente es establecer efectos lineales, cuadráticos e interacciones entre las variables, o polinomios de mayor orden.

Ejemplo 1.4. [42] presentan un estudio realizado en la Universidad Nacional de Colombia sobre remoción por absorción reactiva de CO_2 en un sistema alcalino carbonato-biocarbonato. Consideraron dos concentraciones iniciales de la solución alcalina, dos presiones iniciales en el reactor y dos temperaturas del sistema gas-líquido en la absorción del CO_2 . El objetivo era encontrar la relación funcional entre la variable respuesta (cambio en porcentaje de CO_2 en el biogas simulado en la solución alcalina) y las variables explicativas (concentración inicial, presión inicial y temperatura), para lo cual usaron superficies de respuesta.

4. **Optimización de un sistema.** En este tipo de estudios se busca una estrategia para mover el experimento a una región que contenga el óptimo.
5. **Robustez de un sistema.** Busca mejorar un sistema en cuanto a variaciones por ruido dado un conjunto de factores controlados. En este caso la variable respuesta en el análisis estadístico es usualmente la varianza debida al ruido por lo que las variables explicativas se dividen en variables de ruido¹ y en variables controlables.

Ejemplo 1.5. [44] realizó un estudio acerca de la elaboración de néctar de gulupa y curuba en la Universidad Nacional de Colombia. Midió, entre otros, el pH y la viscosidad como variables respuesta, y consideró como variables explicativas: la ausencia/presencia de mesocarpio, tres porcentajes de pulpa, tres cantidades de sólidos disueltos (en grados Brix), tres temperaturas de pasteurización, tres endulcolorantes y tres estabilizantes, de los cuales la temperatura de pasteurización se consideró como variable de ruido y las otras, variables explicativas, como variables controlables.

¹Corresponden a variables que son difíciles de controlar durante un uso o proceso normal.

Según [40, p. 6], algunos aspectos para tener en cuenta a la hora de planear un experimento son:

1. **Simplicidad** al hacer la selección de los tratamientos.
2. **Grado de precisión.** Escoger el diseño apropiado y el número de réplicas apropiadas para la precisión deseada.
3. **Ausencia de error sistemático.** Las unidades experimentales asignadas a un tratamiento no deben diferir de las unidades experimentales asignadas a otro tratamiento.
4. El rango de validez de las conclusiones debe ser lo más amplio posible.
5. Garantizar que el cálculo del grado de incertidumbre sea solo debido a la aleatoriedad.

Antes de continuar, se va a considerar la diferencia entre modelos de efectos fijos, modelos de efectos aleatorios y modelos de efectos mixtos.

DEFINICIÓN 1.1. Según [40, p. 6]:

- *Modelo de efectos fijos: las conclusiones se formulan sobre un número preestablecido de tratamientos y la inferencia es sobre los efectos medios de los tratamientos.*
- *Modelo de efectos aleatorios: los tratamientos son una muestra aleatoria de una población de tratamientos y la inferencia se hace sobre las varianzas de los tratamientos.*
- *Modelo de efectos mixtos: Es el caso en el cual el modelo contiene tanto efectos fijos como efectos aleatorios.*

1. Unidades experimentales y unidades muestrales

DEFINICIÓN 1.2. La unidad experimental (UE) es la parte del material experimental al cual un tratamiento es asignado y aplicado.

Ejemplo 1.6. Considere las siguientes situaciones:

- A cada paciente seleccionado para el experimento, se le aplica un cierto medicamento. En este caso el paciente es la unidad experimental.
- A cada paciente seleccionado para el experimento, se le aplica un tratamiento en un brazo y en el otro se le aplica otro tratamiento (o el mismo tratamiento). En esta situación cada brazo es una unidad experimental diferente.

DEFINICIÓN 1.3. La unidad de observación o de muestreo (UO) es la unidad sobre la cual las observaciones o mediciones son realizadas.

Ejemplo 1.7. En la aplicación de un método de enseñanza a los alumnos de un curso y posterior aplicación de una prueba de habilidades en el tema del curso, el curso es la unidad experimental y cada estudiante del curso es una unidad de observación.

Nota 1.1. La unidad observacional puede coincidir con la unidad experimental.

Ejemplo 1.8. La unidad experimental es un ratoncito al cual se le aplica un medicamento para disminuir la tensión arterial y, posteriormente, se le mide al ratoncito la tensión arterial. Así, $UE=UO$.

Nota 1.2. Asociado a las unidades experimentales está el error experimental (EE) debido a las diferencias propias de las unidades experimentales. Contribuye al error experimental el error de tratamiento debido a dificultades en la administración del tratamiento. Asociado a la unidad observacional está el error observacional (EO) debido a errores de medición y errores de muestreo.

Ejemplo 1.9. (Sobre error de tratamiento). En un estudio sobre el efecto del dióxido de cloro acuoso (ClO_2) en el control de patógenos en arándanos, se debe aplicar 15ppm de ClO_2 (según el protocolo de la investigación), pero en realidad se aplican entre 14 ppm y 16 ppm por dificultades en la exactitud de la medición del ClO_2 al ser aplicado.

2. Fuentes de variación

La respuesta puede variar de una observación a otra debido a diferentes **causas de variación**, que corresponden a diferencias intrínsecas entre las mismas

unidades experimentales, las diferencias intrínsecas entre las unidades observacionales, al efecto del diseño estadístico aplicado y al efecto del tratamiento aplicado. Así, [31, p. 37]² proponen el siguiente modelo general para la descomposición de la variabilidad en las observaciones:

$$\begin{aligned} \text{una observación} &= \text{efecto del diseño} \\ &+ \text{efecto del tratamiento} \\ &+ \text{error experimental} \\ &+ \text{error observacional} \end{aligned}$$

2.1. Factores y niveles

DEFINICIÓN 1.4. Según [25, p. 150], el término **factor** es muy usado en estadística, pero comunmente se usa para referirse a una variable categórica con un pequeño número de niveles (categorías), que está bajo investigación en un experimento como una posible causa de variación.

Ejemplo 1.10. Retomando el ejemplo 1.1 sobre niveles de histamina en la sangre en perros se tienen en cuenta dos factores: medicamento y suplemento de histamina. Los niveles del factor medicamento son morfina y trimethaphan. Los niveles del factor suplemento de histamina son histamina empobrecida e histamina intacta.

DEFINICIÓN 1.5. Según [40, p. 9], cada combinación específica de niveles de factores se denomina **tratamiento**.

Ejemplo 1.11. En el ejemplo 1.1, el número de tratamientos es cuatro (2×2):

- Morfina e histamina empobrecida.
- Morfina e histamina intacta.
- Trimethaphan e histamina empobrecida.
- Trimethaphan e histamina intacta.

²En versiones posteriores del libro, [33, p. 40] agregan un término al modelo que es el efecto del diseño $\times$ el efecto de la interacción de los tratamientos, señalando que en ciertos diseños experimentales puede ser un componente de interés.

Nota 1.3. *El concepto de tratamiento implica que cualquier unidad experimental está en capacidad de recibir cualquier tratamiento y, al contrario, cualquier tratamiento puede ser aplicado a cualquier unidad experimental. Así, características como sexo no son tratamientos, ya que el sexo no se asigna aleatoriamente a las unidades experimentales, por ejemplo, perros o ratones, ni las unidades experimentales se asignan aleatoriamente a un sexo, sino que el sexo es una característica intrínseca de la unidad. Por otro lado, una dieta se puede asignar aleatoriamente a los perros o ratones, o los perros o ratones se pueden asignar aleatoriamente a una dieta, lo que implica que en este caso se presente un tratamiento.*

Los factores se pueden dividir en dos clases:

- *Factor de clasificación:* es una propiedad inherente a la unidad experimental, por ejemplo, la variable género. Note que un factor de clasificación no cumple con el concepto de tratamiento.
- *Factor de tratamiento:* es asignado por el investigador a la unidad experimental, por ejemplo, la variable medicamento.

Algunas particularidades que deben tener los tratamientos son [40, p. 10]:

- Presentar la finalidad (por qué voy a considerar esos tratamientos y no otros).
- La respuesta versus el medio circulante o terreno (influencia en la población objetivo).
- Los tratamientos propuestos deben ser “bien” aplicados (condiciones de aplicación).
- Considerar un tratamiento testigo o control.

3. Tres aspectos del diseño experimental

[33, p. 45] describen los tres principios del diseño experimental de la siguiente manera:

Diseño y análisis de experimentos

1. **Replicaciones** de algunos o todos los tratamientos para estimar la magnitud del error experimental.
2. **Aleatorización** de los tratamientos o grupos de tratamientos a las unidades experimentales para producir estimaciones insesgadas y de varianza mínima, además de permitir validar el error experimental.
3. Usar **control local o bloqueo** de fuentes de variación extrañas o conocidas para que ayude a disminuir el error experimental.

Nota 1.4. *Sobre replicación y aleatorización:*

1. **Replicación:** *es el proceso de repetir en condiciones similares el experimento para cada tratamiento, en otras palabras, que las unidades experimentales estén en condiciones iguales al recibir los tratamientos. Este proceso puede ser:*
 - *Balanceado: igual número de unidades experimentales para cada tratamiento aplicado.*
 - *Desbalanceado: diferente número de unidades experimentales por tratamiento, i. e., al menos un tratamiento tiene diferente número de unidades experimental con respecto a los otros tratamientos.*
2. **Aleatorización:** *es el procedimiento que permite tener iguales condiciones al recibir cualquier tratamiento. Debe garantizar que cada unidad experimental tiene igual probabilidad de ser asignada a un tratamiento, independientemente de la unidad experimental.*

Algunas recomendaciones para planear un diseño experimental:

1. Tener claros los objetivos del experimento.
2. Seleccionar las variables respuesta según los objetivos del experimento.
3. Escoger cuidadosamente los factores y niveles.
4. Conocer de forma clara el material experimental.
5. Seleccionar el diseño experimental más adecuado.

6. Asegurar que se satisfagan las condiciones necesarias del experimento.
7. *Analizar los datos.* Implica establecer la matriz diseño (o matriz modelo), X , el vector de parámetros, β , y asociar un modelo,

$$Y = X\beta + \varepsilon,$$

con Y , el vector asociado a la variable respuesta y ε el vector asociado a los errores. Para analizar los datos es necesario:

- Estimar los parámetros del modelo (β y σ^2).
- Conocer la estructura distribucional de los estadísticos de prueba para tomar decisiones.
- Realizar las pruebas de validez del modelo propuesto.

8. Conclusiones y recomendaciones.

En la figura 1.1 se puede observar un esquema de los pasos en el diseño de investigaciones. Cabe notar que la ejecución del experimento y la recolección de los datos es posterior a la formulación del modelo estadístico y no al contrario, como usualmente ocurre.

Ejemplo 1.12. Basado en [33, p. 46] y [40, p. 34], suponga que un investigador desea estudiar y comparar los efectos de agentes contaminantes en la plantación de semillas de pino. El autor considera cuatro tipos de agentes contaminantes denotados por: P_1 , P_2 , P_3 y P_4 ($t = 4$). En el experimento se disponen de cuatro plantaciones de semilla para cada agente contaminante, i.e., se dispone de un total de 16 plantaciones.

- **Situación 1.** Considere la distribución del material experimental según la figura 1.2.

Se tiene que:

- 4 camas con agentes contaminantes son usadas (UE), esto es, $r = 1$.
- Cada cama contiene 4 plantaciones (UO), esto es, $n = 4$.
- Agentes asignados aleatoriamente a las camas.

Principios del diseño de experimentos

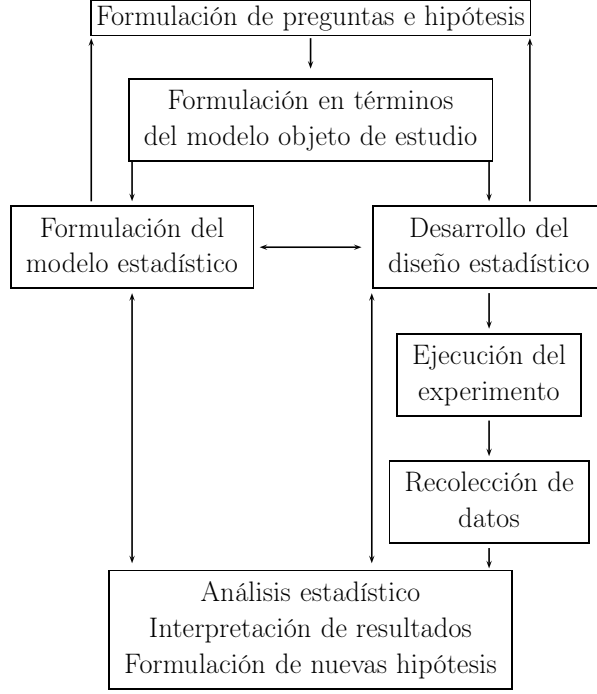


Figura 1.1. Pasos en el diseño de investigaciones

Fuente: [33, p. 31].

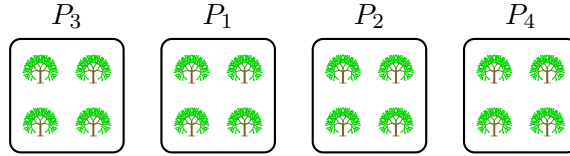


Figura 1.2. Situación 1

En este caso se observa que el efecto de tratamiento y el error experimental están confundidos porque no se sabe si las diferencias se deben a la cama en sí o al agente contaminante. El modelo lineal se puede escribir como:

$$y_{ij} = \mu + P_i + \varepsilon_i + \eta_{ij}, \quad i = 1, 2, 3, 4; \quad j = 1, 2, 3, 4,$$

en que el subíndice i se asocia con la cama y el subíndice j se asocia con la plantación, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ij} \sim N(0, \sigma_\eta^2)$, μ es la media general, P_i es el efecto

to fijo relacionado con el agente contaminante i , y_{ij} es la respuesta asociada a la j -ésima plantación a la cual se le aplica el i -ésimo agente contaminante, y $N(\mu, \sigma^2)$ denota la distribución normal con media μ y varianza σ^2 .

Matricialmente se puede escribir como:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \varepsilon_1 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_2 \\ \varepsilon_2 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_3 \\ \varepsilon_3 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_4 \\ \varepsilon_4 \\ \varepsilon_4 \end{bmatrix} + \begin{bmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \\ \eta_{14} \\ \eta_{21} \\ \eta_{22} \\ \eta_{23} \\ \eta_{24} \\ \eta_{31} \\ \eta_{32} \\ \eta_{33} \\ \eta_{34} \\ \eta_{41} \\ \eta_{42} \\ \eta_{43} \\ \eta_{44} \end{bmatrix}$$

Teniendo en cuenta que el efecto de tratamiento y el error experimental están confundidos, este diseño NO es adecuado para comparar efectos de los agentes contaminantes.

- **Situación 2.** Considere la distribución del material experimental según la figura 1.3.

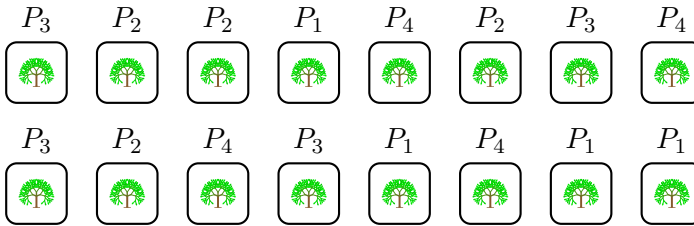


Figura 1.3. Situación 2

Principios del diseño de experimentos

Se tiene que:

- 16 camas con agentes contaminantes son usadas. Cuatro para cada agente contaminante, esto es, $r = 4$.
- Cada cama contiene una plantación (UO), esto es, $n = 1$.
- Los agentes contaminantes son asignados aleatoriamente a las camas.

Entonces, las unidades experimentales son iguales a las unidades observacionales (UE=UO), luego no es posible separar el error observacional del error experimental. El modelo lineal se puede escribir como:

$$y_{ij} = \mu + P_i + \varepsilon_{ij} + \eta_{ij}, \quad i = 1, 2, 3, 4; \quad j = 1, 2, 3, 4,$$

en que $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ij} \sim N(0, \sigma_\eta^2)$, μ es la media general, P_i es el efecto fijo relacionado con el agente contaminante i , y y_{ij} es la respuesta asociada a la j -ésima plantación a la cual se le aplica el i -ésimo agente contaminante.

Matricialmente se puede escribir como:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \end{bmatrix} + \begin{bmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \\ \eta_{14} \\ \eta_{21} \\ \eta_{22} \\ \eta_{23} \\ \eta_{24} \\ \eta_{31} \\ \eta_{32} \\ \eta_{33} \\ \eta_{34} \\ \eta_{41} \\ \eta_{42} \\ \eta_{43} \\ \eta_{44} \end{bmatrix}$$

Con este diseño es posible comparar el efecto de los agentes contaminantes.

- **Situación 3.** Considere la distribución del material experimental según la figura 1.4.

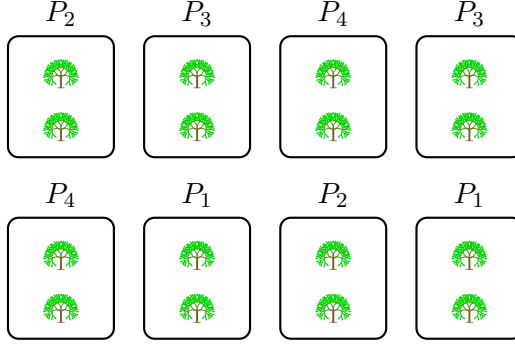


Figura 1.4. Situación 3

Se tiene que:

- Hay ocho camas, dos para cada agente contaminante, esto es $r = 2$.
- En cada cama se asignan dos plantaciones, esto es, $n = 2$.
- Se asignan aleatoriamente los agentes contaminantes a las camas.

Entonces las camas son las unidades experimentales y las plantaciones son las unidades observacionales. El error experimental (EE) se puede medir usando las camas tratadas con el mismo agente contaminante, mientras que el error observacional (EO) se puede medir usando las plantaciones dentro de las camas. El modelo lineal se puede escribir como:

$$y_{ijk} = \mu + P_i + \varepsilon_{ij} + \eta_{ijk}; \quad i = 1, 2, 3, 4; \quad j = 1, 2; \quad k = 1, 2,$$

en que $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ijk} \sim N(0, \sigma_\eta^2)$, μ es la media general, P_i es el efecto fijo relacionado con el agente contaminante i , y_{ijk} es la respuesta asociada al aplicar el i -ésimo agente contaminante en la j -ésima cama (UE) en la k -ésima plantación (UO).

Matricialmente se puede escribir como:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{311} \\ y_{312} \\ y_{321} \\ y_{322} \\ y_{411} \\ y_{412} \\ y_{421} \\ y_{422} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{32} \\ \varepsilon_{41} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{42} \end{bmatrix} + \begin{bmatrix} \eta_{111} \\ \eta_{112} \\ \eta_{121} \\ \eta_{122} \\ \eta_{211} \\ \eta_{212} \\ \eta_{221} \\ \eta_{222} \\ \eta_{311} \\ \eta_{312} \\ \eta_{321} \\ \eta_{322} \\ \eta_{411} \\ \eta_{412} \\ \eta_{421} \\ \eta_{422} \end{bmatrix}$$

Con este diseño se puede tener una estimación de la varianza de los errores experimentales y una estimación de la varianza de los errores observacionales. También se puede comparar el efecto de los agentes contaminantes.

- **Situación 4** Considere la distribución del material experimental según la figura 1.5.

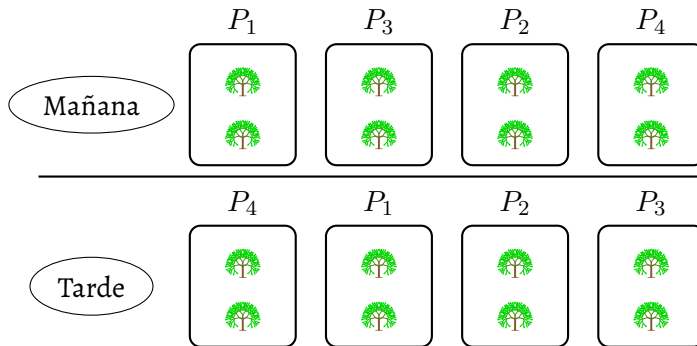


Figura 1.5. Situación 4

Diseño y análisis de experimentos

Se tiene que:

- Hay dos camas para cada agente contaminante, esto es $r = 2$.
- En cada cama se asignan dos plantaciones, esto es, $n = 2$.
- Se asignan aleatoriamente los 4 agentes contaminantes a las camas en la jornada mañana y aleatoriamente los 4 agentes contaminantes a las camas en la jornada tarde. La asignación en la mañana es independiente de la asignación en la tarde.
- Este diseño se usa para considerar en el modelo, por ejemplo, el ritmo diurno de las plantas.

El modelo lineal se puede escribir como:

$$y_{ijk} = \mu + P_i + T_j + \varepsilon_{ij} + \eta_{ijk}; \quad i = 1, 2, 3, 4; \quad j = 1, 2; \quad k = 1, 2,$$

en que $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ijk} \sim N(0, \sigma_\eta^2)$, μ es la media general, P_i es el efecto fijo relacionado con el agente contaminante i , T_j es el efecto fijo relacionado con la jornada de aplicación j , y_{ijk} es la k -ésima observación en la j -ésima jornada (mañana o tarde), del i -ésimo agente contaminante.

Matricialmente sería:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{311} \\ y_{312} \\ y_{321} \\ y_{322} \\ y_{411} \\ y_{412} \\ y_{421} \\ y_{422} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ T_1 \\ T_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{32} \\ \varepsilon_{41} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{42} \end{bmatrix} + \begin{bmatrix} \eta_{111} \\ \eta_{112} \\ \eta_{121} \\ \eta_{122} \\ \eta_{211} \\ \eta_{212} \\ \eta_{221} \\ \eta_{222} \\ \eta_{311} \\ \eta_{312} \\ \eta_{321} \\ \eta_{322} \\ \eta_{411} \\ \eta_{412} \\ \eta_{421} \\ \eta_{422} \end{bmatrix}$$

Con este diseño se puede tener una estimación de la varianza de los errores experimentales y una estimación de la varianza de los errores observacionales. También se está usando bloqueo (jornada mañana y jornada tarde), que al considerarlo en el modelo ayuda a disminuir el error experimental. Usando este diseño se puede comparar el efecto de los agentes contaminantes.

Finalmente, en las tablas 1.1, 1.2 y 1.3 se presentan clasificaciones de los diseños según número de factores de bloqueo, diseño de tratamientos (factoriales) y diseños muestrales (submuestreo).

Tabla 1.1. Jerarquía de los diseños para controlar el error

Número de factores	Clases de diseños
0	Diseño completamente aleatorizado
1	Diseño en bloques aleatorizados
2	Diseños en cuadro latino
3	Diseño en cuadro latino replicado
3	Diseño en cuadro grecolatino
≥ 3	Cuadros latinos mutuamente ortogonales

Fuente: [33, p. 63].

En particular, en el capítulo sobre bloques (capítulo 4) se presentan ejemplos de diseños para controlar el error, en el capítulo sobre factoriales (capítulo 5) se presentan diseños factoriales simétricos, asimétricos y fraccionados; y en la sección de submuestreo (sección 3) se presenta el diseño completamente aleatorizado con submuestreo.

Tabla 1.2. Jerarquía de los diseños de tratamientos

Tipos de factoriales	No. de factores	No. de niveles	No. de tratamientos
Simétrico (o puro)	g	s Con s primo o potencia de primo	s^g
Asimétrico (o mixto)	$g_1 + g_2 + \dots + g_m$	$s_1, s_2, \dots, s_m$	$s_1^{g_1} s_2^{g_2} \dots s_m^{g_m}$
Fraccionado (simétrico)	g	s	s^{g-1}
Fraccionado (asimétrico)	$g_1 + g_2 + \dots + g_m$	$s_1, s_2, \dots, s_m$	$s_1^{g_1-1} s_2^{g_2-1} \dots s_m^{g_m-1}$

Fuente: [33, p. 66].

Tabla 1.3. Jerarquía de los diseños muestrales

Tipos de diseño	No. de UE por tratamiento	No. de submuestras por UE	No. de sub-submuestras por submuestra	No. de observaciones por tratamiento
No hay submuestreo (UE=UO)	r	1	No aplica	r
Hay submuestreo	r'	n^*	No aplica	$r' n^*$
Hay sub-submuestreo	r'	n^*	m	$r' n^* m$

Fuente: [33, p. 67].



Capítulo 2

Introducción a los modelos lineales

En este capítulo se presentan conceptos de modelos lineales necesarios para, una vez propuesto un modelo, estimar el vector de parámetros y, si se considera que los errores provienen de variables aleatorias independientes, homocedásticas y con distribución normal, realizar el análisis de varianza. También se presenta el concepto de estimabilidad y su implicación en el momento de interpretar los parámetros estimados.

Según [25], una *función lineal* es un conjunto de variables, parámetros, etc., que no contiene potencias o productos cruzados de las cantidades y un *modelo lineal* es un modelo en el cual el valor esperado de una variable aleatoria es expresado como una función lineal de los parámetros en el modelo. Por otro lado, [33, p. 71] expresan que un modelo es una explicación de observables en términos de observables.

[33, p.71] clasifican los modelos, en general, desde dos puntos de vista, como:

- Estáticos o dinámicos:
 - **Modelos estáticos.** Corresponden a los modelos que describen una situación, por ejemplo, la medida del tiempo t que demora en caer un objeto desde una determinada altura s , $s = \frac{1}{2}gt^2$ con s altura en metros, g constante gravitacional ($g = 9.8m/s^2$) y t tiempo de caída en segundos.
 - **Modelos dinámicos.** Corresponden a los modelos que en caso de ser correctos, “ayudarían” a *prever* el futuro.
- Explicativos o causales:
 - **Modelos explicativos.** Corresponden a los modelos en que NO hay intervenciones por parte del observador.
 - **Modelos causales.** Corresponden a los modelos en que hay intervención de parte del observador.

Para representar **un modelo lineal**, suponga que una variable Y , que se llamará *variable dependiente*, se puede expresar en términos de las variables $X_1, X_2, \dots, X_p$, que se llamarán *variables independientes*, y que se observan en n unidades.

Organizando las observaciones en matrices se tiene que:

$$\mathbf{y}_{n \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad y \quad \mathbf{X}_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_p \end{pmatrix},$$

con $\mathbf{x}_j = (x_{1j} \ x_{2j} \ x_{3j} \ \dots \ x_{nj})^\top$. A $\mathbf{X}$ se le llama **matriz diseño**.

Un modelo lineal en los parámetros es dado por:

$$\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \quad (2.1)$$

en que $\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_p)^\top$ es un vector de parámetros a estimar.

Nota 2.1. En un modelo lineal NO hay relación entre los coeficientes $\beta_1, \beta_2, \dots, \beta_p$. Sea $\mathbb{E}(\cdot)$ tal que denota el operador esperanza, algunos ejemplos de modelos lineales son:

- $\mathbb{E}(Y_{ij}) = \mu + \gamma_i + \delta_j$, con $i = 1, 2, \dots, a$ y $j = 1, 2, \dots, b$.
- $\mathbb{E}(Y_{ij}) = \mu + \gamma_i + \delta_j + (\gamma\delta)_{ij}$, con $i = 1, 2, \dots, a$ y $j = 1, 2, \dots, b$.
- $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i$, con $i = 1, 2, \dots, n$.
- $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$, con $i = 1, 2, \dots, n$.
- $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}$, con $i = 1, 2, \dots, n$.
- $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i}$, con $i = 1, 2, \dots, n$.
- $\mathbb{E}(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 \ln(x_{1i}) + \beta_3 [\ln(x_{1i})]^2$, con $i = 1, 2, \dots, n$.

Un modelo no lineal en los parámetros es, por ejemplo, $\mathbb{E}(y_i) = \beta_1 x_1 + \beta_1^2 x_2 + \beta_2 x_3$.

El modelo (2.1) se usa para describir un modelo lineal estocástico que es un vector aleatorio tal que $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ con $\boldsymbol{\varepsilon}$ un vector aleatorio que sigue alguna distribución de probabilidad. Cabe notar el cambio de $\mathbf{y}$ por $\mathbf{Y}$. El primer caso denota observaciones (realizaciones de la variable aleatoria $\mathbf{Y}$) y el segundo caso denota variables aleatorias en que la aleatoriedad viene dada por $\boldsymbol{\varepsilon}$.

Otra clasificación dada en [33, p. 74] de los modelos lineales es:

- **Modelos funcionales.** Son, por ejemplo, los modelos de regresión lineal simple/múltiple en que uno de sus problemas es posible multicolinealidad, esto es, $\gamma_1 \mathbf{x}_1 + \gamma_2 \mathbf{x}_2 + \dots + \gamma_p \mathbf{x}_p = 0$.

- **Modelos de clasificación.** Por ejemplo,

$$\mathbb{E}(\mathbf{y}) = \mu + \text{efecto de bloque} + \text{efecto de tratamiento}.$$

- **Modelos de componentes clasificatorias y funcionales.** Por ejemplo,

$$\mathbb{E}(y_{ij}) = \mu + \beta_j + \tau_i + \gamma x_{ij},$$

en que μ , β_j , τ_i y γ son parámetros del modelo a estimar y x_{ij} es una variable concomitante o covariable.

Nota 2.2. Sea β un vector de parámetros, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$. Algunos ejemplos de combinaciones lineales de los parámetros inseridos en β son:

- $\beta_1 - \beta_2$.
- $\beta_1 - \beta_3$.
- $\beta_2 + \beta_4 - 2\beta_3$.

1. Método de estimación por mínimos cuadrados

Suponga que las variables aleatorias $Y_1, Y_2, Y_3, \dots, Y_n$, se pueden escribir de la forma

$$Y_i = g_i(\theta_1, \theta_2, \dots, \theta_r) + \varepsilon_i, \quad 1 \leq i \leq n,$$

con g_i funciones conocidas y $\theta_1, \theta_2, \dots, \theta_r$, parámetros de interés, de valor desconocido. Suponga que el vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_r)^\top$ puede variar libremente sobre un conjunto $A \subset \mathbb{R}^r$. También, suponga que los ε_i satisfacen, al menos en forma aproximada, las siguientes restricciones:

- $\mathbb{E}(\varepsilon_i) = 0, 1 \leq i \leq n.$
- $\mathbb{V}(\varepsilon_i) = \sigma^2 > 0, 1 \leq i \leq n.$
- $\mathbb{Cov}(\varepsilon_i, \varepsilon_j) = 0, 1 \leq i < j \leq n.$

Cabe notar que σ^2 es, también, un parámetro.

Sea $\mathbf{g}(\boldsymbol{\theta})$ dado por:

$$\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_n(\boldsymbol{\theta}))^\top$$

y $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)^\top$, el valor esperado de $\mathbf{Y}$ es:

$$\mathbb{E} \left[\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \right] = \begin{pmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \\ \vdots \\ g_n(\boldsymbol{\theta}) \end{pmatrix}$$

$$\mathbb{E}(\mathbf{Y}) = \mathbf{g}(\boldsymbol{\theta}).$$

Una vez se observa $\mathbf{Y} = \mathbf{y}$, con $\mathbf{y} = (y_1, y_2, y_3, \dots, y_n)^\top$, el estimador obtenido vía mínimos cuadrados resulta de minimizar

$$\sum_{i=1}^n [y_i - g_i(\theta_1, \dots, \theta_r)]^2 = [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})]^\top [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})].$$

Si las $g_i, i = 1, 2, \dots, n$, son diferenciables, $\hat{\boldsymbol{\theta}}$ está bien definido, ver [8, p. 100].

Para minimizar

$$Q = [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})]^\top [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})] = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

se deriva Q con respecto a $\boldsymbol{\theta}$, $\frac{dQ}{d\boldsymbol{\theta}}$, y se iguala a cero, $\frac{dQ}{d\boldsymbol{\theta}} = 0$.

Nota 2.3. Según [30, p. 299], definiendo $\frac{d}{d\boldsymbol{\beta}} = \left[\left(\frac{\partial}{\partial \beta_i} \right) \right]$ se tiene que:

a) para un vector $\mathbf{a}$ de tamaño adecuado:

$$\frac{d(\boldsymbol{\beta}^\top \mathbf{a})}{d\boldsymbol{\beta}} = \mathbf{a}, \quad (2.2)$$

b) y para una matriz $\mathbf{A}$, simétrica, de tamaño adecuado:

$$\frac{d(\beta^\top \mathbf{A} \beta)}{d\beta} = 2\mathbf{A}\beta. \quad (2.3)$$

Usando (2.2) y (2.3):

$$\begin{aligned} \frac{dQ}{d\beta} &= \frac{d}{d\beta} \left(\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X} \beta - \beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta \right) \\ &= \frac{d}{d\beta} \left(\mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta \right) \\ &= -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \beta, \end{aligned}$$

igualando a 0,

$$\begin{aligned} \frac{dQ}{d\beta} &= 0 \\ -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \hat{\beta} &= 0 \\ \mathbf{X}^\top \mathbf{X} \hat{\beta} &= \mathbf{X}^\top \mathbf{y}. \end{aligned} \quad (2.4)$$

A las ecuaciones (2.4) se les llama **ecuaciones normales**.

Si $\mathbf{X}^\top \mathbf{X}$ es no singular,

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.5)$$

Si $\mathbf{X}$ no es de rango columna completo, $\mathbf{X}^\top \mathbf{X}$ es singular, ver anexo A.

2. Estimabilidad

Ejemplo 2.1. Considere los datos presentados en la tabla 2.1.

1. Presente las respectivas matrices $\mathbf{y}$, $\mathbf{X}$, β y ϵ .
2. Presente posibles soluciones para β vía mínimos cuadrados.
3. Estime $\hat{\tau}_3 - \hat{\tau}_1$.

Tabla 2.1. Datos ilustrativos. Una vía de clasificación

Categoría		
A	B	C
8	10	13
10	15	16
12	.	13
13	12	15

Solución

1. Las respectivas matrices $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$ y $\boldsymbol{\varepsilon}$ son dadas por

$$\mathbf{y} = \begin{pmatrix} 8 \\ 10 \\ 12 \\ 13 \\ 10 \\ 15 \\ 12 \\ 13 \\ 16 \\ 13 \\ 15 \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}; \quad \mathbf{y} - \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{pmatrix}.$$

2. Cualquier solución para $\boldsymbol{\beta}$, vía mínimos cuadrados, es de la forma, ver (A.3),

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \begin{pmatrix} \hat{\mu} & \hat{\tau}_1 & \hat{\tau}_2 & \hat{\tau}_3 \end{pmatrix}^\top \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} + (\mathbf{I} - (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X})) \mathbf{z}, \end{aligned}$$

con $\mathbf{z}$ arbitrario.

$$\hat{\beta} = \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 137 \\ 43 \\ 37 \\ 57 \end{pmatrix} + \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix} \right] \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}$$

Así, dos posibles soluciones para $\hat{\beta}$, tomando por ejemplo $z = \mathbf{0}$ son:

$$\begin{aligned} \hat{\beta}_1 &= \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 137 \\ 43 \\ 37 \\ 57 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0.25 \end{pmatrix} \begin{pmatrix} 137 \\ 43 \\ 37 \\ 57 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 10.75 & 12.33 & 14.25 \end{pmatrix}^{\top} \\ \hat{\beta}_2 &= \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 137 \\ 43 \\ 37 \\ 57 \end{pmatrix} \\ &= \begin{pmatrix} 0.25 & -0.25 & -0.25 & 0 \\ -0.25 & 0.5 & 0.25 & 0 \\ -0.25 & 0.25 & 0.583 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 137 \\ 43 \\ 37 \\ 57 \end{pmatrix} = (14.25 \ -3.5 \ -1.917 \ 0)^{\top}. \end{aligned}$$

Si z se toma diferente de cero, por ejemplo $z = (1 \ 1 \ 1 \ 1)^{\top}$, ya que es arbitrario, otros dos posibles valores para la estimación de β pueden ser:

$$\begin{aligned}
 \hat{\beta}_3 &= \begin{pmatrix} 0 \\ 10.75 \\ 12.33 \\ 14.25 \end{pmatrix} + \\
 &\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0.25 \end{pmatrix} \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} 0 \\ 10.75 \\ 12.33 \\ 14.25 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 9.75 \\ 11.33 \\ 13.25 \end{pmatrix} \\
 \hat{\beta}_4 &= \begin{pmatrix} 0 \\ 10.75 \\ 12.33 \\ 14.25 \end{pmatrix} + \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.25 & -0.25 & -0.25 & 0 \\ -0.25 & 0.5 & 0.25 & 0 \\ -0.25 & 0.25 & 0.583 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 11 & 4 & 3 & 4 \\ 4 & 4 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} 0 \\ 10.75 \\ 12.33 \\ 14.25 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 11.75 \\ 13.33 \\ 15.25 \end{pmatrix}
 \end{aligned}$$

Implicando que, por ejemplo, la estimación de τ_1 dé múltiples valores, por ejemplo: $\hat{\tau}_1 = 10.75$ o $\hat{\tau}_1 = -3.5$ o $\hat{\tau}_1 = 9.75$ o $\hat{\tau}_1 = 11.75$, entonces **¿cómo interpretar a $\hat{\tau}_1$?**

3. Para evaluar $\hat{\tau}_3 - \hat{\tau}_1$, usando las cuatro estimaciones de β se tiene que:

$$\begin{aligned}
 \hat{\tau}_3 - \hat{\tau}_1 &= \underbrace{\begin{pmatrix} 0 & -1 & 0 & 1 \end{pmatrix}}_{\lambda^\top} \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{pmatrix} = \underbrace{14.25 - 10.75}_{\text{usando } \hat{\beta}_1} = 3.5 \\
 &= \underbrace{0 - (-3.5)}_{\text{usando } \hat{\beta}_2} = 3.5 \\
 &= \underbrace{13.25 - 9.75}_{\text{usando } \hat{\beta}_3} = 3.5 \\
 &= \underbrace{15.25 - 11.75}_{\text{usando } \hat{\beta}_4} = 3.5
 \end{aligned}$$

esto es, se obtiene el mismo resultado.

✓

DEFINICIÓN 2.1. Sea $\lambda^\top \beta$, una combinación lineal de β , el vector de parámetros de interés. Una **función lineal estimable** es una función lineal de los parámetros para los cuales se puede encontrar un **estimador invariante** a soluciones de las ecuaciones normales, $X^\top X \hat{\beta} = X^\top Y$.

Ejemplo 2.2. Sea

$$X_1 \sim N(\beta_1 + V, 1) \quad y \quad X_2 \sim N(\beta_2 + V, 1), \quad \text{luego} \quad \theta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ V \end{pmatrix}$$

tal que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_1 + V \\ \beta_2 + V \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],$$

$$\text{esto es, } \mu = \begin{pmatrix} \beta_1 + V \\ \beta_2 + V \end{pmatrix} y V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Suponga

$$\theta_1 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \quad \text{entonces} \quad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$$

y

$$\boldsymbol{\theta}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{entonces} \quad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],$$

por tanto

$$f_{X_1, X_2}(x_1, x_2; \boldsymbol{\theta}_1) = f_{X_1, X_2}(x_1, x_2; \boldsymbol{\theta}_2)$$

a pesar de que $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$.

Entonces $\boldsymbol{\theta}$ no es estimable¹.

En el caso de un modelo lineal,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.6)$$

con $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, se dice que $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ es linealmente **estimable** si existe un vector $\mathbf{a}$ tal que

$$\begin{aligned} \mathbb{E}(\mathbf{a}^\top \mathbf{Y}) &= \boldsymbol{\lambda}^\top \boldsymbol{\beta} \\ \mathbb{E}(\mathbf{a}^\top [\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}]) &= \boldsymbol{\lambda}^\top \boldsymbol{\beta} \\ \mathbf{a}^\top \mathbf{X}\boldsymbol{\beta} + \mathbf{a}^\top \mathbb{E}[\boldsymbol{\varepsilon}] &= \boldsymbol{\lambda}^\top \boldsymbol{\beta} \\ \mathbf{a}^\top \mathbf{X}\boldsymbol{\beta} &= \boldsymbol{\lambda}^\top \boldsymbol{\beta}. \end{aligned}$$

Así, dado $\boldsymbol{\beta}$, un vector completamente libre, $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ es **estimable** si existe un vector $\mathbf{a}$ tal que $\boldsymbol{\lambda}^\top = \mathbf{a}^\top \mathbf{X}$, esto es, $\boldsymbol{\lambda} \in \mathcal{R}(\mathbf{X})$, el espacio fila de $\mathbf{X}$ [33, p. 78].

Ejemplo 2.3. Según [40, p. 82], sean

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \\ 1 & 3 & 6 \end{pmatrix} \quad y \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Determine si β_0 es estimable.

¹Ni *identificable* en un contexto más amplio al de los modelos lineales [22].

Se dice que una familia paramétrica $\{P_\theta : \theta \in \Theta\}$ es **identificable** si y solo si $\theta_1 \neq \theta_2$ y $\theta_i \in \Theta$, $i = 1, 2$, implica que $P_{\theta_1} \neq P_{\theta_2}$ (ver [58, p. 94]).

Solución Se tiene que:

$$\beta_0 = \underbrace{\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}}_{\lambda^\top} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \lambda^\top \beta,$$

luego $\lambda = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^\top$. La pregunta es ¿existe $\mathbf{a}$ tal que $\lambda^\top = \mathbf{a}^\top \mathbf{X}$?, ¿quien sería $\mathbf{a}$?

$$\text{Sea } \mathbf{a} = \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathbf{X}^\top \mathbf{a} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 3 \\ 2 & 4 & 2 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \lambda,$$

entonces β_0 es estimable. ✓

Nota 2.4. Sea $\lambda^\top \beta$ una combinación lineal de β ,

$$\lambda^\top \beta \text{ es estimable si y solo si } \lambda^\top (\mathbf{X}^\top \mathbf{X})^- (\mathbf{X}^\top \mathbf{X}) = \lambda^\top,$$

con $(\mathbf{X}^\top \mathbf{X})^-$ una inversa generalizada de $\mathbf{X}^\top \mathbf{X}$.

3. Mejor estimador linealmente insesgado

DEFINICIÓN 2.2. Si $\lambda^\top \beta$ es estimable y $\mathbf{a}^\top \mathbf{y}$ es un estimador lineal insesgado, que tiene menor varianza entre todos los estimadores lineales insesgados, entonces $\mathbf{a}^\top \mathbf{y}$ es el mejor estimador lineal insesgado de $\lambda^\top \beta$, denotado por $MELI(\lambda^\top \beta)$ [40, p. 85].

DEFINICIÓN 2.3. El modelo lineal de Gauss-Markov es dado por el modelo lineal estocástico, dado en (2.6), adicionando la condición de $\mathbb{V}(\boldsymbol{\varepsilon}) = \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \sigma^2 \mathbf{I}^2$ [33, p. 128].

Teorema 2.1 (Teorema de Gauss- Markov). Bajo el modelo lineal de Gauss Markov, si $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ es estimable entonces el MELI $(\boldsymbol{\lambda}^\top \boldsymbol{\beta}) = \boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}$ para todo $\hat{\boldsymbol{\beta}}$, solución de las ecuaciones $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{Y}$.

Nota 2.5. El valor esperado de $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}$ y su varianza son

$$\begin{aligned} \mathbb{E}(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) &= \mathbb{E} \left(\boldsymbol{\lambda}^\top \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} + \left[\mathbf{I} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \right] \mathbf{z} \right\} \right) \\ &= \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}(\mathbf{Y}) + \underbrace{\left[\boldsymbol{\lambda}^\top - \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \right]}_{\boldsymbol{\lambda}^\top} \mathbf{z} \\ &= \underbrace{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}}_{\boldsymbol{\lambda}^\top} \boldsymbol{\beta} \quad \text{porque } \boldsymbol{\lambda}^\top \boldsymbol{\beta} \text{ es estimable} \\ &= \boldsymbol{\lambda}^\top \boldsymbol{\beta}. \end{aligned}$$

Si $\mathbb{V}(\mathbf{Y}) = \sigma^2 \mathbf{I}$,

$$\begin{aligned} \mathbb{V}(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) &= \mathbb{V} \left(\boldsymbol{\lambda}^\top \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} + \left[\mathbf{I} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \right] \mathbf{z} \right\} \right) \\ &= \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{V}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \\ &= \sigma^2 \underbrace{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}}_{\boldsymbol{\lambda}^\top} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} = \sigma^2 \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}. \end{aligned}$$

Nota 2.6. Si $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ es estimable con $\boldsymbol{\lambda}^\top$ de rango fila completo, se puede plantear la hipótesis

$$H_0 : \boldsymbol{\lambda}^\top \boldsymbol{\beta} = 0,$$

la cual tiene suma de cuadrados asociada [56, p. 192]:

$$SC_{H_0} = (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})^\top \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}).$$

²Según [34, p. 10], la matriz de covarianza $\mathbf{V} = \mathbb{V}(\mathbf{Y})$ del vector aleatorio $\mathbf{Y}$ es dada por:

$$\mathbf{V} = \mathbb{V}(\mathbf{Y}) = \mathbb{E} \left\{ [\mathbf{Y} - \mathbb{E}(\mathbf{Y})] [\mathbf{Y} - \mathbb{E}(\mathbf{Y})]^\top \right\}. \quad (2.7)$$

Ejemplo 2.4. Considere el modelo:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

con $i = 1, 2, 3; j = 1, 2, \dots, n_i; n_1 = 4, n_2 = 2, n_3 = 1$, tal que $\mathbb{E}(\varepsilon_{ij}) = 0$ y $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$, tal que:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{pmatrix} 13 \\ 12 \\ 5 \\ 8 \\ 2 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{31} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \end{pmatrix}$$

1. Presente una solución para $\boldsymbol{\beta}$.
2. ¿ $\tau_1 - \tau_2$ es estimable?
3. Si $\tau_1 - \tau_2$ es estimable, ¿cuál es el MELI($\tau_1 - \tau_2$)?
4. Si $\tau_1 - \tau_2$ es estimable, ¿cuál es la $\mathbb{V}[\text{MELI}(\tau_1 - \tau_2)]$?
5. Si $\tau_1 - \tau_2$ es estimable, ¿cuál es la suma de cuadrados asociada a la $H_0 : \tau_1 - \tau_2 = 0$?

Solución

1. Una posible solución de las ecuaciones normales es:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} + [\mathbf{I} - (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X})] \mathbf{z}, \quad \text{con } \mathbf{z} \text{ arbitrario}$$

$$= \begin{pmatrix} 7 & 4 & 2 & 1 \\ 4 & 4 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 46 \\ 38 \\ 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 46 \\ 38 \\ 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 9.5 \\ 3 \\ 2 \end{pmatrix}.$$

2. Para la estimabilidad,

$$\begin{aligned}
 (\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X}) &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & 4 & 2 & 1 \\ 4 & 4 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \\
 \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X}) &= \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} = \boldsymbol{\lambda}^\top,
 \end{aligned}$$

por tanto $\tau_1 - \tau_2$ sí es estimable.

3. El $MELI(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}$, entonces:

$$MELI(\tau_1 - \tau_2) = \boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} = \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 9.5 \\ 3 \\ 2 \end{pmatrix} = 6.5.$$

4. Se tiene que $\mathbb{V}[MELI(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})] = \mathbb{V}(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}$.

$$\begin{aligned}
 \mathbb{V}[MELI(\tau_1 - \tau_2)] &= \sigma^2 \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \\
 &= \sigma^2 \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \frac{3}{4} \sigma^2.
 \end{aligned}$$

5. La suma de cuadrados asociada a $H_0 : \boldsymbol{\lambda}^\top \boldsymbol{\beta} = \mathbf{0}$ es dada por

$SC_{H_0} = (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})^\top (\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda})^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})$. Entonces:

$$SC(MELI(\tau_1 - \tau_2)) = SC(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = (6.5) \left(\frac{3}{4}\right)^{-1} (6.5) = 56.3333.$$

✓

Retomando la ecuación (2.4) sobre las ecuaciones normales, se tiene que:

- Las ecuaciones normales son consistentes para $\mathbf{y} \in \mathbb{R}^n$, esto es, existe un vector $\hat{\beta}$ que es solución de las ecuaciones normales dado un $\mathbf{y}$.
- Cualquiera sea la solución $\hat{\beta}$, $\mathbf{X}\hat{\beta}$ es única dado un $\mathbf{y}$ particular porque

$$\mathbf{X}\hat{\beta} = \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\text{Invariante a } (\mathbf{X}^\top \mathbf{X})^{-1}} \mathbf{y} = \mathbf{P}_x \mathbf{y},$$

con $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$, a $\mathbf{P}_x$ se le llama matriz de proyección.

4. Matriz de proyección ($\mathbf{P}_x$)

La matriz $\mathbf{P}_x$, que es simétrica e idempotente³, se puede usar para reescribir $\mathbf{y}$ como:

$$\mathbf{y} = \mathbf{P}_x \mathbf{y} + (\mathbf{I} - \mathbf{P}_x) \mathbf{y}, \quad (2.8)$$

en que $\mathbf{P}_x \mathbf{y}$ es el ajuste de $\mathbf{y}$ usando $\mathbf{X}\beta$, e $(\mathbf{I} - \mathbf{P}_x) \mathbf{y}$ es el residual.

Nota 2.7. Observaciones sobre $\mathbf{P}_x$:

- $\mathbf{P}_x(\mathbf{I} - \mathbf{P}_x) = \mathbf{P}_x - \mathbf{P}_x^2 = \mathbf{P}_x - \mathbf{P}_x = \mathbf{0}$ y por tanto $\mathbf{P}_x \mathbf{y}$ y $(\mathbf{I} - \mathbf{P}_x) \mathbf{y}$ son vectores perpendiculares de orden $n \times 1$.
- $\text{rango}(\mathbf{P}_x) = \text{rango}(\mathbf{X})$ y $\text{rango}(\mathbf{I} - \mathbf{P}_x) = n - \text{rango}(\mathbf{P}_x) = n - \text{rango}(\mathbf{X})$.

Premultiplicando a (2.8) por $\mathbf{y}^\top$ se tiene que:

$$\underbrace{\mathbf{y}^\top \mathbf{y}}_{SCT} = \underbrace{\mathbf{y}^\top \mathbf{P}_x \mathbf{y}}_{\underbrace{SC\mathbf{X}\beta}_{SCM}} + \underbrace{\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{y}}_{SC\text{residual}},$$

³Una matriz cuadrada $\mathbf{A}$, de orden n , es idempotente si $\mathbf{A}\mathbf{A} = \mathbf{A}$. Entre sus propiedades se tiene que $\text{rango}(\mathbf{A}) = \text{traza}(\mathbf{A})$ y $\text{rango}(\mathbf{I} - \mathbf{A}) = \text{traza}(\mathbf{I} - \mathbf{A}) = n - \text{traza}(\mathbf{A})$ [30, pp. 133-135].

en que SCT denota la **suma de cuadrados total**, SCM denota la **suma de cuadrados del modelo**, esto es, del ajuste de $\mathbf{y}$ usando $\mathbf{X}\beta$, y $SC_{residual}$ denota la **suma de cuadrados del residual**. En la tabla 2.2 se presenta una organización de estas suma de cuadrados.

Tabla 2.2. Análisis de varianza simple

Causa de variación	Grados de libertad	Suma de cuadrados
$\mathbf{X}$	$\text{rango}(\mathbf{X})$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y}$
Residual	$n - \text{rango}(\mathbf{X})$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$
Total	n	$\mathbf{Y}^\top \mathbf{Y}$

Fuente: [33, p. 80].

Nota 2.8. Un modelo lineal $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\beta$ de rango r sobre un vector de parámetros $p \times 1$ ($p \geq r$) se puede escribir como un modelo $\mathbb{E}(\mathbf{Y}) = \mathbf{Z}\theta$ con $\mathbf{Z}_{n \times r}$ de rango r . Esto significa que θ es un conjunto de r funciones estimables o identificables linealmente independientes. A esta reescritura se le llama **reparametrización de rango completo** [33, p. 81].

Nota 2.9. Otro proceso para obtener la solución de las ecuaciones consistentes, $\mathbf{A}\mathbf{x} = \mathbf{d}$, es agregar restricciones (ecuaciones consistentes de la forma $\mathbf{C}\mathbf{x} = \mathbf{c}$), que ayudan a obtener una solución única [33, p. 82].

5. Modelo lineal en dos partes

Suponga que se desea explicar o aproximar un vector de variables $\mathbf{Y}$ usando un modelo lineal sobre las variables explicativas $x_1, x_2, \dots, x_p$ y $z_1, z_2, \dots, z_q$. Dos preguntas interesantes en este caso son:

- ¿Las variables $z_1, z_2, \dots, z_q$ ayudan a explicar $\mathbf{Y}$ después de ajustar $x_1, x_2, \dots, x_p$?

- ¿Las variables $x_1, x_2, \dots, x_p$ ayudan a explicar Y después de ajustar $z_1, z_2, \dots, z_q$?

En otras palabras, ¿ajustar Y por $X_1\beta_1$ y luego por $X_2\beta_2$ es diferente de ajustar Y por $X_2\beta_2$ y luego por $X_1\beta_1$?, en que X_1 está asociado a las variables explicativas $x_1, x_2, \dots, x_p$ y X_2 está asociado a las variables explicativas $z_1, z_2, \dots, z_q$.

La respuesta es que en general es diferente, esto es:

$$\mathbb{E}(Y) = X_1\beta_1 + X_2\beta_2 \neq \mathbb{E}(Y) = X_2\beta_2 + X_1\beta_1, \quad (2.9)$$

que corresponden a dos **modelos lineales ordenados en dos partes**. En la tabla 2.3 se presenta el respectivo análisis de varianza.

Tabla 2.3. Tabla de análisis de varianza para un modelo lineal ordenado en dos partes

Causa de variación	Grados de libertad	Suma de cuadrados
$X_1: X_1\beta_1$	$\text{rango}(P_1) = r_1$	$Y^\top P_1 Y$
X_2/X_1	$\text{rango}(P_{12} - P_1) = r_{12} - r_1$	$Y^\top (P_{12} - P_1) Y$
$I/X_1 X_2$:Residual	$\text{rango}(I - P_{12}) = n - r_{12}$	$Y^\top (I - P_{12}) Y$
Total	$\text{rango}(I) = n$	$Y^\top Y$

Fuente: [33, p. 93].

$$P_1 = X_1(X_1^\top X_1)^{-1} X_1^\top, P_2 = X_2(X_2^\top X_2)^{-1} X_2^\top \text{ y } P_{12} = P_x.$$

Se obtiene un ANOVA ortogonal, esto es la igualdad en (2.9), cuando:

$$Y^\top (P_{12} - P_1) Y = Y^\top P_2 Y \quad \text{para todo } Y$$

y

$$Y^\top (P_{12} - P_2) Y = Y^\top P_1 Y \quad \text{para todo } Y,$$

o

$$\mathbf{P}_{12} = \mathbf{P}_1 + \mathbf{P}_2$$

o

$$\mathbf{X}_1^\top \mathbf{P}_{12} \mathbf{X}_2 = \mathbf{X}_1^\top \mathbf{P}_1 \mathbf{X}_2 + \mathbf{X}_1^\top \mathbf{P}_2 \mathbf{X}_2$$

o

$$\mathbf{X}_1^\top \mathbf{X}_2 = \mathbf{X}_1^\top \mathbf{X}_2 + \mathbf{X}_1^\top \mathbf{X}_2$$

o

$$\mathbf{X}_1^\top \mathbf{X}_2 = \mathbf{0}.$$

6. Distribución de formas lineales y cuadráticas

DEFINICIÓN 2.4. Según [34, p. 40], sea el vector de variables aleatorias $\mathbf{Y}$, y sean $\mathbf{A} = \{a_{ij}\}$ y $\mathbf{B} = \{b_{kl}\}$ matrices de dimensión $N \times N$ y $t \times N$, respectivamente, en que los elementos de $\mathbf{A}$ y $\mathbf{B}$ son constantes conocidas y $\mathbf{A}$ es simétrica, entonces:

1. $\mathbf{L} = \mathbf{B}\mathbf{Y}$ es llamada una **forma lineal** en $\mathbf{Y}$.
2. $\mathbf{Q} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ es llamada una **forma cuadrática** en $\mathbf{Y}$.

Ejemplo 2.5. Escribir el promedio y la varianza muestral como una forma lineal y como una forma cuadrática, respectivamente.

Solución

1. $\bar{Y} = \sum_{i=1}^n \frac{Y_i}{n}$ se puede escribir como una forma lineal usando $\mathbf{B} = \frac{1}{n} \mathbf{1}_n^\top$.
2. $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ se puede escribir como una forma cuadrática usando $\mathbf{A} = \frac{1}{n-1} \left(\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)$. ✓

Nota 2.10. Si $\mathbf{A}$ no es simétrica, esto es, $a_{ij} \neq a_{ji}$, siempre se puede escribir $\mathbf{Q}$ (forma cuadrática) usando la matriz simétrica [56, p. 34]:

$$a_{ij}^* = a_{ji}^* = \frac{1}{2} (a_{ij} + a_{ji}).$$

6.1. Momentos de $\mathbf{L}$ y $\mathbf{Q}$

Si $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}$ y $\mathbb{V}(\mathbf{Y}) = \mathbf{V}_{n \times n}$,

$$\mathbb{E}(\mathbf{L}) = \mathbb{E}(\mathbf{BY}) = \mathbf{B}\mathbb{E}(\mathbf{Y}) = \mathbf{B}\boldsymbol{\mu}.$$

$$\mathbb{V}(\mathbf{L}) = \mathbb{V}(\mathbf{BY}) = \mathbf{B}\mathbb{V}(\mathbf{Y})\mathbf{B}^\top = \mathbf{B}\mathbf{V}\mathbf{B}^\top.$$

Si $\mathbf{L}_1 = \mathbf{B}_1\mathbf{Y}$ y $\mathbf{L}_2 = \mathbf{B}_2\mathbf{Y}$ son dos formas lineales en $\mathbf{Y}$, entonces:

$$\begin{aligned} \mathbb{Cov}(\mathbf{L}_1, \mathbf{L}_2) &= \mathbb{E}[(\mathbf{L}_1 - \mathbf{B}_1\boldsymbol{\mu})(\mathbf{L}_2 - \mathbf{B}_2\boldsymbol{\mu})^\top] \\ &= \mathbb{E}[(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_1\boldsymbol{\mu})(\mathbf{B}_2\mathbf{Y} - \mathbf{B}_2\boldsymbol{\mu})^\top] \\ &= \mathbb{E}(\mathbf{B}_1\mathbf{Y}\mathbf{Y}^\top\mathbf{B}_2^\top - \mathbf{B}_1\mathbf{Y}\boldsymbol{\mu}^\top\mathbf{B}_2^\top - \mathbf{B}_1\boldsymbol{\mu}\mathbf{Y}^\top\mathbf{B}_2^\top + \mathbf{B}_1\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{B}_2^\top) \\ &= \mathbf{B}_1\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)\mathbf{B}_2^\top - \mathbf{B}_1\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{B}_2^\top - \mathbf{B}_1\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{B}_2^\top + \mathbf{B}_1\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{B}_2^\top \\ &= \mathbf{B}_1[\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top]\mathbf{B}_2^\top \\ &= \mathbf{B}_1\mathbf{V}\mathbf{B}_2^\top. \end{aligned}$$

El valor esperado de la forma cuadrática $\mathbf{Q}$ es:

$$\begin{aligned} \mathbb{E}(\mathbf{Q}) &= \mathbb{E}[\mathbf{Y}^\top \mathbf{A} \mathbf{Y}] = \mathbb{E}[\text{traza}(\mathbf{Y}^\top \mathbf{A} \mathbf{Y})] \underbrace{=}_{\text{4}} \mathbb{E}[\text{traza}(\mathbf{A} \mathbf{Y} \mathbf{Y}^\top)] \\ &= \mathbb{E}[\text{traza}(\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top + \mathbf{A} \mathbf{Y} \boldsymbol{\mu}^\top + \mathbf{A} \boldsymbol{\mu} \mathbf{Y}^\top - \mathbf{A} \boldsymbol{\mu} \boldsymbol{\mu}^\top)] \\ &= \mathbb{E}[\text{traza}(\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top) + \text{traza}(\mathbf{A} \mathbf{Y} \boldsymbol{\mu}^\top) + \text{traza}(\mathbf{A} \boldsymbol{\mu} \mathbf{Y}^\top) - \text{traza}(\mathbf{A} \boldsymbol{\mu} \boldsymbol{\mu}^\top)] \\ &= \mathbb{E}[\text{traza}(\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top)] + \mathbb{E}[\text{traza}(\mathbf{A} \mathbf{Y} \boldsymbol{\mu}^\top)] + \\ &\quad \mathbb{E}[\text{traza}(\mathbf{A} \boldsymbol{\mu} \mathbf{Y}^\top)] - \mathbb{E}[\text{traza}(\mathbf{A} \boldsymbol{\mu} \boldsymbol{\mu}^\top)] \\ &= \text{traza}\{\mathbb{E}[\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top]\} + \mathbb{E}[\boldsymbol{\mu}^\top \mathbf{A} \mathbf{Y}] + \mathbb{E}[\mathbf{Y}^\top \mathbf{A} \boldsymbol{\mu}] - \mathbb{E}[\boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}] \\ &= \text{traza}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \\ \mathbb{E}(\mathbf{Q}) &= \text{traza}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}. \end{aligned} \tag{2.10}$$

Para obtener la esperanza de $\mathbf{L}$ y $\mathbf{Q}$, y la varianza de $\mathbf{L}$, solo se necesita conocer $\mathbb{E}(\mathbf{Y})$ y $\mathbb{V}(\mathbf{Y})$, sin embargo para encontrar $\mathbb{V}(\mathbf{Q})$, se va a asumir que además

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V}).$$

⁴ $\text{traza}(\mathbf{ABC}) = \text{traza}(\mathbf{CAB}) = \text{traza}(\mathbf{BCA})$ y $\text{traza}(\mathbf{A} + \mathbf{B}) = \text{traza}(\mathbf{A}) + \text{traza}(\mathbf{B})$.

Algunos resultados de distribuciones de formas cuadráticas se presentan a continuación:

$$\mathbb{V}(\mathbf{Q}) = 2\text{traza}(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}^\top \mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}.$$

Si $\mathbf{Q}_1$ y $\mathbf{Q}_2$ son formas cuadráticas con matrices $\mathbf{A}_1$ y $\mathbf{A}_2$, entonces:

$$\mathbb{C}_{\mathbb{O}\mathbb{V}}(\mathbf{Q}_1, \mathbf{Q}_2) = 2\text{traza}(\mathbf{A}_1\mathbf{V}\mathbf{A}_2\mathbf{V}) + 4\boldsymbol{\mu}^\top \mathbf{A}_1\mathbf{V}\mathbf{A}_2\boldsymbol{\mu}.$$

Si $\mathbf{L} = \mathbf{B}\mathbf{Y}$ es una forma lineal, entonces:

$$\mathbb{C}_{\mathbb{O}\mathbb{V}}(\mathbf{L}, \mathbf{Q}) = 2\mathbf{B}\mathbf{V}\mathbf{A}\boldsymbol{\mu}.$$

Para la distribución de las formas cuadráticas se usará el teorema 2.2, cuya demostración se puede consultar en [34, p. 51].

Teorema 2.2. Si $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ y $\mathbf{Q} = \mathbf{Y}^\top \mathbf{A}\mathbf{Y}$, entonces $\mathbf{Q} \sim \chi^2(r, \lambda)$ con $r = \text{rango}(\mathbf{A})$ y $\lambda = \frac{1}{2}\boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu}$, parámetro de no centralidad, si y solo si $\mathbf{A}\mathbf{V}$ es idempotente.

Ejemplo 2.6. Encuentre la distribución de $\frac{n-1}{\sigma^2} S^2$, con $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$, en que $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, si $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ es un vector aleatorio tal que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ con

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = \mu \mathbf{1}_n \quad y \quad \mathbb{V}(\mathbf{Y}) = \mathbf{V} = \sigma^2 \mathbf{I}_n.$$

Solución. Se tiene que:

$$\begin{aligned} \frac{n-1}{\sigma^2} S^2 &= \frac{1}{\sigma^2} \mathbf{Y}^\top \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{Y} \\ &= \underbrace{\left(\frac{1}{\sigma} \mathbf{Y} \right)^\top}_{\mathbf{Y}^{*\top}} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \underbrace{\left(\frac{1}{\sigma} \mathbf{Y} \right)}_{\mathbf{Y}^*} = \mathbf{Y}^{*\top} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{Y}^*, \end{aligned}$$

en que $\mathbf{Y}^* = \frac{1}{\sigma} \mathbf{Y}$, entonces $\mathbb{E}(\mathbf{Y}^*) = \boldsymbol{\mu}^* = \frac{\mu}{\sigma} \mathbf{1}_n$ y $\mathbb{V}(\mathbf{Y}^*) = \mathbf{V}^* = \frac{\sigma^2}{\sigma^2} \mathbf{I}_n = \mathbf{I}_n$.

Sea $\mathbf{A} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$. Para usar el teorema 2.2, se verifica si $\mathbf{A}\mathbf{V}^*$ es idempotente:

$$\begin{aligned} \mathbf{A}\mathbf{V}^* &= \left[\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \mathbf{I}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \\ \mathbf{A}\mathbf{V}^* \mathbf{A}\mathbf{V}^* &= \left[\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \left[\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \\ &= \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top + \frac{1}{n^2} \mathbf{1}_n \mathbf{1}_n^\top \mathbf{1}_n \mathbf{1}_n^\top \\ &= \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top = \mathbf{A}\mathbf{V}^*, \text{ entonces } \mathbf{A}\mathbf{V}^* \text{ es idempotente.} \end{aligned}$$

Del teorema 2.2 se tiene que:

$$\mathbf{Y}^{*\top} \mathbf{A}\mathbf{Y}^* = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(r, \lambda),$$

en que

$$\begin{aligned} r = \text{rango}(\mathbf{A}) &= \text{rango} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) = \underbrace{\text{traza} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)}_{\text{porque es idempotente}} \\ &= n - \frac{n}{n} = n - 1. \\ \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* = \frac{1}{2} \frac{\mu^2}{\sigma^2} \mathbf{1}_n^\top \left[\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \mathbf{1}_n \\ &= \frac{\mu^2}{2\sigma^2} \left[\mathbf{1}_n^\top \mathbf{I}_n \mathbf{1}_n - \frac{1}{n} \mathbf{1}_n^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{1}_n \right] = 0. \end{aligned}$$

Por tanto

$$\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2.$$

✓

Para determinar la independencia entre dos formas cuadráticas se usará el teorema 2.3, y para determinar la distribución de particiones de sumas de cuadrados es útil el teorema 2.4. Las demostraciones se pueden consultar en [34, pp. 52-53].

Teorema 2.3. Si $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$, $Q_1 = \mathbf{Y}^\top \mathbf{A}_1 \mathbf{Y}$ y $Q_2 = \mathbf{Y}^\top \mathbf{A}_2 \mathbf{Y}$, entonces Q_1 y Q_2 son independientes si y solo si $\mathbf{A}_1 \mathbf{V} \mathbf{A}_2 = \mathbf{0}$.

Teorema 2.4 (Teorema de Cochran). Sea $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ y $\mathbf{A}_i, i = 1, \dots, k$, matrices simétricas con $\text{rango}(\mathbf{A}_i) = r_i$. Sea:

$$\mathbf{A} = \sum_{i=1}^k \mathbf{A}_i,$$

con $r = \text{rango}(\mathbf{A})$. Si $\mathbf{A}\mathbf{V}$ es idempotente y

$$r = \sum_{i=1}^k r_i,$$

entonces las formas cuadráticas $Q_i = \mathbf{Y}^\top \mathbf{A}_i \mathbf{Y}, i = 1, \dots, k$, son mutuamente independientes, variables aleatorias chi-cuadrado no centrales con

$$Q_i \sim \chi^2 \left(r_i, \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A}_i \boldsymbol{\mu} \right).$$

DEFINICIÓN 2.5. Si $Y \sim N(\delta, 1)$ y $X \sim \chi_d^2$ son variables aleatorias independientes, la razón:

$$t = \frac{Y}{\sqrt{\frac{X}{d}}}$$

se llama una t -estadística no central con d grados de libertad y parámetro de no centralidad δ . Se denota por $t(d, \delta)$ (ver [34, p. 57]).

DEFINICIÓN 2.6. Sea $X_1 \sim \chi^2(d_1, \lambda)$ y $X_2 \sim \chi^2(d_2, 0)$, variables aleatorias independientes. La razón:

$$F = \frac{X_1/d_1}{X_2/d_2}$$

es llamada una F -estadística no central con d_1 y d_2 grados de libertad y parámetro de no centralidad λ . Se denota $F(d_1, d_2, \lambda)$ (ver [34, p. 58]).

7. Ejercicios

7.1 Para la ecuación $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$ en que

$$\mathbf{y} = \begin{pmatrix} 10 \\ 5 \\ 12 \\ 3 \\ 4 \\ 3 \\ 8 \\ 7 \\ 6 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{y} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{pmatrix}.$$

- Encontrar al menos dos soluciones de $\hat{\boldsymbol{\beta}}$.
- Encontrar la estimación de $\tau_1 - \tau_3$ usando las dos estimaciones de $\hat{\boldsymbol{\beta}}$. ¿ $\tau_1 - \tau_3$ es estimable?

7.2 Considere un experimento, en que los resultados obtenidos son [40, p. 124]:

α_1	α_2	α_3
4	3	10
6	5	15
9	2	11

y considere el modelo $Y_{ij} = \mu + \alpha_i + e_{ij}$.

- Escriba el modelo en forma matricial.
- Verifique cuáles, entre el siguiente conjunto de funciones paramétricas, son estimables.

- 1) $\boldsymbol{\lambda}^t \boldsymbol{\beta} = \alpha_1 - \alpha_2$
- 2) $\boldsymbol{\lambda}^t \boldsymbol{\beta} = \mu + 2\alpha_1 - \alpha_2$
- 3) $\boldsymbol{\lambda}^t \boldsymbol{\beta} = 2\alpha_1 - \alpha_2 - \alpha_3$

- c) Para cada una de las funciones paramétricas estimables que encontró en (b), encontrar el MELI, su varianza y su suma de cuadrados.

7.3 Demostrar el teorema 2.1.

7.4 Considere la matriz de proyección $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$.

- Demostrar que $\mathbf{P}_x$ es simétrica e idempotente.
- Demostrar que $\text{rango}(\mathbf{P}_x) = \text{rango}(\mathbf{X})$.
- Demostrar que

$$\text{rango}(\mathbf{I} - \mathbf{P}_x) = n - \text{rango}(\mathbf{P}_x) = n - \text{rango}(\mathbf{X}).$$

7.5 Suponga que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_3)$, y sea $q_i = \mathbf{Y}^\top \mathbf{A}_i \mathbf{Y}$, $i = 1, 2$, $n = 3$ en que $\mathbf{A}_1 = \frac{1}{3} \mathbf{1} \mathbf{1}^\top$ con $\mathbf{1} = (1, 1, 1)^\top$, y

$$\mathbf{A}_2 = \frac{1}{6} \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{pmatrix}.$$

- Encuentre la distribución de cada q_i . Justifique adecuadamente su respuesta.
- ¿Son q_1 y q_2 independientes? Justifique adecuadamente su respuesta.

7.6 Dada la definición de la distribución t y la distribución χ^2 [56, p. 72], muestre que

$$\frac{\bar{Y} - \mu}{1/\sqrt{n}} \sqrt{\frac{n-1}{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \sim t_{n-1},$$

cuando $\mathbf{Y} \sim N(\mu \mathbf{1}, \sigma^2 \mathbf{I})$.

7.7 Sea el modelo $Y_{ij} = \mu + i\alpha + e_{ij}$, con $i = 1, \dots, a$ y $j = 1, \dots, r$.

- Encontrar $\begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \end{pmatrix}$ vía mínimos cuadrados ordinarios en términos de Y_{**} y $\bar{Y}_{i*}$.
- Entontrar $SC(\hat{\mu})$.



Capítulo 3

Diseño

completamente

aleatorizado

En este capítulo se presenta el diseño completamente aleatorizado (DCA) que, además de ser el más simple de los diseños experimentales, da una base para diseños más complejos. La primera sección (1) considera el tema de aleatorización de los tratamientos a las unidades experimentales o la aleatorización de las unidades experimentales a los tratamientos, que es uno de los tres aspectos importantes en el diseño de experimentos (ver la sección 3). Esta aleatorización se presenta desde el punto de vista conceptual.

En la sección 2 se presenta, como tal, el diseño completamente aleatorizado cuando el interés es sobre efectos fijos. Las secciones 3 y 4 presentan variantes que se pueden tener del diseño como presencia de submuestreo, o que, en vez de ser un modelo de efectos fijos, sea un modelo de efectos aleatorios. La sección 5 está dedicada a algunas indicaciones para la verificación de supuestos del modelo. Las secciones 6 y 7 tratan el tema de comparación de efectos de tratamientos para el caso en que la medida de escala de estos es nominal u ordinal, o el ajuste de curvas usando polinomios ortogonales cuando la escala de medida de los tratamientos es de razón, respectivamente. En la sección 8 se presenta el análisis de varianza cuando hay una variable dependiente regresora y en la sección 9 se muestra una alternativa de análisis no paramétrica.

1. Aleatorización

Antes de presentar el concepto de aleatorización, se recuerda que la diferencia entre estudios observacionales y estudios de intervención radica que en los segundos se ejerce control (hay influencia al aplicar un tratamiento) por parte del experimentador en las unidades experimentales.

Por otro lado, los datos históricos ayudan a conocer/establecer una distribución de probabilidad. Sin embargo, cuando no hay datos históricos, se puede pensar en una población de repeticiones de la cual se toman muestras que ayudarán a establecer la distribución de probabilidad [33, p. 139].

Ejemplo 3.1. Se consideran **las respuestas potenciales o conceptuales**, y_{ij} , provenientes de 6 individuos y dos tratamientos (ver tabla 3.1).

Tabla 3.1. Tabla de respuestas potenciales o conceptuales para 6 individuos y dos tratamientos

Tratamiento	Unidad						Total	Media
	1	2	3	4	5	6		
Trat. 1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	$\sum_{j=1}^6 y_{1j}$	$\bar{y}_{1*}$
Trat. 2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}	$\sum_{j=1}^6 y_{2j}$	$\bar{y}_{2*}$

Fuente: [33, p. 145].

El interés está en estimar $\bar{y}_{1*} - \bar{y}_{2*}$, sin embargo, como solo se puede observar la respuesta del individuo bajo uno de los dos tratamientos, para la estimación se seleccionan, por ejemplo, las respuestas, y_{ij} , de 3 individuos de cada tratamiento, esto es, $N = 6, n = 3$. Se sabe que:

$$\mathbb{E}(\bar{Y}_{1*}) = \bar{y}_{1*} \quad y \quad \mathbb{E}(\bar{Y}_{2*}) = \bar{y}_{2*},$$

con $\bar{Y}_{1*} = \frac{1}{3} \sum_{j \in s_1} y_{1j}$ y $\bar{Y}_{2*} = \frac{1}{3} \sum_{j \in s_2} y_{2j}$, en que s_1 denota el conjunto de los 3 individuos seleccionados bajo el tratamiento 1 y s_2 denota el conjunto de los 3 individuos seleccionados bajo el tratamiento 2. $\square$

Nota 3.1. La varianza de la media muestral bajo población finita de tamaño N y muestreo aleatorio simple sin reemplazo es:

$$\mathbb{V}(\bar{X}) = \frac{1-f}{n} S_{xU}^2 = \frac{1-f}{n} \sum_{i=1}^N \frac{(x_i - \bar{X})^2}{N-1},$$

con $f = n/N$ [54, p. 68].

Adicionalmente, se debe tener en cuenta que las unidades experimentales pueden ser asignadas tanto al tratamiento 1 como al tratamiento 2, pero no a los dos tratamientos al tiempo, por lo tanto

$$\mathbb{V}(\bar{Y}_{1*} - \bar{Y}_{2*}) \neq \mathbb{V}(\bar{Y}_{1*}) + \mathbb{V}(\bar{Y}_{2*}).$$

Este proceso de aleatorización a los tratamientos se puede formalizar usando la variable indicadora [33, p. 146]:

$$\delta_j = \begin{cases} 1, & \text{si la unidad } j \text{ recibe el tratamiento 1} \\ 0, & \text{en otro caso.} \end{cases}$$

Entre las propiedades de la variable indicadora δ_j están:

$$\mathbb{E}(\delta_j) = P(\delta_j = 1)$$

$$\mathbb{E}(\delta_j^2) = 1^2 P(\delta_j = 1) + 0^2 P(\delta_j = 0) = P(\delta_j = 1)$$

$$\begin{aligned} \mathbb{V}(\delta_j) &= \mathbb{E}(\delta_j^2) - \mathbb{E}(\delta_j)^2 = P(\delta_j = 1) - [P(\delta_j = 1)]^2 \\ &= P(\delta_j = 1)[1 - P(\delta_j = 1)] \end{aligned}$$

$$\mathbb{E}(\delta_j \delta_{j'}) = P(\delta_j = 1)P(\delta_{j'} = 1/\delta_j = 1)$$

$$\begin{aligned} \mathbb{C}_{\odot \mathbb{V}}(\delta_j, \delta_{j'}) &= \mathbb{E}(\delta_j \delta_{j'}) - \mathbb{E}(\delta_j)\mathbb{E}(\delta_{j'}) \\ &= P(\delta_j = 1)P(\delta_{j'} = 1/\delta_j = 1) - P(\delta_j = 1)P(\delta_{j'} = 1). \end{aligned}$$

Ejemplo 3.2. Retomando el ejemplo 3.1,

1. Presente $\mathbb{E}(\delta_j)$, $\mathbb{V}(\delta_j)$ y $\mathbb{C}_{\odot \mathbb{V}}(\delta_j, \delta_{j'})$.
2. Si se está interesado en estimar $T = \bar{y}_{1*} - \bar{y}_{2*}$, ¿cuál es la $\mathbb{E}(T)$ y la $\mathbb{V}(T)$

Solución

1.

$$\begin{aligned} \mathbb{E}(\delta_j) &= \frac{1}{2}, \quad \mathbb{E}(\delta_j^2) = \frac{1}{2} \\ \mathbb{V}(\delta_j) &= \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4} \\ \mathbb{E}(\delta_j \delta_{j'}) &= \frac{1}{2} \times \frac{2}{5} = \frac{1}{5} \\ \mathbb{C}_{\odot \mathbb{V}}(\delta_j, \delta_{j'}) &= \frac{1}{5} - \left(\frac{1}{2}\right)^2 = \frac{4-5}{20} = \frac{-1}{20}. \end{aligned}$$

2. Como se está interesado en estimar $\bar{y}_{1*} - \bar{y}_{2*}$, el estimador se puede escribir usando la indicadora δ_j , esto es,

$$T = \frac{1}{3} \sum_{j=1}^6 \delta_j y_{1j} - \frac{1}{3} \sum_{j=1}^6 (1 - \delta_j) y_{2j},$$

con

$$\mathbb{E}(T) = \frac{1}{3} \sum_{j=1}^6 \mathbb{E}(\delta_j y_{1j} - y_{2j} + \delta_j y_{2j}) \quad (3.1)$$

$$\begin{aligned} &= \frac{1}{3} \sum_{j=1}^6 \left[\mathbb{E}(\delta_j) (y_{1j} + y_{2j}) - y_{2j} \right] \\ &= \frac{1}{6} \sum_{j=1}^6 [y_{1j} + y_{2j}] - \frac{2}{6} \sum_{j=1}^6 y_{2j} = \bar{y}_{1*} + \bar{y}_{2*} - 2\bar{y}_{2*} \quad (3.2) \\ &= \bar{y}_{1*} - \bar{y}_{2*} \end{aligned}$$

$$\mathbb{V}(T) = \frac{1}{30} \sum_{j=1}^6 \left(y_{*j} - \frac{y_{**}}{6} \right)^2. \quad (3.3)$$

$$\text{con } y_{**} = \sum_{j=1}^6 y_{ij}.$$

✓

2. Diseño a una vía de clasificación

Algunas características del diseño completamente aleatorizado son [40, p. 158]:

- Los tratamientos se ensayan en condiciones homogéneas del material experimental, esto es, las unidades experimentales son homogéneas.
- El único factor a considerar son los tratamientos.
- Las unidades experimentales se asignan aleatoriamente a los tratamientos ó los tratamientos son asignados aleatoriamente a las unidades experimentales.

Nota 3.2. Si n es el número total de unidades experimentales, a el número de tratamientos y r el número de réplicas, esto es, $n = a \times r$, se dice que el **diseño es balanceado**, esto es, es un experimento completamente aleatorizado con igual número de réplicas.

Ventaja del DCA. Flexibilidad porque:

- Se ajusta a cualquier número de tratamientos.
- Puede tener igual o desigual número de réplicas.

Desventaja del DCA. Exige homogeneidad. La homogeneidad es difícil de conseguir cuando hay muchos tratamientos o muchas réplicas.

Nota 3.3. Si los **tratamientos son preestablecidos** y las conclusiones son solo sobre los tratamientos aplicados, la **inferencia es con respecto a efectos medios de tratamientos** sobre una variable respuesta.

Ejemplo 3.3. [20] realizaron un experimento en la Universidad Nacional de Colombia sobre el efecto de algunos fertilizantes en la producción de semilla de colza, en el que se midió, entre otros elementos, la altura de la planta (en metros). Se investigaron 12 fertilizantes, sin embargo para efectos **ilustrativos** solo se van a considerar 6 (ver tabla 3.2). Las alturas de las plantas se muestran en la tabla 3.3 y un gráfico de puntos por fertilizante en la figura 3.1.

Tabla 3.2. Tratamientos considerados para el ejemplo de los fertilizantes

Tratamiento	Fertilizante (kg/ha)	
	Nivel de Nitrógeno (N)	P_2O_5
1	0	0
2	75	0
3	150	0
4	0	40
5	75	40
6	150	40

En este ejemplo, el número de niveles del factor fertilizante es 6, lo que coincide con el número de tratamientos. En cuanto al número de réplicas, como todos los tratamientos tienen el mismo número de unidades experimentales, es un diseño balanceado con 3 réplicas. El interés está en determinar si el fertilizante tiene un efecto estadísticamente

Tabla 3.3. Altura de la planta de colza (*Brassica sp.*) bajo algunos niveles de fertilizante

Fertilizante					
$N = 0$ $P_2O_5 = 0$	$N = 75$ $P_2O_5 = 0$	$N = 150$ $P_2O_5 = 0$	$N = 0$ $P_2O_5 = 40$	$N = 75$ $P_2O_5 = 40$	$N = 150$ $P_2O_5 = 40$
1.84	2.00	2.08	1.77	1.91	2.10
1.78	1.90	2.11	1.81	1.88	1.90
1.84	1.97	2.08	1.70	1.97	2.01

N : Nivel de nitrógeno, P_2O_5 : Nivel de fósforo.

Fuente: [20].

significativo sobre la altura de la planta de colza. La estimación de parámetros, el análisis de varianza y las conclusiones se presentan más adelante. ✓

2.1. El modelo

El modelo estadístico es dado por (para deducción del modelo ver anexos 1 y 1.2):

$$Y_{ij} = \mu + \tau_i + e_{ij}, \quad i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, r, \quad (3.4)$$

en que Y_{ij} corresponde a la respuesta de la j -ésima unidad experimental asignada al i -ésimo tratamiento y los e_{ij} se asume que son variables aleatorias independientes e idénticamente distribuidas con media cero y varianza σ_e^2 . Para dar una interpretación a los otros parámetros es necesario especificar una restricción. Las dos restricciones más usadas son:

1. La suma de los τ es igual a cero, esto es, $\sum_{i=1}^a \tau_i = 0$. En este caso:
 - μ es la media general.
 - τ_i es el efecto del tratamiento i con respecto a la media general, μ .
2. Uno de los τ es cero, por ejemplo, $\tau_1 = 0$. En este caso:
 - μ es la media del tratamiento 1.

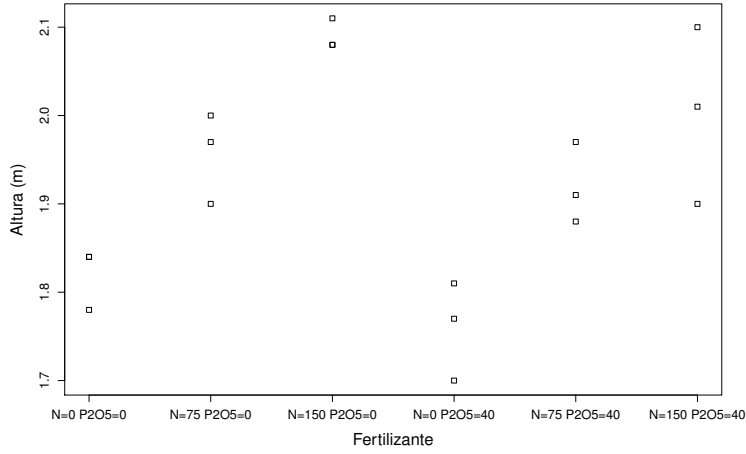


Figura 3.1. Gráfico de puntos para el ejemplo sobre la altura de las plantas de colza

- $\tau_i, i = 2, \dots, a$, es el efecto del tratamiento i con respecto al efecto del tratamiento 1.

Nota 3.4. Es usual asumir este tipo de restricción si entre los tratamientos hay un tratamiento control, que sería el asignado a τ_1 .

Matricialmente

Considerando la primera restricción, esto es,

$$\tau_1 + \tau_2 + \tau_3 + \dots + \tau_{a-1} + \tau_a = 0,$$

entonces

$$\tau_a = -\tau_1 - \tau_2 - \tau_3 - \dots - \tau_{a-1}.$$

Así, en el vector de parámetros β no es necesario incluir a τ_a puesto que τ_a se puede escribir y obtener en términos de los otros τ ,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E},$$

con

$$\boldsymbol{\beta}_{a \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \dots \\ \tau_{a-1} \end{pmatrix}$$

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1r} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2r} \\ \vdots \\ Y_{a1} \\ Y_{a2} \\ \vdots \\ Y_{ar} \end{pmatrix}; \quad \mathbf{X}_{n \times a} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \end{pmatrix}; \quad \mathbf{E}_{n \times 1} = \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1r} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2r} \\ \vdots \\ e_{a1} \\ e_{a2} \\ \vdots \\ e_{ar} \end{pmatrix}.$$

El estimador vía mínimos cuadrados de $\boldsymbol{\beta}$, teniendo en cuenta la restricción de la suma, es dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

con

$$\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma_e^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Ejemplo 3.4. Continuando con el ejemplo 3.3, considere la restricción de la suma de los τ igual a cero:

1. Presente el modelo escrito en forma matricial.
2. Estime el vector de parámetros $\boldsymbol{\beta}$ e interprete.
3. Estime σ_e^2 y $\mathbb{V}(\boldsymbol{\beta})$.

Solución

1. Considerando la restricción de la suma de los τ igual a cero, el modelo es:

$$\begin{aligned}
 \mathbf{Y} &= \mathbf{X} \boldsymbol{\beta} + \mathbf{E} \\
 \begin{pmatrix} 1.84 \\ 1.78 \\ 1.84 \\ 2.00 \\ 1.90 \\ 1.97 \\ 2.08 \\ 2.11 \\ 2.08 \\ 1.77 \\ 1.81 \\ 1.70 \\ 1.91 \\ 1.88 \\ 1.97 \\ 2.10 \\ 1.90 \\ 2.01 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{pmatrix} + \underbrace{\begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{51} \\ e_{52} \\ e_{53} \\ e_{61} \\ e_{62} \\ e_{63} \end{pmatrix}}_{\substack{\text{una} \\ \text{realización} \\ \text{de } \mathbf{E}}}
 \end{aligned}$$

2. Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 3 & 3 & 3 & 3 \\ 0 & 3 & 6 & 3 & 3 & 3 \\ 0 & 3 & 3 & 6 & 3 & 3 \\ 0 & 3 & 3 & 3 & 6 & 3 \\ 0 & 3 & 3 & 3 & 3 & 6 \end{pmatrix}; \quad \mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 34.65 \\ -0.55 \\ -0.14 \\ 0.26 \\ -0.73 \\ -0.25 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.06 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.28 & -0.06 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & 0.28 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & 0.28 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & 0.28 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & -0.06 & 0.28 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} 1.925 \\ -0.105 \\ 0.032 \\ 0.165 \\ -0.165 \\ -0.005 \end{pmatrix}$$

De $\hat{\beta}$ se puede concluir que:

- $\hat{\mu} = 1.925$, esto es, la altura media de las plantas de colza en la fracción del experimento que se está considerando fue de 1.925 metros.
- $\hat{\tau}_1 = -0.105$, esto es, el efecto medio del fertilizante $N = 0$, $P_2O_5 = 0$ sobre la altura de la planta de colza es -0.105 metros con respecto a la media general. Así, el fertilizante $N = 0$, $P_2O_5 = 0$ reduce en promedio la altura en 10.5 centímetros.
- $\hat{\tau}_2 = 0.032$, esto es, el efecto medio del fertilizante $N = 75$, $P_2O_5 = 0$ sobre la altura de la planta de colza es 0.032 metros con respecto a la media general. Así, el fertilizante $N = 75$, $P_2O_5 = 0$ aumenta en promedio la altura en 3.2 centímetros.
- $\hat{\tau}_3 = 0.165$, esto es, el efecto medio del fertilizante $N = 150$, $P_2O_5 = 0$ sobre la altura de la planta de colza es 0.165 metros con respecto a la media general. Así, el fertilizante con $N = 150$ y $P_2O_5 = 0$ aumenta en promedio la altura en 16.5 centímetros.
- $\hat{\tau}_4 = -0.165$, esto es, el efecto medio del fertilizante $N = 0$, $P_2O_5 = 40$ sobre la altura de la planta de colza es -0.165 metros con respecto a la media general. Así, el fertilizante con $N = 0$ y $P_2O_5 = 40$ reduce en promedio la altura en 16.5 centímetros.

- $\hat{\tau}_5 = -0.005$, esto es, el efecto medio del fertilizante $N = 75$, $P_2O_5 = 40$ sobre la altura de la planta de colza es -0.005 metros con respecto a la media general. Así, el fertilizante $N = 75$, $P_2O_5 = 40$ reduce en promedio la altura en 0.5 centímetros.
- $\hat{\tau}_6 = -\hat{\tau}_1 - \hat{\tau}_2 - \hat{\tau}_3 - \hat{\tau}_4 - \hat{\tau}_5 = 0.105 - 0.032 - 0.165 + 0.165 + 0.005 = 0.078$, esto es, el efecto medio del fertilizante $N = 150$, $P_2O_5 = 40$ sobre la altura de la planta de colza es 0.078 metros con respecto a la media general. Así, el fertilizante $N = 150$, $P_2O_5 = 40$ aumenta en promedio la altura en 7.8 centímetros.

3. La varianza del vector de parámetros estimado sería:

$$\begin{aligned} \mathbb{V}(\hat{\beta}) &= \sigma_e^2 (\mathbf{X}^\top \mathbf{X})^{-1} \\ &= \sigma_e^2 \begin{pmatrix} 0.06 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.28 & -0.06 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & 0.28 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & 0.28 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & 0.28 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & -0.06 & 0.28 \end{pmatrix} \end{aligned}$$

Para la estimación de σ_e^2 se puede usar, ver (B.13), que:

$$\hat{\sigma}_e^2 = CME = \frac{\mathbf{y}^\top \mathbf{y} - \hat{\beta}^\top \mathbf{X}^\top \mathbf{y}}{n - a} = \frac{\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}}{n - a}, \quad (3.5)$$

en que:

$$\mathbf{y}^\top \mathbf{y} = 66.9579; \quad \hat{\beta}^\top \mathbf{X}^\top \mathbf{y} = 66.91917 \quad \text{y} \quad n - a = 18 - 6 = 12,$$

entonces $\hat{\sigma}_e^2 = 0.00323$ y una estimación de la varianza del vector de parámetros estimado es:

$$\begin{aligned}\widehat{\mathbb{V}}(\widehat{\beta}) &= 0.00323 \begin{pmatrix} 0.06 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.28 & -0.06 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & 0.28 & -0.06 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & 0.28 & -0.06 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & 0.28 & -0.06 \\ 0 & -0.06 & -0.06 & -0.06 & -0.06 & 0.28 \end{pmatrix} \\ &= \begin{pmatrix} 0.00018 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.00090 & -0.00018 & -0.00018 & -0.00018 & -0.00018 \\ 0 & -0.00018 & 0.00090 & -0.00018 & -0.00018 & -0.00018 \\ 0 & -0.00018 & -0.00018 & 0.00090 & -0.00018 & -0.00018 \\ 0 & -0.00018 & -0.00018 & -0.00018 & 0.00090 & -0.00018 \\ 0 & -0.00018 & -0.00018 & -0.00018 & -0.00018 & 0.00090 \end{pmatrix}.\end{aligned}$$

Por tanto, $\widehat{\mathbb{V}}(\widehat{\mu}) = 0.00018$; $\widehat{\mathbb{V}}(\widehat{\tau}_1) = \widehat{\mathbb{V}}(\widehat{\tau}_2) = \dots = \widehat{\mathbb{V}}(\widehat{\tau}_6) = 0.00090$. $\checkmark$

Ejemplo 3.5. Continuando con el ejemplo 3.3, considere la restricción de uno de los τ igual a cero,

1. Presente el modelo escrito en forma matricial.
2. Estime el vector de parámetros β e interprete.
3. Estime σ_e^2 y $\mathbb{V}(\beta)$.

Solución

1. Considerando la restricción de uno de los τ igual a cero:

$$\begin{aligned}
 \mathbf{Y} &= \mathbf{X} \boldsymbol{\beta} + \mathbf{E} \\
 \begin{pmatrix} 1.84 \\ 1.78 \\ 1.84 \\ 2.00 \\ 1.90 \\ 1.97 \\ 2.08 \\ 2.11 \\ 2.08 \\ 1.77 \\ 1.81 \\ 1.70 \\ 1.91 \\ 1.88 \\ 1.97 \\ 2.10 \\ 1.90 \\ 2.01 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{pmatrix} + \underbrace{\begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{51} \\ e_{52} \\ e_{53} \\ e_{61} \\ e_{62} \\ e_{63} \end{pmatrix}}_{\text{una realización de } \mathbf{E}}
 \end{aligned}$$

2. Para calcular el vector de parámetros estimado:

$$\begin{aligned}
 \mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} 18 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 0 & 0 & 0 & 0 \\ 3 & 0 & 3 & 0 & 0 & 0 \\ 3 & 0 & 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 0 & 3 & 0 \\ 3 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}; \quad \mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 34.65 \\ 5.87 \\ 6.27 \\ 5.28 \\ 5.76 \\ 6.01 \end{pmatrix} \\
 (\mathbf{X}^\top \mathbf{X})^{-1} &= \begin{pmatrix} 0.33 & -0.33 & -0.33 & -0.33 & -0.33 & -0.33 \\ -0.33 & 0.67 & 0.33 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.67 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.67 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.67 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.67 \end{pmatrix}
 \end{aligned}$$

$$\hat{\beta} = \begin{pmatrix} 1.820 \\ 0.137 \\ 0.270 \\ -0.060 \\ 0.100 \\ 0.183 \end{pmatrix}$$

De $\hat{\beta}$ se puede concluir que:

- $\hat{\mu} = 1.82$, esto es, la altura media de las plantas de colza bajo el fertilizante $N = 0, P_2O_5 = 0$ (fertilizante control) fue de 1.82 metros.
- $\hat{\tau}_2 = 0.137$, esto es, el efecto medio del fertilizante $N = 75, P_2O_5 = 0$ sobre la altura de la planta de colza es 0.137 metros con respecto a la altura media bajo el fertilizante control. Así, el fertilizante $N = 75, P_2O_5 = 0$ aumenta en promedio la altura de la planta en 13.7 centímetros con respecto al fertilizante control.
- $\hat{\tau}_3 = 0.27$, esto es, el efecto medio del fertilizante $N = 150, P_2O_5 = 0$ sobre la altura de la planta de colza es 0.27 metros con respecto a la altura media bajo el fertilizante control. Así, el fertilizante $N = 150, P_2O_5 = 0$ aumenta en promedio la altura de la planta en 27 centímetros con respecto al fertilizante control.
- $\hat{\tau}_4 = -0.06$, esto es, el efecto medio del fertilizante con $N = 0$ y $P_2O_5 = 40$ sobre la altura de la planta de colza es -0.06 metros con respecto a la altura media bajo el fertilizante control. Así, el fertilizante $N = 0, P_2O_5 = 40$ disminuye en promedio la altura de la planta en 6 centímetros con respecto al fertilizante control.
- $\hat{\tau}_5 = 0.10$, esto es, el efecto medio del fertilizante con $N = 75$ y $P_2O_5 = 40$ sobre la altura de la planta de colza es 0.10 metros con respecto a la altura media bajo el fertilizante control. Así, el fertilizante $N = 75, P_2O_5 = 40$ aumenta en promedio la altura de la planta en 10 centímetros con respecto al fertilizante control.
- $\hat{\tau}_6 = 0.183$, esto es, el efecto medio del fertilizante $N = 150, P_2O_5 = 40$ sobre la altura de la planta de colza es 0.183 metros

con respecto a la altura media bajo el fertilizante control. Así, el fertilizante $N = 150$, $P_2O_5 = 40$ aumenta en promedio la altura de la planta en 18.3 centímetros con respecto al fertilizante control.

3. La varianza del vector de parámetros estimado sería:

$$\mathbb{V}(\hat{\beta}) = \sigma_e^2 \begin{pmatrix} 0.33 & -0.33 & -0.33 & -0.33 & -0.33 & -0.33 \\ -0.33 & 0.67 & 0.33 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.67 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.67 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.67 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.67 \end{pmatrix}$$

Para la estimación de σ_e^2 se puede usar (3.5), esto es, $\hat{\sigma}_e^2 = 0.00323$ y una estimación de la varianza del vector de parámetros estimados es:

$$\begin{aligned} \hat{\mathbb{V}}(\hat{\beta}) &= 0.00323 \begin{pmatrix} 0.33 & -0.33 & -0.33 & -0.33 & -0.33 & -0.33 \\ -0.33 & 0.67 & 0.33 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.67 & 0.33 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.67 & 0.33 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.67 & 0.33 \\ -0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.67 \end{pmatrix} \\ &= \begin{pmatrix} 0.0011 & -0.0011 & -0.0011 & -0.0011 & -0.0011 & -0.0011 \\ -0.0011 & 0.0022 & 0.0011 & 0.0011 & 0.0011 & 0.0011 \\ -0.0011 & 0.0011 & 0.0022 & 0.0011 & 0.0011 & 0.0011 \\ -0.0011 & 0.0011 & 0.0011 & 0.0022 & 0.0011 & 0.0011 \\ -0.0011 & 0.0011 & 0.0011 & 0.0011 & 0.0022 & 0.0011 \\ -0.0011 & 0.0011 & 0.0011 & 0.0011 & 0.0011 & 0.0022 \end{pmatrix}. \end{aligned}$$

Por tanto,

$$\hat{\mathbb{V}}(\hat{\mu}) = 0.0011; \quad \hat{\mathbb{V}}(\hat{\tau}_1) = \hat{\mathbb{V}}(\hat{\tau}_2) = \dots = \hat{\mathbb{V}}(\hat{\tau}_6) = 0.0022.$$

□

2.2. Intervalos de confianza para la media muestral de un tratamiento e intervalos de confianza para la diferencia de medias entre dos tratamientos

Partiendo del modelo (3.4):

$$\begin{aligned} Y_{ij} &= \mu + \tau_i + e_{ij}, \quad i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, r, \\ \mathbb{E}(Y_{ij}) &= \mathbb{E}(\mu + \tau_i + e_{ij}) = \mu + \tau_i \\ \mathbb{V}(Y_{ij}) &= \mathbb{V}(\mu + \tau_i + e_{ij}) = \mathbb{V}(e_{ij}) = \sigma_e^2, \end{aligned}$$

si la media muestral del i -ésimo tratamiento se denota por $\bar{Y}_{i*}$,

$$\bar{Y}_{i*} = \frac{1}{r} \sum_{j=1}^r Y_{ij}, \quad (3.6)$$

entonces:

$$\mathbb{E}(\bar{Y}_{i*}) = \mathbb{E} \left[\frac{1}{r} \sum_{j=1}^r Y_{ij} \right] = \frac{1}{r} \sum_{j=1}^r \mathbb{E}(Y_{ij}) = \frac{1}{r} \sum_{j=1}^r (\mu + \tau_i) = \mu + \tau_i$$

y

$$\mathbb{V}(\bar{Y}_{i*}) = \mathbb{V} \left[\frac{1}{r} \sum_{j=1}^r Y_{ij} \right] \stackrel{\text{1}}{=} \frac{1}{r^2} \sum_{j=1}^r \mathbb{V}(Y_{ij}) = \frac{1}{r^2} \sum_{j=1}^r \sigma_e^2 = \frac{\sigma_e^2}{r}.$$

Si además se asume que los errores, e_{ij} , provienen de una distribución normal, se tiene que:

$$\bar{Y}_{i*} \sim N \left(\mu + \tau_i, \frac{\sigma_e^2}{r} \right).$$

Como usualmente no se conoce σ_e^2 , esta se puede estimar usando (3.5), ver (B.13).

¹Asumiendo independencia entre los Y_{ij} .

Con base en la distribución de las medias de tratamiento, $\bar{Y}_{i*}$, y la estimación de σ_e^2 , se puede construir un intervalo de confianza de $(1 - \alpha) \times 100 \%$ para la media del i -ésimo tratamiento [41, p. 74]:

$$\left[\bar{Y}_{i*} - t_{1-\frac{\alpha}{2}, n-a} \sqrt{\frac{CME}{r}}; \quad \bar{Y}_{i*} + t_{1-\frac{\alpha}{2}, n-a} \sqrt{\frac{CME}{r}} \right]. \quad (3.7)$$

En cuanto al intervalo de confianza para las diferencias de las medias de tratamientos, $\bar{Y}_{i*} - \bar{Y}_{i' *}$ con $i \neq i'$, se tiene:

$$\mathbb{E} (\bar{Y}_{i*} - \bar{Y}_{i' *}) = \mathbb{E} (\bar{Y}_{i*}) - \mathbb{E} (\bar{Y}_{i' *}) = \mu + \tau_i - (\mu + \tau_{i'}) = \tau_i - \tau_{i'}$$

y

$$\mathbb{V} (\bar{Y}_{i*} - \bar{Y}_{i' *}) = \mathbb{V} (\bar{Y}_{i*}) + \mathbb{V} (\bar{Y}_{i' *}) = \frac{\sigma_e^2}{r} + \frac{\sigma_e^2}{r} = \frac{2\sigma_e^2}{r},$$

por tanto

$$\bar{Y}_{i*} - \bar{Y}_{i' *} \sim N \left(\tau_i - \tau_{i'}, \frac{2\sigma_e^2}{r} \right).$$

A partir de la distribución de $\bar{Y}_{i*} - \bar{Y}_{i' *}$ se puede establecer un intervalo de confianza de $(1 - \alpha) \times 100 \%$ para la diferencia entre la media del i -ésimo tratamiento y la media del i' -ésimo tratamiento [41, p. 74]:

$$\left[\bar{Y}_{i*} - \bar{Y}_{i' *} - t_{1-\frac{\alpha}{2}, n-a} \sqrt{\frac{2CME}{r}}; \quad \bar{Y}_{i*} - \bar{Y}_{i' *} + t_{1-\frac{\alpha}{2}, n-a} \sqrt{\frac{2CME}{r}} \right]. \quad (3.8)$$

Nota 3.5. Los intervalos de confianza dados en (3.7) y (3.8) garantizan el $(1 - \alpha) \times 100 \%$ uno a la vez, pero no con varios intervalos al tiempo. Para controlar el nivel de confianza de varios intervalos al tiempo, se sugiere usar en vez de α , usar $\frac{\alpha}{k}$, con k el número de intervalos simultáneos. Esta elección de α garantiza que la probabilidad de que los k intervalos sean correctos simultáneamente sea al menos de $(1 - \alpha) \times 100 \%$.

Ejemplo 3.6. Continuando con el ejemplo 3.3, presente el intervalo de confianza al 95 % para la altura media de la planta de colza para cada uno de los tratamientos.

Solución. Teniendo en cuenta que $n = 18$ y $a = 6$, entonces $n - a = 18 - 6 = 12$ y $(t_{0.975;12} = 2.179)$,

Diseño completamente aleatorizado

- $N = 0, P_2O_5 = 0$: $\left[1.82 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [1.75 ; 1.89]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 0 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ está entre 1.75 y 1.89 metros con un nivel de confianza del 95 %.

- $N = 75, P_2O_5 = 0$: $\left[1.96 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [1.89 ; 2.03]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 75 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ está entre 1.89 y 2.03 metros con un nivel de confianza del 95 %.

- $N = 150, P_2O_5 = 0$: $\left[2.09 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [2.02 ; 2.16]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ está entre 2.02 y 2.16 metros con un nivel de confianza del 95 %.

- $N = 0, P_2O_5 = 40$: $\left[1.76 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [1.69 ; 1.83]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 0 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$ está entre 1.69 y 1.83 metros con un nivel de confianza del 95 %.

- $N = 75, P_2O_5 = 40$: $\left[1.92 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [1.85 ; 1.99]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 75 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$ está entre 1.85 y 1.99 metros con un nivel de confianza del 95 %.

- $N = 150, P_2O_5 = 40$: $\left[2.00 \mp t_{0.975;12} \sqrt{\frac{0.00323}{3}} \right] = [1.93 ; 2.07]$

La altura de la planta de colza cuando se considera el fertilizante con $N = 150 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$ está entre 1.93 y 2.07 metros con un nivel de confianza del 95 %.



Ejemplo 3.7. Continuando con el ejemplo 3.3, encuentre los intervalos de confianza al 95 % para la diferencia media entre el fertilizante control ($N = 0 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$) y el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$, y para la diferencia media entre el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ y el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$.

Solución. Como son dos intervalos, se usa $\alpha/2 = 0.025$, por tanto $t_{0.9875;12} = 2.56$. Los dos intervalos son:

- Para el fertilizante $N = 0 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ y el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$:

$$\left[2.09 - 1.82 \mp t_{0.9875;12} \sqrt{\frac{2 \times 0.00323}{3}} \right] = [0.15 ; 0.39].$$

- Para el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ y el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$:

$$\left[2.09 - 2.00 \mp t_{0.9875;12} \sqrt{\frac{2 \times 0.00323}{3}} \right] = [-0.03 ; 0.21].$$

Se concluye que el fertilizante control y el fertilizante con $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ producen estadísticamente diferentes alturas (en media) de la planta de colza, esto es, el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ da una mayor altura (en promedio) de la planta de colza. Sobre considerar el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 0 \text{ kg/ha}$ y el fertilizante $N = 150 \text{ kg/ha}$ y $P_2O_5 = 40 \text{ kg/ha}$, estos dos producen estadísticamente la misma altura (en media) de la planta de colza. El nivel de confianza para los dos intervalos simultáneamente es de al menos un 95 %. $\checkmark$

Nota 3.6. El error estándar del estimador de un contraste $\sum_{i=1}^a c_i \tau_i$ es dado por

$$e.e \left(\sum_{i=1}^a c_i \bar{Y}_{i*} \right) = \sqrt{\frac{\hat{\sigma}_e^2}{r} \sum_{i=1}^a c_i^2}.$$

2.3. Análisis de varianza

Un procedimiento adecuado para probar la igualdad de las medias de los a tratamientos o probar que los efectos medios de los a tratamientos es cero es usar análisis de varianza, ver tabla 3.4, y anexo 2 para la deducción de la distribución del F-calculado. Las hipótesis a probar son:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_{a-1} = \tau_a = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \tau_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si $P\text{-valor} < \alpha$, en que α es un nivel de significancia preestablecido por el investigador, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio (sobre la respuesta) de los tratamientos en consideración. Es importante resaltar que la validez de las conclusiones depende de si se cumplen o no los supuestos del modelo.

Tabla 3.4. Análisis de varianza para un diseño completamente aleatorizado

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$a - 1$	$\hat{\beta}^\top \mathbf{X}^\top \mathbf{Y} - n\bar{Y}^2$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CME}$	$P(F_{t^*} > F_c)$
Error	$n - a$	$\mathbf{Y}^\top \mathbf{Y} - \hat{\beta}^\top \mathbf{X}^\top \mathbf{Y}$	$\frac{SCE}{n - a}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{Y}^2$			

$$n = \sum_{i=1}^a r_i, \quad t^* = \{a - 1, gle\} \quad \text{y} \quad gle = \text{grados de libertad del error.}$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} Y_{ij}$$

Ejemplo 3.8. Continuando con el ejemplo 3.3, en la tabla 3.5 se presenta el análisis de varianza para los datos de fertilizantes y su efecto en la altura de las plantas de colza.

Si el modelo cumple con los supuestos, se concluye que hay diferencias significativas en la altura de la planta de colza al aplicar diferentes niveles del fertilizante. $\checkmark$

Tabla 3.5. Análisis de Varianza para los datos del ejemplo 3.3

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Fertilizantes	5	0.21792	0.04358	13.50	$P(F_{5;12} > 13.50) = < 0.01$
Error	12	0.03873	0.00323		
Total	17	0.25665			

$$n\bar{y}^2 = 66.70125$$

2.3.1. Instrucciones en R - ejemplo 3.3

```
remove(list=ls())
fertilizante=gl(6,3) # Datos de la variable fertilizante
Altura=c(1.84,1.78,1.84,2,1.9,1.97,2.08,2.11,2.08,1.77,1.81,1.7,
        1.91,1.88,1.97,2.10,1.90,2.01)
# Datos de la variable altura
fertilizante=as.factor(fertilizante)
# Indica que fertilizante es una variable categórica
fertilizante = factor(fertilizante,levels = c(1,2,3,4,5,6),
                      labels = c("N=0 P205=0","N=75 P205=0",
                                "N=150 P205=0","N=0 P205=40",
                                "N=75 P205=40","N=150 P205=40"))
# Asigna los labels a la variable fertilizante
stripchart(Altura~fertilizante,vertical=TRUE,xlab="Fertilizante")
# Hace un gráfico de puntos para fertilizantes
contrasts(fertilizante)=contr.sum(levels(fertilizante))
# Indica que se asuma la primera restricción
# - restricción de la suma
contrasts(fertilizante) # Muestra la asignación de los
# parámetros a los tratamientos
datos=data.frame(fertilizante,Altura) # coloca en una sola
# base de datos las dos variables
ajuste1 = aov(Altura ~ fertilizante, data=datos)
# ajusta el modelo
ajuste1$coefficients # Imprime los coeficientes estimados (beta)
```

Diseño completamente aleatorizado

```
summary(ajuste1) # Imprime el análisis de varianza
## colocando la restricción de  $\tau_1=0$ .
contrasts(fertilizante)=contr.treatment(levels(fertilizante),
    base=1) # Indica que se asuma la restricción  $\tau_1=0$ 
datos=data.frame(fertilizante,Altura)
ajuste2 = aov(Altura ~ fertilizante, data=datos)
    # ajusta el modelo
ajuste2$coefficients # Imprime los coeficientes estimados (beta)
summary(ajuste2) # Imprime el análisis de varianza
```

Nota 3.7. En el caso en que el diseño sea desbalanceado, esto es, el número de réplicas por tratamiento no sea igual para todos los tratamientos, el análisis de varianza se construye de forma similar a la tabla 3.4, teniendo en cuenta el desbalanceo, ver anexo 3.

2.4. Tamaño de muestra

Para determinar un tamaño de muestra mínimo, es necesario conocer:

- El número de tratamientos a aplicar, a .
- La diferencia mínima que se desea detectar entre dos efectos de tratamiento, se denotará como Δ .
- El nivel de significancia deseado, α .
- La potencia deseada, $1 - \beta$.
- Una estimación de la varianza de los errores, $s^2 = \hat{\sigma}_e^2$ (usualmente se toma de un estudio anterior o de un estudio piloto).

Una deducción de la fórmula para tamaño de muestra se presenta en el anexo 4. Basado en la tabla de potencia del test F , ver programa del anexo 4.1, el proceso consiste en [21, p. 64]:

1. Generar la tabla usando el código dado en el anexo 4.1 teniendo en cuenta el número de tratamientos " a " y α .

2. Ubicar $r = 250$ en la primera iteración y sobre esa fila, ubicar la potencia que se desea tener (igual a $1 - \beta$ o mayor) y obtener el respectivo ϕ (nombre de la columna en la tabla). De esta iteración sale un primer r (número de réplicas por tratamiento).
3. Calcular:

$$r = 2 \frac{a \hat{\sigma}_e^2 \phi^2}{\Delta^2}, \quad (3.9)$$

redondeando al entero superior.

4. Ubicar el r encontrado en la tabla generada en el paso 1 y en la fila r ubicar un valor igual o mayor a la potencia requerida, $1 - \beta$, e identificar el respectivo ϕ (parte superior de la tabla).
5. Repetir 3) y 4) hasta que r no cambie o alterne entre dos valores. Seleccionar el mayor valor.

Ejemplo 3.9. Continuando con el ejemplo 3.3, establecer un tamaño de muestra por fertilizante, considerando un nivel de significancia de 0.05, una potencia de 0.90 y que se desea detectar una diferencia de 15 centímetros en la altura. Utilice las alturas de las plantas de colza recolectadas como un estudio piloto, en particular $\hat{\sigma}_e^2 = 0.00323$.

Solución. Se tiene $a = 6$, $\alpha = 0.05$, $1 - \beta = 0.9$, $\Delta = 0.15$ y $\hat{\sigma}_e^2 = 0.00323$.

Las iteraciones serán:

$$r = \frac{2 \times 6 \times 0.00323 \times \phi^2}{0.15^2}$$

r	$v_2 = 6(r - 1)$	ϕ	r
	1000	1.65	$4.69 \approx 5$
5	24	1.85	$5.90 \approx 6$
6	30	1.85	$5.90 \approx 6$

Entonces $r = 6$, luego el total de unidades experimentales necesarias serían $6 \times 6 = 36$ U.E. ✓

Ejemplo 3.10. [53] realizaron un experimento en la Universidad Nacional de Colombia sobre rendimiento de variedades de cebada. En el experimento consideraron 20 variedades, pero para efectos **ilustrativos**, en este ejemplo considere solo 5 variedades. De la tesis

de grado se obtiene como varianza estimada del rendimiento en gramos por parcela de 7.5 m^2 igual a 24 947 gramos. Si se desea detectar una diferencia mínima de 300 gramos, ¿cuál sería el tamaño de muestra? Considere un nivel de significancia de 0.05 y una potencia de 0.80.

Solución Se tiene $a = 5$, $\Delta = 300$, $\hat{\sigma}_e^2 = 24\,947$, $\alpha = 0.05$ y $1 - \beta = 0.80$. Entonces:

$$r = \frac{2 \times 5 \times 24\,947 \times \phi^2}{300^2}.$$

Las iteraciones serán:

r	$v_2 = a(r - 1)$	ϕ	r
	1245	1.55	$6.66 \approx 7$
7	30	1.7	$8.01 \approx 9$
9	40	1.65	$7.55 \approx 8$
8	35	1.65	$7.55 \approx 8$

Así $r = 8$, luego el total de parcelas necesarias para el experimento serían $5 \times 8 = 40$ U.E. 

Las instrucciones en R para el tamaño de muestra relativas a los ejemplos 3.9 y 3.10 son, respectivamente:

```
library(pwr2)
ss.1way(k=6, alpha=0.05, beta=0.10, delta=0.15,
        sigma=sqrt(0.00323), B=100)
ss.1way(k=5, alpha=0.05, beta=0.20, delta=300,
        sigma=sqrt(24947), B=100)
```

3. Submuestreo en un diseño a una vía de clasificación

En el análisis de varianza, el submuestreo se tiene en cuenta cuando las unidades experimentales y las unidades observacionales no son idénticas (no son las mismas).

Ejemplo 3.11. En un estudio sobre el efecto del nitrógeno en el rendimiento de frijol (kg/ha), se consideraron dos tratamientos:

- *Aplicar todo el nitrógeno en el momento de sembrar.*
- *Aplicar el 50 % del nitrógeno en el momento de sembrar y el restante 50 % de nitrógeno, 18 días más tarde.*

Se consideraron tres parcelas por tratamiento y de cada parcela se seleccionaron tres surcos. Los datos se presentan en la tabla 3.6, un boxplot por tratamiento en la figura 3.2, y un gráfico de puntos por parcela en la figura 3.3.

Tabla 3.6. Ejemplo del nitrógeno y el frijol

100 % al inicio			50 % al inicio		
Parcela 1	Parcela 2	Parcela 3	Parcela 4	Parcela 5	Parcela 6
1367	1086	1166	1491	1487	1579
1342	1187	1431	1290	1468	1230
1156	1539	1464	1476	1444	1228

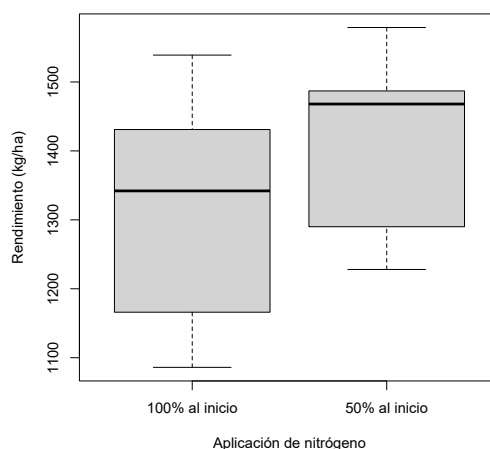


Figura 3.2. Boxplot para el ejemplo sobre rendimiento en frijol

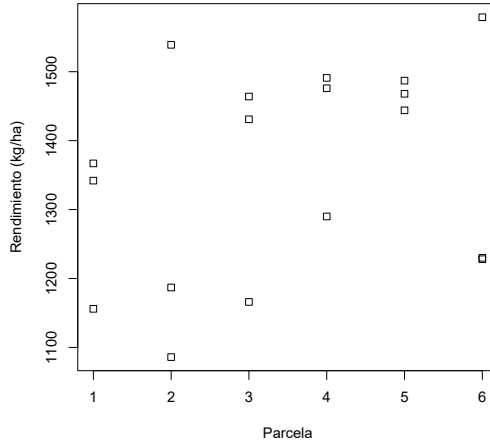


Figura 3.3. Gráfico de puntos para el ejemplo sobre rendimiento en frijol por parcela

3.1. El modelo

Suponga que se tienen a tratamientos, cada uno con r' replicas (unidades experimentales) y cada réplica (UE) tiene n^* unidades observacionales, el modelo es dado por [33, p. 191]:

$$Y_{ijk} = \mu + \tau_i + \varepsilon_{j(i)} + \eta_{k(ij)} \quad (3.10)$$

con

- $i = 1, 2, \dots, a; j = 1, 2, \dots, r' \text{ y } k = 1, 2, \dots, n^*.$
- y_{ijk} es la respuesta de la unidad observacional k en la unidad experimental j en el tratamiento i .
- $\varepsilon_{j(i)}$ es el error experimental tal que $\varepsilon_{j(i)} \sim N(0, \sigma_\varepsilon^2).$
- $\eta_{k(ij)}$ es el error observacional tal que $\eta_{k(ij)} \sim N(0, \sigma_\eta^2).$
- $n = ar'n^*.$

Asumiendo la restricción $\sum_{i=1}^a \tau_i = 0$ en el modelo (3.10),

- μ es la media general.
- τ_i es el efecto del tratamiento i con respecto a la media general, μ .

Matricialmente

Considerando la restricción de la suma para los τ , en el modelo (3.10), se tiene que:

$$\tau_1 + \tau_2 + \tau_3 + \cdots + \tau_{a-1} + \tau_a = 0$$

y despejando, por ejemplo, τ_a en términos de los otros τ ,

$$\tau_a = -\tau_1 - \tau_2 - \tau_3 - \cdots - \tau_{a-1},$$

el modelo (3.10) se puede escribir matricialmente como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta}, \quad (3.11)$$

con

$$\underbrace{\boldsymbol{\beta}}_{a \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{a-1} \end{pmatrix}; \underbrace{\mathbf{Z}}_{n \times ar'} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{pmatrix}; \underbrace{\mathbf{0}}_{r' n^* \times r'} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix};$$

Diseño completamente aleatorizado

$$\underbrace{\mathbf{Y}}_{n \times 1} = \begin{pmatrix} Y_{111} \\ \vdots \\ Y_{11n^*} \\ \vdots \\ Y_{1r'1} \\ \vdots \\ Y_{1r'n^*} \\ Y_{211} \\ \vdots \\ Y_{21n^*} \\ \vdots \\ Y_{a11} \\ \vdots \\ Y_{a1n^*} \\ \vdots \\ Y_{ar'1} \\ \vdots \\ Y_{ar'n^*} \end{pmatrix}; \underbrace{\mathbf{X}}_{n \times a} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \end{pmatrix}; \underbrace{\boldsymbol{\eta}}_{n \times 1} = \begin{pmatrix} \eta_{1(11)} \\ \vdots \\ \eta_{n^*(11)} \\ \vdots \\ \eta_{1(1r')} \\ \vdots \\ \eta_{n^*(1r')} \\ \eta_{1(21)} \\ \vdots \\ \eta_{n^*(21)} \\ \vdots \\ \eta_{1(a1)} \\ \vdots \\ \eta_{n^*(a1)} \\ \vdots \\ \eta_{1(ar')} \\ \vdots \\ \eta_{n^*(ar')} \end{pmatrix};$$

$$\underbrace{\mathbf{A}}_{r'n^* \times r'} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}; \underbrace{\boldsymbol{\varepsilon}}_{ar' \times 1} = \begin{pmatrix} \varepsilon_{1(1)} \\ \varepsilon_{2(1)} \\ \vdots \\ \varepsilon_{r'(1)} \\ \varepsilon_{1(2)} \\ \varepsilon_{2(2)} \\ \vdots \\ \varepsilon_{r'(2)} \\ \vdots \\ \varepsilon_{1(a)} \\ \vdots \\ \varepsilon_{r'(a)} \end{pmatrix};$$

El estimador mínimos cuadrados de $\boldsymbol{\beta}$ es dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Ejemplo 3.12. Continuando con el ejemplo 3.11, presente el modelo en forma matricial, estime β e interprete.

Solución. El respectivo modelo es dado por:

$$\begin{pmatrix} 1367 \\ 1342 \\ 1156 \\ 1086 \\ 1187 \\ 1539 \\ 1166 \\ 1431 \\ 1464 \\ 1491 \\ 1290 \\ 1476 \\ 1487 \\ 1468 \\ 1444 \\ 1579 \\ 1230 \\ 1228 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1(1)} \\ \varepsilon_{2(1)} \\ \varepsilon_{3(1)} \\ \varepsilon_{1(2)} \\ \varepsilon_{2(2)} \\ \varepsilon_{3(2)} \end{pmatrix} + \begin{pmatrix} \eta_{1(11)} \\ \eta_{2(11)} \\ \eta_{3(11)} \\ \eta_{1(12)} \\ \eta_{2(12)} \\ \eta_{3(12)} \\ \eta_{1(31)} \\ \eta_{2(13)} \\ \eta_{3(13)} \\ \eta_{1(21)} \\ \eta_{2(21)} \\ \eta_{3(21)} \\ \eta_{1(22)} \\ \eta_{2(22)} \\ \eta_{3(22)} \\ \eta_{1(23)} \\ \eta_{2(23)} \\ \eta_{3(23)} \end{pmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 18 & 0 \\ 0 & 18 \end{pmatrix}; (\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{18} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{Y} = \begin{pmatrix} 24431 \\ -955 \end{pmatrix}; \hat{\beta} = \begin{pmatrix} 1357.2778 \\ -53.0556 \end{pmatrix}.$$

De $\hat{\beta}$ se puede concluir que:

- $\mu = 1357.3$, esto es, el rendimiento en media es 1357 kg/ha.
- $\tau_1 = -53.2$, esto es, el efecto medio de aplicar el 100 % del nitrógeno en el momento de la siembra sobre el rendimiento es de -53 kg/ha con respecto a la media general. Así, aplicar el 100 % del nitrógeno en el momento de la siembra *disminuye* el rendimiento en 53 kg/ha con respecto al rendimiento medio.

- $\tau_2 = 53.2$, esto es, el efecto medio de aplicar el 50 % del nitrógeno en el momento de la siembra y aplicar el restante 50 % del nitrógeno 18 días más tarde sobre el rendimiento es de 53 kg/ha con respecto a la media general. Por tanto, aplicar el 50 %-50 % del nitrógeno *aumenta* el rendimiento en 53 kg/ha con respecto al rendimiento medio. ✓

3.2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que la variación del error se debe dividir en la variación debida a las unidades experimentales y a la variación debida a las unidades observacionales, ver tabla 3.7. Ver anexo 5 para deducción de la tabla 3.7.

La hipótesis de interés es:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_{a-1} = \tau_a = 0 \quad \text{versus} \\ H_a : \text{Al menos un } \tau_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado a los tratamientos es menor a α en la tabla del análisis de varianza, ver tabla 3.7, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los tratamientos en consideración.

Ejemplo 3.13. Continuando con el ejemplo 3.11, presente el análisis de varianza y concluya.

Solución. Se tiene que:

$$\begin{aligned} \mathbf{y}^\top \mathbf{X}\hat{\boldsymbol{\beta}} &= 33\,210\,321.44 & \mathbf{y}^\top \mathbf{Z}\hat{\boldsymbol{\varepsilon}} &= 33\,243\,969.67 \\ \mathbf{y}^\top \mathbf{y} &= 33\,544\,699 & n\bar{y}^2 &= 33\,159\,653.39 \end{aligned}$$

De la tabla 3.8 se concluye que no hay diferencias significativas entre las dos tipos de aplicación de nitrógeno al frijol con un nivel de significancia del 5 %. ✓

Tabla 3.7. Tabla de análisis de varianza para un diseño completamente aleatorizado en presencia de submuestreo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$a - 1$	$\mathbf{Y}^\top \mathbf{X} \hat{\beta} - n \bar{y}^2$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CMEE}$	$P(F_{a-1, glee} > F_c)$
Error Exper.	$a(r' - 1)$	$\mathbf{Y}^\top \mathbf{Z} \hat{\epsilon} - \mathbf{Y}^\top \mathbf{X} \hat{\beta}$	$\frac{SCEE}{glee}$		
Error Ob-serv.	$ar'(n^* - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{Z} \hat{\epsilon}$	$\frac{SCEO}{gleo}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n \bar{y}^2$			
$n = a * r' * n^*$ y $glee$ =grados de libertad del error experimental. $\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r'} \sum_{k=1}^{n^*} y_{ijk}$ $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad \hat{\epsilon} = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{Y}$					

Nota 3.8. Para hacer las estimaciones de los componentes de varianza σ_ϵ^2 y σ_η^2 , se debe tener en cuenta que, ver (B.27):

$$\hat{\sigma}_\eta^2 = CMEO$$

y

$$\hat{\sigma}_\epsilon^2 = \frac{CMEE - CMEO}{n^*},$$

en que $CMEO$ es el cuadrado medio del error observacional y $CMEE$ es el cuadrado medio del error experimental.

La ventaja de usar submuestreo es poder separar la variación debida a las unidades experimentales y la variación debida a las unidades observacionales. Esto ayuda a observar la calidad de las técnicas de medición y experimentación. Si $\hat{\sigma}_\eta^2$ es muy grande, se debe mejorar el procedimiento de medición y si $\hat{\sigma}_\epsilon^2$ es muy grande, se debe revisar la homogeneidad de las unidades experimentales.

En el ejemplo 3.11, $\hat{\sigma}_\eta^2 = 25\,060.778$, $\hat{\sigma}_\epsilon^2 = \frac{8412.056 - 25\,060.778}{3} < 0$, indicando que la varianza se debe solo a variación dentro de las parcelas y no entre las parcelas.

Tabla 3.8. Análisis de varianza para el ejemplo de los porcentajes de nitrógeno y el frijol

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratamientos	1	50 668.06	50 668.06	6.02	0.05<P-valor<0.10
Error Exper.	4	33 648.22	8412.06		
Error Observ.	12	300 729.33	25 060.78		
Total	17	385 045.61			

3.2.1. Instrucciones en R - ejemplo [3.11](#)

```
remove(list=ls())
porcentaje=c(rep("100% al inicio",9),rep("50% al inicio",9))
# Variable independiente
UE=gl(6,3) # unidades experimentales
rendimiento=c(1367,1342,1156,1086,1187,1539,1166,1431,1464,1491,
1290,1476,1487,1468,1444,1579,1230,1228) # variable dependiente
porcentaje=as.factor(porcentaje) # especifica que 'porcentaje' es
# una variable categorica
UE=as.factor(UE) # especifica que UE es una variable categórica
plot(porcentaje,rendimiento,ylab="Rendimiento (kg/ha)",
xlab="Aplicación de nitrógeno") # Boxplot por tratamiento
stripchart(rendimiento~UE,vertical=TRUE,
ylab="Rendimiento (kg/ha)",xlab="Parcela")
submues=data.frame(porcentaje,UE,rendimiento) # une en un solo
# data.frame las variables independiente y dependientes
ajuste.subm=aov(rendimiento ~ porcentaje+Error(UE),data=submues)
# ajusta el modelo de submuestreo en que el término de error
# corresponde al asociado a las unidades experimentales
summary(ajuste.subm) #imprime en pantalla el análisis de varianza
1-pf(6.023,1,4) # P-valor
```

3.3. Tamaño de muestra en submuestreo

Se pueden relacionar r , r' y n^* usando [33, p. 195]:

$$n^* = \frac{r\delta}{r'(1+\delta) - r},$$

en que

$$\delta = \frac{\sigma_\eta^2}{\sigma_\epsilon^2}.$$

Para calcular un tamaño de muestra mínimo, se selecciona r' cercano a r , por ejemplo $r' = r - 1$, y se usa la relación anterior entre r y r' para encontrar n^* .

Ejemplo 3.14. *Asumir que los datos del ejemplo de los fertilizantes y la altura de las plantas de colza ($a = 6$) son prueba piloto para otro estudio. Calcular un tamaño de muestra adecuado para $\Delta = 0.10$ con $\alpha = 0.05$ y $1 - \beta = 0.80$.*

Solución. Del estudio piloto se sabe que $\hat{\sigma}_\epsilon^2 = 0.00323$ y calculando el tamaño de muestra:

```
library(pwr2)
ss.1way(k=6, alpha=0.05, beta=0.20, delta=0.10,
        sigma=sqrt(0.00323), B=100)
```

se tiene que $r = 10$. Asumiendo diferentes δ , se tienen algunos escenarios con tamaños de muestra en la tabla 3.9. ✓

4. Modelo de efectos aleatorios

Se utiliza este modelo cuando los niveles que se usan en el experimento son una **muestra aleatoria** de una población de niveles del factor en estudio. En este caso no es de interés los efectos medios de tratamientos, sino la varianza del factor aleatorio. La inferencia se realiza sobre la población de niveles del factor.

Ejemplo 3.15. *Ejercicio 12 de [40, p. 225] sobre la concentración de calcio en hojas de una especie de árbol. La variable respuesta es porcentaje de peso seco. Los datos se presentan en la tabla 3.10.*

Tabla 3.9. Algunos escenarios de tamaños de muestra usando submuestreo en el ejemplo de las plantas de colza

$\delta = 0.50$		$\delta = 0.75$		$\delta = 1.00$		$\delta = 1.5$	
r'	n^*	r'	n^*	r'	n^*	r'	n^*
9	2	9	2	9	2	9	2
8	3	8	2	8	2	8	2
7	10	7	4	7	3	7	2
		6	15	6	4	6	3
						5	6

Tabla 3.10. Ejemplo de las hojas del árbol y el peso seco

Hojas	Porcentaje de peso seco			
1	3.28	3.09	3.03	3.03
2	3.52	3.48	3.38	3.38
3	2.88	2.8	2.81	2.76
4	3.34	3.38	3.23	3.26

Fuente: [40, p. 225].

4.1. El modelo

Sea

$$Y_{ij} = \mu + \tau_i + e_{ij} \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, r, \quad (3.12)$$

con

- Y_{ij} , la respuesta de la unidad experimental j en el tratamiento i .
- μ , la media general.
- τ_i , una variable aleatoria tal que $\tau_i \sim N(0, \sigma_\tau^2)$.
- e_{ij} , el error aleatorio tal que $e_{ij} \sim N(0, \sigma_e^2)$.
- e_{ij} y τ_i son no correlacionados.

La hipótesis de interés [41, p. 512] es:

$$H_0 : \sigma_\tau^2 = 0 \quad \text{versus} \quad H_a : \sigma_\tau^2 > 0.$$

Nota 3.9. Este diseño es semejante al diseño completamente aleatorizado de efectos fijos en cuanto al cálculo del P-valor del test. Se diferencia del diseño completamente aleatorizado de efectos fijos en las hipótesis de interés, en que los efectos de tratamiento son variables aleatorias, debido a las esperanzas de los cuadrados medios, y en la estimación de sus componentes de varianza. Ver detalles en el anexo 6.

Ejemplo 3.16. Continuando con el ejemplo 3.15. Presente el análisis de varianza y concluya.

Solución. Se tiene que:

$$\beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} \quad \text{entonces} \quad \hat{\beta} = \begin{pmatrix} 3.1656 \\ -0.0581 \\ 0.2744 \\ -0.3531 \end{pmatrix}.$$

En que $\mathbb{E}[(\tau_1 \ \tau_2 \ \dots, \tau_a)^\top] = \mathbf{0}$. Note que $\hat{\beta}$ es una realización del vector aleatorio β . Las hipótesis a probar son:

$$H_0 : \sigma_\tau^2 = 0 \quad \text{versus} \quad H_a : \sigma_\tau^2 > 0,$$

en que la hipótesis nula significa que la varianza debida a las hojas en el peso seco no es significativa y se debe solo a las replicaciones.

Para el análisis de varianza, ver tabla 3.11, se tiene que:

$$\mathbf{Y}^\top \mathbf{X} \hat{\beta} = 161.2273, \quad \mathbf{Y}^\top \mathbf{Y} = 161.3065 \quad \text{y} \quad n\bar{y}^2 = 160.3389.$$

Se concluye de la tabla 3.11 que la varianza del peso seco de las hojas es significativa.

Tabla 3.11. Tabla de análisis de varianza para el ejemplo 3.15

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Hoja	3	0.888	0.296	44.85	<0.001
Error	12	0.079	0.007		
Total	15	0.968			

Instrucciones en R para el ejemplo 3.15.

```
remove(list=ls())
hoja=gl(4,4) # Variable independiente
peso=c(3.28,3.09,3.03,3.03,3.52,3.48,3.38,3.38,2.88,2.8,2.81,
       2.76,3.34,3.38,3.23,3.26) # variable dependiente
hoja=as.factor(hoja) # especifica que hoja es una variable
                        # categórica
ajuste.hoj = aov(peso ~ hoja) # ajusta el modelo
summary(ajuste.hoj) #imprime en pantalla el análisis de varianza
```



Nota 3.10. Para estimar las componentes de varianza σ_τ^2 y σ_e^2 , se usa (B.35),

$$\hat{\sigma}_\tau^2 = \frac{CMM - CME}{r}$$

$$\hat{\sigma}_e^2 = CME.$$

Al estimar las componentes de varianza, puede ocurrir que $\hat{\sigma}_\tau^2 < 0$, que sería una contradicción porque las varianzas son números mayores o iguales a cero [41, p. 514] y [57, p. 60]. Las posibles causas de esta estimación negativa son:

1. El modelo que se está usando es errado.
2. $\sigma_\tau^2 = 0$ y el modelo se debería reducir a:

$$Y_{ij} = \mu + e_{ij}.$$

Algunas soluciones para este tipo de problema son:

1. Aumentar el número de unidades experimentales.
2. Usar otros métodos de estimación como, por ejemplo, máxima verosimilitud, máxima verosimilitud restringida o estimación por Bayes.

Ejemplo 3.17. Continuando con el ejemplo 3.15, estime los componentes de varianza.

Solución. Las componentes de varianza estimadas son:

$$\hat{\sigma}_\tau^2 = \frac{0.296 - 0.007}{4} = 0.07225$$

$$\hat{\sigma}_e^2 = 0.007.$$

✓

Nota 3.11. Para diseños no balanceados, en la nota 3.10 se reemplaza r por r_0 [41, p. 513]:

$$r_0 = \frac{1}{a-1} \left[\sum_{i=1}^a r_i - \frac{\sum_{i=1}^a r_i^2}{\sum_{i=1}^a r_i} \right].$$

4.2. Tamaño de muestra [21, p. 628]

En este caso es necesario buscar tanto a r como a a , el número de tratamientos a seleccionar².

Se sabe que, ver (B.36):

$$\frac{\sigma_e^2}{r\sigma_\tau^2 + \sigma_e^2} \frac{CMM}{CME} \sim F_{(a-1); a(r-1)}$$

$$\frac{1}{r \frac{\sigma_\tau^2}{\sigma_e^2} + 1} \times \frac{CMM}{CME} \sim F_{(a-1); a(r-1)}.$$

Sea $\gamma = \frac{\sigma_\tau^2}{\sigma_e^2}$ entonces:

$$\frac{1}{r\gamma + 1} \frac{CMM}{CME} \sim F_{(a-1); a(r-1)}.$$

²Puesto que el número de tratamientos a usar en el experimento es una muestra aleatoria de tratamientos.

Diseño completamente aleatorizado

Note que probar:

$$H_0 : \sigma_\tau^2 = 0 \quad \text{versus} \quad H_a : \sigma_\tau^2 > 0,$$

es equivalente a probar:

$$H_0 : \gamma = 0 \quad \text{versus} \quad H_a : \gamma = \gamma_1,$$

para $\gamma_1 > 0$, en que se rechaza $H_0 : \gamma = 0$ si

$$\begin{aligned} \frac{1}{r\gamma + 1} \frac{CMM}{CME} &> F_{(a-1);[a(r-1)];(1-\alpha)} \\ \frac{CMM}{CME} &> k, \end{aligned}$$

con $k = (r\gamma + 1)F_{(a-1);[a(r-1)];(1-\alpha)}$.

Por otro lado,

$$\begin{aligned} 1 - \beta &= P(\text{Rechazar } H_0 / H_0 \text{ falsa}) \\ 1 - \beta &= P\left(\frac{CMM}{CME} > k \mid \gamma = \gamma_1\right) \\ 1 - \beta &= P\left(\underbrace{\frac{1}{r\gamma_1 + 1} \times \frac{CMM}{CME}}_{\sim F_{(a-1);[a(r-1)]}} > \frac{k}{r\gamma_1 + 1}\right) \\ \beta &= P\left(\frac{1}{r\gamma_1 + 1} \times \frac{CMM}{CME} \leq \frac{k}{r\gamma_1 + 1}\right), \end{aligned}$$

entonces

$$F_{(a-1);[a(r-1)];\beta} = \frac{k}{r\gamma_1 + 1}.$$

Reemplazando k se tiene que:

$$\begin{aligned} F_{(a-1);[a(r-1)];\beta} &= \frac{(r\gamma + 1)F_{(a-1);[a(r-1)];(1-\alpha)}}{r\gamma_1 + 1} \\ \frac{r\gamma_1 + 1}{r\gamma + 1} &= \frac{F_{(a-1);[a(r-1)];(1-\alpha)}}{F_{(a-1);[a(r-1)];\beta}} \\ \frac{r\gamma_1 + 1}{r\gamma + 1} &= F_{(a-1);[a(r-1)];(1-\alpha)} F_{[a(r-1)];(a-1);(1-\beta)}. \end{aligned}$$

El tamaño de muestra se obtiene por ensayo y error, variando a y r .

Ejemplo 3.18. Suponga que se desea probar la hipótesis:

$$H_0 : \sigma_\tau^2 = \sigma_e^2 \quad \text{versus} \quad H_a : \sigma_\tau^2 > \sigma_e^2, \quad \text{es decir, } \gamma = 1,$$

para $\alpha = 0.05$ y $\beta = 0.10$. Se desea rechazar H_0 si $\frac{\sigma_\tau^2}{\sigma_e^2} = 2$, esto es, $\gamma_1 = 2$. En un experimento anterior se uso $r = 11$ y $a = 4$.

1. ¿Cuánto sería el tamaño de muestra si $r = 11$?
2. ¿Cuánto sería el tamaño de muestra si $r = 20$?
3. ¿Cuánto sería el tamaño de muestra si $r = 60$?
4. ¿Cuánto sería el tamaño de muestra si $r = 5$?

Solución

1. ¿Cuánto sería el tamaño de muestra si $r = 11$?

Se va a iniciar con $a = 4$ que era el número de tratamientos que se usaron en el experimento anterior que se está usando como piloto. Las iteraciones (ensayo y error) para la obtención del tamaño de muestra se presentan en la tabla 3.12.

$$\frac{r\gamma_1 + 1}{r\gamma + 1} = 1.9167 \approx 1.92.$$

Tabla 3.12. Tabla obtención tamaño de muestra del ejemplo para $r = 11$

a	$F_{(a-1);[a(r-1)];(1-\alpha)}$	$F_{[a(r-1)];(a-1);(1-\beta)}$	Producto	Decisión
4	$F_{3;40;0.95} = 2.84$	$F_{40;3;0.90} = 5.16$	$14.7396 > 1.92$	aumentar a
21	$F_{20;210;0.95} \approx 1.62$	$F_{210;20;0.90} \approx 1.63$	$2.64 > 1.92$	aumentar a
41	$F_{40;410;0.95} \approx 1.42$	$F_{410;40;0.90} \approx 1.39$	$1.97 > 1.92$	aumentar a
51	$F_{50;510;0.95} \approx 1.38$	$F_{510;50;0.90} \approx 1.34$	$1.85 < 1.92$	disminuir a
45	$F_{44;450;0.95} \approx 1.40$	$F_{450;44;0.90} \approx 1.37$	$1.918 < 1.92$	parar

Para $r = 11$ se tendrían que usar $a = 45$, entonces $n = 495$.

Diseño completamente aleatorizado

2. ¿Cuánto sería el tamaño de muestra si $r = 20$?

Las iteraciones para la obtención del tamaño de muestra se presentan en la tabla 3.13, esta vez se inició con $a = 45$.

$$\frac{r\gamma_1 + 1}{r\gamma + 1} = 1.9524 \approx 1.95.$$

Tabla 3.13. Tabla obtención tamaño de muestra del ejemplo para $r = 20$

a	$F_{(a-1);[a(r-1)];(1-\alpha)}$	$F_{[a(r-1)];(a-1);(1-\beta)}$	Producto	Decisión
45	$F_{44;855;0.95} \approx 1.39$	$F_{855;44;0.90} \approx 1.36$	$1.89 < 1.95$	disminuir a
20	$F_{19;380;0.95} \approx 1.61$	$F_{380;19;0.90} \approx 1.64$	$2.64 > 1.95$	aumentar a
30	$F_{29;570;0.95} \approx 1.49$	$F_{570;29;0.90} \approx 1.48$	$2.20 > 1.95$	aumentar a
40	$F_{39;760;0.95} \approx 1.42$	$F_{760;39;0.90} \approx 1.39$	$1.97 > 1.95$	aumentar a
43	$F_{42;817;0.95} \approx 1.40$	$F_{817;42;0.90} \approx 1.37$	$1.92 > 1.95$	parar a

Para $r = 20$ se tendrían que usar $a = 43$, entonces $n = 860$.

3. ¿Cuánto sería el tamaño de muestra si $r = 60$?

Las iteraciones para la obtención del tamaño de muestra se presentan en la tabla 3.14, se inició con $a = 40$.

$$\frac{r\gamma_1 + 1}{r\gamma + 1} = 1.9836 \approx 1.98.$$

Tabla 3.14. Tabla obtención tamaño de muestra del ejemplo para $r = 60$

a	$F_{(a-1);[a(r-1)];(1-\alpha)}$	$F_{[a(r-1)];(a-1);(1-\beta)}$	Producto	Decisión
40	$F_{39;2360;0.95} \approx 1.40$	$F_{2360;39;0.90} \approx 1.39$	$1.95 < 1.98$	disminuir a
30	$F_{29;1770;0.95} \approx 1.47$	$F_{1770;29;0.90} \approx 1.47$	$2.16 > 1.98$	aumentar a
35	$F_{34;2065;0.95} \approx 1.435$	$F_{2065;34;0.90} \approx 1.422$	$2.04 < 1.98$	aumentar a
38	$F_{37;2242;0.95} \approx 1.416$	$F_{2242;37;0.90} \approx 1.40$	$1.9824 < 1.9836$	parar

Para $r = 60$ se tendrían que usar $a = 38$, entonces $n = 2280$.

4. ¿Cuánto sería el tamaño de muestra si $r = 5$?

Las iteraciones para la obtención del tamaño de muestra se presentan en la tabla 3.15, se inició con $a = 50$.

$$\frac{r\gamma_1 + 1}{r\gamma + 1} = 1.8333 \approx 1.83.$$

Tabla 3.15. Tabla obtención tamaño de muestra del ejemplo para $r = 5$

a	$F_{(a-1);[a(r-1)];(1-\alpha)}$	$F_{[a(r-1)];(a-1);(1-\beta)}$	Producto	Decisión
50	$F_{49;200;0.95} \approx 1.418$	$F_{200;49;0.90} \approx 1.36$	$1.93 > 1.83$	aumentar a
100	$F_{99;400;0.95} \approx 1.284$	$F_{400;99;0.90} \approx 1.24$	$1.59 < 1.83$	disminuir a
70	$F_{69;280;0.95} \approx 1.346$	$F_{280;69;0.90} \approx 1.294$	$1.74 < 1.83$	disminuir a
60	$F_{59;240;0.95} \approx 1.377$	$F_{240;59;0.90} \approx 1.324$	$1.82 < 1.83$	parar

Para $r = 5$ se tendrían que usar $a = 60$, entonces $n = 300$.



5. Verificación de supuestos y transformaciones en un diseño completamente aleatorizado

Los supuestos que se deben verificar [61, p. 35] son:

- Errores independientes e idénticamente distribuidos.
- Homogeneidad de varianza.
- Errores provienen de una distribución normal³.

La verificación de supuestos usualmente se realiza sobre los residuales,

$$r_{ij} = y_{ij} - \hat{y}_{ij} \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, r.$$

Matricialmente se obtienen de $\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}}$ (otra notación usual para el vector de residuales es $\hat{\mathbf{E}}$). Los **residuos estandarizados** son definidos por:

$$t_{r_{ij}} = \frac{r_{ij}}{d.e. (r_{ij})} = \frac{r_{ij}}{\sqrt{\mathbb{V}(r_{ij})}}, \quad (3.13)$$

donde $d.e. (r_{ij})$ corresponde a la desviación estándar.

Nota 3.12. Note que:

$$\begin{aligned} \mathbb{V}(\mathbf{Y} - \hat{\mathbf{Y}}) &= \mathbb{V}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbb{V}\left(\mathbf{Y} - \mathbf{X}\left\{\left(\mathbf{X}^\top \mathbf{X}\right)^{-} \mathbf{X}^\top \mathbf{Y} + \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X}\right)^{-} \mathbf{X}^\top \mathbf{X}\right] \mathbf{z}\right\}\right) \\ &= \mathbb{V}(\mathbf{Y} - \mathbf{P}_x \mathbf{Y}) = (\mathbf{I}_n - \mathbf{P}_x) \mathbb{V}(\mathbf{Y}) (\mathbf{I}_n - \mathbf{P}_x)^\top \\ &= (\mathbf{I}_n - \mathbf{P}_x) \sigma_e^2 \mathbf{I}_n (\mathbf{I}_n - \mathbf{P}_x) = \sigma_e^2 (\mathbf{I}_n - \mathbf{P}_x). \end{aligned}$$

Si h_{kl} es el elemento (k, l) de la matriz de proyección $\mathbf{P}_x$, se tiene que:

$$\begin{aligned} \mathbb{V}(r_k) &= \sigma_e^2 (1 - h_{kk}); \quad k = 1, 2, 3, \dots, n, \quad (n = ar) \\ \mathbb{Cov}(r_k, r_l) &= -\sigma_e^2 h_{kl}, \end{aligned}$$

³El análisis de varianza es robusto ante desviaciones pequeñas de la normalidad.

esto es, los **residuos**, r_{ij} , **son correlacionados** (en comparación con los errores, e_{ij} , los cuales son no correlacionados). Luego (3.13) es dado por:

$$t_{r_k} = \frac{r_k}{\sigma_e \sqrt{1 - h_{kk}}}. \quad (3.14)$$

Como en la práctica σ_e^2 es desconocido, una alternativa es usar una estimación de σ_e , ver (3.5), que se va a denotar por S ,

$$t_{s_k} = \frac{r_k}{S \sqrt{1 - h_{kk}}},$$

con $S^2 = CME$, la varianza estimada de los residuales. Estos residuales también son conocidos como **residuales estudentizados internamente**. Un problema con este tipo de residuales es que el valor de S puede ser grande a causa de residuales grandes.

Otra alternativa es usar los **residuales estudentizados**:

$$t_{s_k}^* = \frac{r_k}{S_{(-k)} \sqrt{1 - h_{kk}}}, \quad (3.15)$$

siendo $S_{(-k)}^2$ la varianza estimada de los residuales cuando el modelo es ajustado sin la k -ésima observación⁴. La distribución de $t_{s_k}^*$ es:

$$t_{s_k}^* \sim t_{n-q-2},$$

siendo n el número de observaciones y q el número de parámetros. Si n es suficientemente grande

$$t_{s_k}^* \sim N(0, 1).$$

⁴Se puede encontrar $S_{(-k)}^2$ sin retirar el k -ésimo dato [2], usando la relación:

$$\begin{aligned} (n-p-1)S_{(-k)}^2 &= (n-p)S^2 - \frac{r_k^2}{1-h_{kk}} = (n-p)S^2 - \frac{S^2 r_k^2}{S^2(1-h_{kk})} \\ &= S^2 \left[(n-p) - \underbrace{\frac{r_k^2}{S^2(1-h_{kk})}}_{t_{s_k}^2} \right] \\ S_{(-k)}^2 &= S^2 \left(\frac{n-p-t_{s_k}^2}{n-p-1} \right), \end{aligned}$$

en que $p = \text{rango}(X)$.

Así, son considerados **marginalmente aberrantes**⁵ los puntos tales que:

$$|t_{s_k}^*| > 2.$$

Entre los gráficos más usados para hacer un chequeo de residuales están:

- **Gráfica de $\hat{y}_k$ versus r_k .** Sirve para verificar que no hayan problemas de homogeneidad de varianza. No debe haber ningún patrón ni creciente, ni decreciente, ni en forma de embudo. En caso de haber algún patrón, es posible que el supuesto de homogeneidad de varianza no se cumpla.
- **Gráfica de t versus r_t ,** siendo t la secuencia en que fueron tomados los datos. Sirve para verificar que no hayan problemas de dependencia. No debe haber ningún patrón ni creciente, ni decreciente. En caso de haber algún patrón, es posible que el supuesto de independencia no se cumpla. Es importante notar que para este gráfico es necesario *tener el orden* en que los datos fueron tomados.
- **Q-Qplot** sobre los residuales. Sirve para verificar que los residuales provienen de una distribución normal. En particular, los puntos atípicos afectan el análisis de varianza, por lo cual es necesario detectarlos y estudiarlos. Una forma de detectar puntos atípicos es verificar si

$$|r_{ij}| > 3\sqrt{CME},$$

en cuyo caso se considerará que r_{ij} es un punto atípico a estudiar.

- **Envelope.** Fue propuesto inicialmente por [2] y se refiere a una banda de confianza empírica para comparar residuales y percentiles de la distribución normal estándar, ver anexo 7. Si los residuales estudentizados están dentro de la banda empírica, la normalidad no estaría siendo violada, esto es, no hay nada en contra del modelo en revisión. Este gráfico sirve para detectar errores asimétricos a derecha o izquierda o si hay colas pesadas. También se puede detectar falta de coeficientes en el modelo.

⁵Observaciones atípicas aberrantes corresponden a observaciones mal ajustadas con residuo alto.

Ejemplo 3.19. Continuando con el ejemplo 3.3, sobre los fertilizantes y la altura de la planta de colza, un envelope se presenta en la figura 3.4.

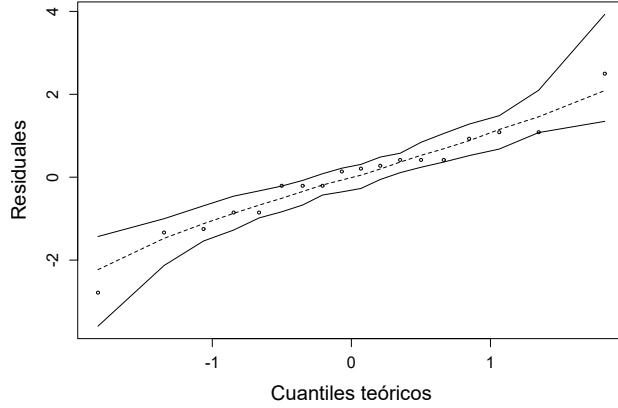


Figura 3.4. Envelope para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 3.3

De igual forma, en la figura 3.5 se pueden observar los tres gráficos mencionados para evaluar homogeneidad de varianza 3.5(a), independencia 3.5(b) y normalidad 3.5(c). Vale aclarar que el orden tomado para el gráfico de independencia (b) es **ilustrativo**, ya que no se dispone del orden original en que se tomaron las alturas de las plantas.

5.1. Gráfico Q-Q plot para los datos $x_1, x_2, \dots, x_n$

El gráfico se construye siguiendo estos pasos:

1. Se ordenan los datos de menor a mayor, generando las estadísticas ordenadas $x_{(1)}, x_{(2)}, \dots, x_{(m)}, \dots, x_{(n-1)}, x_{(n)}$, en que $x_{(1)}$ es el menor dato y $x_{(n)}$ es el mayor dato.
2. Para cada $m, m = 1, 2, \dots, n$, se calcula:

$$f_m = \frac{m - 0.5}{n} \quad \text{y} \quad z_k = \Phi^{-1}(f_m),$$

que es el percentil de la distribución normal estándar asociado a f_m .

Diseño completamente aleatorizado

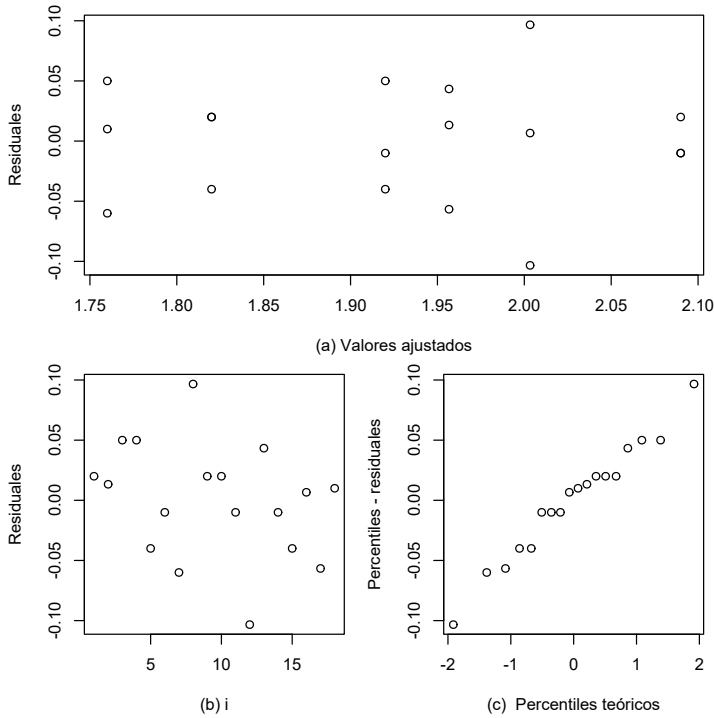


Figura 3.5. Graficos para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 3.3

3. Se grafica los z_m versus $x_{(m)}$.

Si los datos quedan sobre una línea se considera que no hay problemas de normalidad.

Ejemplo 3.20. Realice un Q-Q plot para los residuales del ejemplo 3.3.

Solución. Los residuales, $r_k = y_k - \hat{y}_k$, obtenidos son:

$$\begin{aligned}
 r_1 &= 0.020, r_2 = -0.040, r_3 = 0.020, r_4 = 0.043, r_5 = -0.057, \\
 r_6 &= 0.013, r_7 = -0.010, r_8 = 0.020, r_9 = -0.010, r_{10} = 0.010, \\
 r_{11} &= 0.050, r_{12} = -0.060, r_{13} = -0.010, r_{14} = -0.040, r_{15} = 0.050, \\
 r_{16} &= 0.097, r_{17} = -0.103, r_{18} = 0.007.
 \end{aligned}$$

1. Residuales ordenados:

$$\begin{aligned} r_{(1)} &= -0.103, r_{(2)} = -0.060, r_{(3)} = -0.057, r_{(4)} = -0.040, \\ r_{(5)} &= -0.040, r_{(6)} = -0.010, r_{(7)} = -0.010, r_{(8)} = -0.010, \\ r_{(9)} &= 0.007, r_{(10)} = 0.010, r_{(11)} = 0.013, r_{(12)} = 0.020, \\ r_{(13)} &= 0.020, r_{(14)} = 0.020, r_{(15)} = 0.043, r_{(16)} = 0.050, \\ r_{(17)} &= 0.050, r_{(18)} = 0.097. \end{aligned}$$

2. f_m y z_m son, respectivamente,

$$\begin{aligned} f_1 &= \frac{1 - 0.5}{18} = 0.028, f_2 = \frac{2 - 0.5}{18} = 0.083, f_3 = 0.139, \\ f_4 &= 0.194, f_5 = 0.250, f_6 = 0.306, f_7 = 0.361, f_8 = 0.417, \\ f_9 &= 0.472, f_{10} = 0.528, f_{11} = 0.583, f_{12} = 0.639, f_{13} = 0.694, \\ f_{14} &= 0.750, f_{15} = 0.806, f_{16} = 0.861, f_{17} = 0.917, f_{18} = 0.972. \end{aligned}$$

$$\begin{aligned} z_1 &= \Phi^{-1}(0.028) = -1.91, z_2 = \Phi^{-1}(0.083) = -1.38, \\ z_3 &= -1.09, z_4 = -0.86, z_5 = -0.67, z_6 = -0.51, z_7 = -0.36, \\ z_8 &= -0.21, z_9 = -0.07, z_{10} = 0.07, z_{11} = 0.21, z_{12} = 0.36, \\ z_{13} &= 0.51, z_{14} = 0.67, z_{15} = 0.86, z_{16} = 1.09, z_{17} = 1.38, \\ z_{18} &= 1.91. \end{aligned}$$

3. La gráfica de z_m versus $r_{(m)}$ se puede observar en la figura 3.5(c). ✓

5.2. Test para normalidad y homogeneidad de varianza

1. Entre los test para verificar normalidad están [50]:

- El test de Shapiro-Wilk [59].
- El test de Anderson-Darling [1].
- El test de Lilliefors (Kolmogorov-Smirnov) [39].

- El test de Pearson chi-cuadrado.
- El test de Shapiro-Francia [60].

De estos test, los más clásicos son el test de Shapiro-Wilk y el test de Kolmogorov-Smirnov.

2. Hay varios test para evaluar homogeneidad de varianzas, entre ellos están:

- El test de Levene [38] y [9].
- El test de Bartlett [7] y [13].
- El test de Fligner-Killeen [26] y [18].

Sin embargo, el test paramétrico más conocido es el test de Bartlett.

5.2.1. Validación de supuestos en R

Continuando con el ejemplo 3.3 sobre la altura de plantas de colza.

1. Para la validación del supuesto de normalidad, se pueden usar varios test que vienen en la librería *nortest*, una vez sea ajustado el modelo. El código en R para verificar normalidad es:

```
### correr instrucciones página 67 hasta obtener ###
### el modelo ajuste1 ###
library(nortest)
ad.test(ajuste1$residuals) # Test de Anderson-Darling
cvm.test(ajuste1$residuals) # Test de Cramer-von Mises
lillie.test(ajuste1$residuals) # Test de Lilliefors
                                # (Kolmogorov-Smirnov)
pearson.test(ajuste1$residuals) # Test de Pearson
                                # chi-square
sf.test(ajuste1$residuals) # Test de Shapiro-Francia
shapiro.test(ajuste1$residuals) # Test de Shapiro-Wilk
```

2. Para verificar homogeneidad de varianzas:

```
bartlett.test(ajuste1$residuals~fertilizante) # Test de
# Bartlett para homogeneidad de Varianzas - paramétrico
fligner.test(ajuste1$residuals~fertilizante) # Test
# Figner-Killeen para homogeneidad de
# Varianzas - no paramétrico
library(lawstat)
levene.test(ajuste1$residuals,fertilizante) # Test de
# Levene para homogeneidad de Varianzas
```

3. Instrucciones para realizar los gráficos de las figuras 3.5 y 3.4:

```
tiempo = c(1,5,9,13,17,2,6,10,14,18,3,7,11,15,4,8,12,16)
# orden "ilustrativo" correspondiente a la
# recolección de los datos
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
plot(ajuste1$fit,ajuste1$residuals,
     xlab="(a) Valores Ajustados",ylab="Residuales")

plot(tiempo,ajuste1$residuals,xlab="(b) t",
     ylab="Residuales")
qqnorm(ajuste1$residuals,xlab="(c) Percentiles teóricos",
     ylab="Percentiles - residuales",main= " ")
library(hnp) #libreria para realizar un envelope
par(mfrow=c(1,1))
ajuste.hnp=hnp(ajuste1,sim=99,conf=0.95,
               resid.type="student",halfnormal=FALSE)
# elaboración del envelope
```

5.3. Algunas transformaciones

Un alternativa al problema del cumplimiento de supuestos es considerar la transformación de los datos. Algunas transformaciones usadas para validación de supuestos se presentan en la tabla 3.16. Vale resaltar que el mayor cuidado al usar transformaciones de los datos es no perder la interpretabilidad de los resultados. Otra alternativa es usar modelos lineales generalizados, por ejemplo, que tienen en cuenta la distribución de la variable respuesta dentro de una familia de distribuciones más amplia que la distribución normal.

Tabla 3.16. Algunas transformaciones

Variable respuesta tipo:	Transformación
Conteo	$\sqrt{y}$
s_{ij}^2	$\ln(y)$
Número de éxitos	$\arcsen\left(\sqrt{\frac{y}{n}}\right)$
Correlaciones (r_{ij})	$\frac{1}{2}\ln\left(\frac{1+y}{1-y}\right)$

5.3.1. Transformaciones de potencia

Este tipo de transformación es muy utilizado y se trata de encontrar λ tal que minimice la suma de cuadrados del error. Un método es graficar la suma de cuadrados del error para diferentes valores de λ y escoger el que minimice dicha suma.

$$f(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{Si } \lambda \neq 0 \\ \ln(y) & \text{Si } \lambda = 0. \end{cases}$$

6. Contrastes

Sea una función $\lambda^\top \beta$ estimable, ver definición 2.1, tal que:

$$\Gamma = \lambda^\top \beta = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_a \end{pmatrix} \begin{pmatrix} \mu & \tau_1 & \tau_2 & \dots & \tau_a \end{pmatrix}^\top.$$

Su estimador es dado por, ver teorema 2.1, $\hat{\Gamma} = \lambda^\top \hat{\beta}$, con β solución de las ecuaciones normales $X^\top X \beta = X^\top y$,

$$\hat{\Gamma} = \lambda^\top \hat{\beta} = \begin{pmatrix} c_0 & c_1 & \dots & c_a \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_a \end{pmatrix} = c_0 \hat{\mu} + \sum_{i=1}^a c_i \hat{\tau}_i.$$

Como $\Gamma = \boldsymbol{\lambda}^\top \boldsymbol{\beta}$ es estimable, $\hat{\Gamma}$ es invariante a soluciones de las ecuaciones normales. Si, por ejemplo,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} \mathbf{1}_n^\top \\ - - - \\ \mathbf{I}_a \otimes \mathbf{1}_r^\top \end{pmatrix} \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{ar} \end{pmatrix} \\ &= \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} \sum_{i=1}^a \sum_{j=1}^r Y_{ij} \\ \sum_{j=1}^r Y_{1j} \\ \vdots \\ \sum_{j=1}^r Y_{aj} \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{Y}_{1*} \\ \vdots \\ \bar{Y}_{a*} \end{pmatrix},\end{aligned}$$

la notación $\otimes$, que indica producto Kronecker, se presenta en nota a pie de página 2 del apéndice B.

$\hat{\Gamma}$ sería dado por:

$$\hat{\Gamma} = c_0 \hat{\mu} + \sum_{i=1}^a c_i \hat{\tau}_i = \sum_{i=1}^a c_i \bar{Y}_{i*}, \quad (3.16)$$

siendo, ver nota 2.5:

$$\mathbb{E}(\hat{\Gamma}) = \mathbb{E}(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = \Gamma \quad (3.17)$$

y

$$\begin{aligned}\mathbb{V}(\hat{\Gamma}) &= \mathbb{V}(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = \sigma_e^2 \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} = \sigma_e^2 \boldsymbol{\lambda}^\top \left[\frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \right] \boldsymbol{\lambda} \\ &= \frac{\sigma_e^2}{r} \begin{pmatrix} c_0 & c_1 & \dots & c_a \end{pmatrix} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_a \end{pmatrix} = \frac{\sigma_e^2}{r} \begin{pmatrix} c_0 & c_1 & \dots & c_a \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_a \end{pmatrix} \\ &= \frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2. \quad (3.18)\end{aligned}$$

Luego

$$\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} = \frac{1}{r} \sum_{i=1}^a c_i^2. \quad (3.19)$$

La respectiva suma de cuadrados asociada a la hipótesis $H_0 : \Gamma = 0$ es dada por, ver nota 2.6:

$$\begin{aligned} SC(\hat{\Gamma}) &= (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})^\top \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}) = \hat{\Gamma}^\top \left(\frac{1}{r} \sum_{i=1}^a c_i^2 \right)^{-1} \hat{\Gamma} \\ &\stackrel{\text{Usando (3.19) y (3.16)}}{=} r \left(\sum_{i=1}^a c_i \bar{Y}_{i*} \right) \left(\sum_{i=1}^a c_i^2 \right)^{-1} \left(\sum_{i=1}^a c_i \bar{Y}_{i*} \right) \\ SC(\hat{\Gamma}) &= r \frac{\left(\sum_{i=1}^a c_i \bar{Y}_{i*} \right)^2}{\sum_{i=1}^a c_i^2}. \end{aligned} \quad (3.20)$$

6.1. Distribución y esperanza de las suma de cuadrados de $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}$

Sea

$$SC(\hat{\Gamma}) = (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}})^\top \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}),$$

en que:

$$\begin{aligned} \boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} &= \boldsymbol{\lambda}^\top \left\{ \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{Y} + \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{X} \right] \mathbf{z} \right\} \\ &= \boldsymbol{\lambda}^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{Y} + \underbrace{\boldsymbol{\lambda}^\top \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{X} \right] \mathbf{z}}_0 \\ &= \boldsymbol{\lambda}^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{Y}, \end{aligned}$$

entonces:

$$\begin{aligned} SC(\hat{\Gamma}) &= \mathbf{Y}^\top \underbrace{\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}_A \\ &= \mathbf{Y}^\top \mathbf{A} \mathbf{Y}, \end{aligned}$$

con

$$\mathbf{A} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top. \quad (3.21)$$

Sea $\mathbf{Y}^* = \frac{1}{\sigma_e} \mathbf{Y}$, luego $\frac{1}{\sigma_e} \mathbf{Y} \sim N(\frac{1}{\sigma_e} \mathbf{X}\boldsymbol{\beta}, \mathbf{I})$, ver (B.11). Para encontrar la distribución de la suma de cuadrados asociada a la hipótesis $H_0 : \Gamma = 0$, $SC(\hat{\Gamma})$, note que:

- La matriz $\mathbf{A} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ es simétrica.
- La idempotencia de $\mathbf{A}\mathbf{V}^*$:

$$\begin{aligned} \mathbf{A}\mathbf{V}^* \mathbf{A}\mathbf{V}^* &= \\ \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left[\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right]^{-1} &\underbrace{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}}_{\boldsymbol{\lambda}^\top} \times \\ &\left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left[\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right]^{-1} \cancel{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}} \left[\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right]^{-1} \times \\ &\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{A}\mathbf{V}^*. \end{aligned}$$

Por tanto, $\mathbf{A}\mathbf{V}^*$ es idempotente al igual que $\mathbf{A}$ es idempotente y

$$\frac{SC(\hat{\Gamma})}{\sigma_e^2} \sim \chi^2(rango(\mathbf{A}), \lambda),$$

siendo:

$$\begin{aligned} rango(\mathbf{A}) &= traza(\mathbf{A}) \\ &= traza \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \\ &= traza \left(\left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \underbrace{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}}_{\boldsymbol{\lambda}^\top} \right) \\ &= traza(1) = 1 \end{aligned}$$

y

$$\begin{aligned}
 \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* \\
 &= \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \underbrace{\mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}}_{\boldsymbol{\lambda}} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \underbrace{\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}}_{\boldsymbol{\lambda}^\top} \\
 &= \frac{1}{2\sigma_e^2} \underbrace{\boldsymbol{\beta}^\top \boldsymbol{\lambda}}_{\Gamma} \left(\boldsymbol{\lambda}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda} \right)^{-1} \underbrace{\boldsymbol{\lambda}^\top \boldsymbol{\beta}}_{\Gamma} \\
 &\stackrel{\text{Usando (3.19)}}{=} \frac{1}{2\sigma_e^2} \Gamma^2 \left(\frac{1}{r} \sum_{i=1}^a c_i^2 \right)^{-1} = \frac{r\Gamma^2}{2\sigma_e^2 \sum_{i=1}^a c_i^2},
 \end{aligned}$$

por tanto,

$$\frac{SC(\hat{\Gamma})}{\sigma_e^2} \sim \chi^2 \left(1, \frac{r\Gamma^2}{2\sigma_e^2 \sum_{i=1}^a c_i^2} \right).$$

La esperanza de $SC(\hat{\Gamma})$ es dada por:

$$\begin{aligned}
 \mathbb{E} [SC(\hat{\Gamma})] &= \text{traza}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \\
 &= \text{traza}(\sigma_e^2 \mathbf{A}) + \frac{r}{\sum_{i=1}^a c_i^2} \Gamma^2 = \sigma_e^2 + \frac{r\Gamma^2}{\sum_{i=1}^a c_i^2}.
 \end{aligned}$$

Sea $\mathbf{A}_2 = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{I} - \mathbf{P}_x$. El cociente:

$$\frac{CM(\hat{\Gamma})}{CME} \sim F(1, a(r-1), \lambda), \quad (3.22)$$

con $\lambda = \frac{r\Gamma^2}{2\sigma_e^2 \sum_{i=1}^a c_i^2}$. Bajo la hipótesis $H_0 : \Gamma = 0$ se tiene que:

$$\frac{CM(\hat{\Gamma})}{CME} \sim F(1, a(r-1)) = t_{a(r-1)}^2.$$

Así, se rechaza $H_0 : \Gamma = 0$ si:

$$\frac{CM(\hat{\Gamma})}{CME} \geq F(1, a(r-1), 1-\alpha). \quad (3.23)$$

Finalmente, a Γ se le considera un **contraste** si $c_0 = 0$ y

$$\sum_{i=1}^a c_i = 0,$$

cuando $r_i = r$, para todo i . Si $r_i \neq r_j$ para algún $i \neq j$,

$$\sum_{i=1}^a r_i c_i = 0.$$

Para un diseño a una vía de clasificación, un **contraste siempre es estimable**.

Veamos: Sea

$$\lambda^\top = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_a \end{pmatrix} = \begin{pmatrix} 0 & c_1 & c_2 & \dots & c_a \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{c}^\top \end{pmatrix},$$

tal que $\sum_{i=1}^a c_i = \mathbf{c}^\top \mathbf{1}_a = 0$. Se va a usar la nota 2.4 para mostrar que un contraste siempre es estimable.

La matriz diseño (para el caso balanceado), en el caso de un diseño a una vía de clasificación, es dada por:

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix},$$

luego

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix}^\top \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix} \\ &= \begin{pmatrix} ra & r\mathbf{1}_a^\top \\ r\mathbf{1}_a & r\mathbf{I}_a \end{pmatrix} = r \begin{pmatrix} a & \mathbf{1}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} \\ (\mathbf{X}^\top \mathbf{X})^- &= \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \quad \text{es una de las posibles inversas} \end{aligned}$$

generalizadas. Para más generalidad, se puede usar una inversa generalizada a partir de una partición de $\mathbf{X}^\top \mathbf{X}$.

$$\begin{aligned} (\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top \mathbf{X} &= \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} a & \mathbf{1}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} \\ \lambda^\top (\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top \mathbf{X} &= \lambda^\top \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{c}^\top \end{pmatrix} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} \\ &= \begin{pmatrix} 0 + \cancel{\mathbf{c}^\top \mathbf{1}_a} \mathbf{0}_a^\top + \mathbf{c}^\top \mathbf{I}_a \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{c}^\top \end{pmatrix} = \lambda^\top. \end{aligned}$$

Luego un contraste siempre es estimable.



6.2. Contrastes ortogonales

Los contrastes:

$$\Gamma = \sum_{i=1}^a c_i \tau_i \quad y \quad \Gamma' = \sum_{i=1}^a d_i \tau_i,$$

son ortogonales si:

$$\begin{cases} \sum_{i=1}^a c_i d_i = 0 & \text{cuando } r_i = r \quad \forall i \text{ o} \\ \sum_{i=1}^a r_i c_i d_i = 0 & \text{cuando } r_i \neq r_j \text{ para algún } i \neq j. \end{cases}$$

Para a tratamientos existen conjuntos de $a - 1$ contrastes ortogonales. Esto implica que la suma de cuadrados de tratamiento se puede dividir en $a - 1$ sumas de cuadrados, cada una con un grado de libertad (ver tabla 3.17).

Tabla 3.17. Análisis de varianza para un diseño completamente aleatorizado y contrastes ortogonales

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	$\mathbb{E}(CM)$
Tratam.	$a - 1$	$SC(Tratam)$	$\frac{SC_{Trat}}{gl_{Trat}}$	$\sigma_e^2 + \frac{r}{a-1} \sum_{i=1}^a \tau_i^2$
C_1	1	$SC(C_1)$	$SC(C_1)$	$\sigma_e^2 + \frac{r(C_1^\top \beta)^2}{\sum_{i=1}^a c_{i1}^2}$
C_2	1	$SC(C_2)$	$SC(C_2)$	$\sigma_e^2 + \frac{r(C_2^\top \beta)^2}{\sum_{i=1}^a c_{i2}^2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_{a-1}	1	$SC(C_{a-1})$	$SC(C_{a-1})$	$\sigma_e^2 + \frac{r(C_{a-1}^\top \beta)^2}{\sum_{i=1}^a c_{i,a-1}^2}$
Error	$a(r - 1)$	$SC(Error)$		σ_e^2
Total	$ar - 1$	$SC(Total)$		

Ejemplo 3.21. Continuando con el ejemplo 3.3 sobre los fertilizantes y la altura de las plantas de colza ($\alpha = 6$), presente un conjunto de contrastes ortogonales y su respectivo análisis de varianza.

Solución. El conjunto de contrastes ortogonales puede ser:

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$C_1 :$	1	1	1	-1	-1	-1
$C_2 :$	2	-1	-1	0	0	0
$C_3 :$	0	0	0	2	-1	-1
$C_4 :$	0	1	-1	0	0	0
$C_5 :$	0	0	0	0	1	-1

En la tabla 3.18 se presenta su respectivo análisis de varianza.

Tabla 3.18. Análisis de varianza para los datos sobre fertilizantes y altura de las plantas de colza en contrastes ortogonales

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	$\mathbb{E}(CM)$
Fertiliz.	5	0.21792	0.04358	$\sigma_e^2 + \frac{3}{5} \sum_{i=1}^6 \tau_i^2$
C_1	1	0.01681	0.01681	$\sigma_e^2 + \frac{3(\tau_1 + \tau_2 + \tau_3 - \tau_4 - \tau_5 - \tau_6)^2}{6}$
C_2	1	0.08269	0.08269	$\sigma_e^2 + \frac{3(2\tau_1 - \tau_2 - \tau_3)^2}{6}$
C_3	1	0.08134	0.08134	$\sigma_e^2 + \frac{3(2\tau_4 - \tau_5 - \tau_6)^2}{6}$
C_4	1	0.02667	0.02667	$\sigma_e^2 + \frac{3(\tau_2 - \tau_3)^2}{2}$
C_5	1	0.01042	0.01042	$\sigma_e^2 + \frac{3(\tau_5 - \tau_6)^2}{2}$
Error	$6(3 - 1)$	0.03873	0.00323	σ_e^2
Total	$18 - 1$	0.25665		

Las instrucciones en R para estimar los contrastes propuestos en el ejemplo 3.21 y hacer inferencia sobre los mismos son:

Diseño completamente aleatorizado

```
### correr instrucciones página 67 hasta obtener ###
### el modelo ajuste1 ###
c1=c(1,1,1,-1,-1,-1)
c2=c(2,-1,-1,0,0,0)
c3=c(0,0,0,2,-1,-1)
c4=c(0,1,-1,0,0,0)
c5=c(0,0,0,0,1,-1)
union=cbind(c1,c2,c3,c4,c5)
library(emmeans)
misContr.emmc <- function(levs, ...) {
  M <- as.data.frame(union)
  names(M) <- c("C1","C2","C3","C4","C5")
  attr(M, "desc") <- "Mis contrastes"
  M
}
estimador=emmeans(ajuste1,"fertilizante",level=0.95) # presenta
# el intervalo de confianza para cada tratamiento
contrast(estimador, method="misContr")
a=contrast(estimador, method="misContr")
CME=deviance(ajuste1)/df.residual(ajuste1) # Estimación
# varianza de los errores
(summary(a)$t.ratio)^2*CME # Suma de cuadrados para cada
# contraste
```



6.3. Pruebas de comparación múltiple

6.3.1. Método de Scheffe [55]

Considere el contraste:

$$\Gamma = \sum_{i=1}^a c_i \tau_i.$$

De (3.16), (3.18), (3.20) y (3.22) se sabe que:

$$\begin{aligned}\hat{\Gamma} &= \sum_{i=1}^a c_i \bar{Y}_{i*}, \quad \mathbb{V}(\hat{\Gamma}) = \frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2 \\ CM(\hat{\Gamma}) &= SC(\hat{\Gamma}) = r \frac{\left(\sum_{i=1}^a c_i \bar{Y}_{i*} \right)^2}{\sum_{i=1}^a c_i^2} = \frac{r \hat{\Gamma}^2}{\sum_{i=1}^a c_i^2} \\ \frac{CM(\hat{\Gamma})}{CME} &\sim F_{1,a(r-1),\lambda},\end{aligned}$$

$$\text{con } \lambda = \frac{r\Gamma^2}{2\sigma_e^2 \sum_{i=1}^a c_i^2}.$$

Denotando $\hat{\mathbb{V}}(\hat{\Gamma}) = S_{c_u}^2$ y usando que $\hat{\sigma}_e^2 = CME$, se tiene que:

$$S_{c_u}^2 = \frac{CME}{r} \sum_{i=1}^a c_i^2$$

$$CME = \frac{r S_{c_u}^2}{\sum_{i=1}^a c_i^2}.$$

También se sabe que se rechaza $H_0 : \Gamma = 0$ si, ver (3.23),

$$\begin{aligned}F_{1,a(r-1),1-\alpha} &\leq \frac{CM(\hat{\Gamma})}{CME} \\ &= \frac{\hat{\Gamma}^2}{\sum_{i=1}^a c_i^2} \\ F_{1,a(r-1),1-\alpha} &\leq \frac{\hat{\Gamma}^2}{S_{c_u}^2} \\ &= \frac{\hat{\Gamma}^2}{S_{c_u}^2} \\ S_{c_u}^2 F_{1,a(r-1),1-\alpha} &\leq \hat{\Gamma}^2 \\ S_{c_u} \sqrt{F_{1,a(r-1),1-\alpha}} &\leq |\hat{\Gamma}|.\end{aligned}\tag{3.24}$$

En (3.24), [55] sugirió reemplazar $F_{1,a(r-1),1-\alpha}$ por $(a-1) \times F_{a-1,a(r-1),1-\alpha}$ y, por tanto, el test de Scheffe rechaza $H_0 : \Gamma = 0$ si

$$|\hat{\Gamma}| \geq S_{cu} \sqrt{(a-1)F_{a-1,a(r-1),1-\alpha}}.$$

Ejemplo 3.22. Continuando con el ejemplo sobre los fertilizantes, ejemplo 3.3,

1. Verificar la hipótesis de que el efecto del fertilizante control, $N = 0$, $P_2O_5 = 0$, es igual al efecto del fertilizante $N = 0$, $P_2O_5 = 40\text{kg/ha}$ sobre la altura de la planta de colza, con un nivel de significancia de 0.05.
2. Verificar la hipótesis de que el promedio de los efectos de los fertilizantes con $N = 75$, $P_2O_5 = 0\text{kg/ha}$ y $N = 150$, $P_2O_5 = 0\text{kg/ha}$ es igual al efecto del fertilizante $N = 150$, $P_2O_5 = 40\text{kg/ha}$, sobre la altura de las plantas de colza, con un nivel de significancia de 0.05.

Solución

1. Del enunciado se tiene que:

$$\Gamma = \tau_4 - \tau_1, \text{ es decir, } c_1 = -1, c_2 = 0, c_3 = 0, c_4 = 1, c_5 = 0, c_6 = 0,$$

y

$$\hat{\Gamma} = \sum_{i=1}^a c_i \bar{y}_{i*} = \bar{y}_{4*} - \bar{y}_{1*} = 1.76 - 1.82 = -0.06.$$

La hipótesis nula es: $H_0 : \Gamma = 0$, es decir,

$$H_0 : \tau_4 - \tau_1 = 0 \text{ versus } H_a : \tau_4 - \tau_1 \neq 0.$$

Se rechaza H_0 si

$$|\hat{\Gamma}| \geq \sqrt{S_{cu}^2 (6-1) F_{6-1,6(3-1);0.95}},$$

siendo

$$S_{cu}^2 = \frac{CME}{r} \sum_{i=1}^a c_i^2 = \frac{0.00323}{3} [1^2 + (-1)^2] = \frac{2 \times 0.00323}{3}.$$

Como

$$F_{5,12;0.95} = 3.11$$

y

$$\sqrt{S_{c_u}^2 (6-1) F_{5;12;0.95}} = \sqrt{\frac{2 \times 0.00323}{3} \times 5 \times 3.11} = 0.1830,$$

se concluye⁶ que no se puede rechazar la hipótesis nula ya que no se tiene suficientes evidencias para considerar que hay diferencia entre el efecto del fertilizante control y el fertilizante $N = 0, P_2O_5 = 40kg/ha$ sobre la altura de las plantas de colza, con un nivel de significancia de 0.05.

2. Del enunciado se tiene que:

$$\Gamma = \frac{\tau_2 + \tau_3}{2} - \tau_6, \text{ es decir, } c_1 = 0, c_2 = \frac{1}{2}, c_3 = \frac{1}{2}, c_4 = 0, c_5 = 0, \\ c_6 = -1$$

y

$$\hat{\Gamma} = \sum_{i=1}^a c_i \bar{y}_{i*} = \frac{\bar{y}_{2*}}{2} + \frac{\bar{y}_{3*}}{2} - \bar{y}_{6*} = \frac{1.957}{2} + \frac{2.09}{2} - 2.003 = 0.02.$$

La hipótesis nula es: $H_0 : \Gamma = 0$, es decir,

$$H_0 : \frac{\tau_2 + \tau_3}{2} - \tau_6 = 0 \text{ versus } H_a : \frac{\tau_2 + \tau_3}{2} - \tau_6 \neq 0.$$

Se rechaza H_0 si

$$|\hat{\Gamma}| \geq \sqrt{S_{c_u}^2 (6-1) F_{6-1,6(3-1),0.95}},$$

en que

$$S_{c_u}^2 = \frac{CME}{r} \sum_{i=1}^a c_i^2 = \frac{0.00323}{3} \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 + (-1)^2 \right] \\ = \frac{0.00323}{3} \left[\frac{6}{4} \right] = \frac{0.00323}{2}.$$

Como

$$F_{5;12;0.95} = 3.11$$

$$^{6}0.06 = |\hat{\Gamma}| < \sqrt{S_{c_u}^2 (6-1) F_{5;12;0.95}} = 0.1830.$$

y

$$\sqrt{S_{cu}^2 (6 - 1) F_{5;12;0.95}} = \sqrt{\frac{0.00323}{2} \times 5 \times 3.11} = 0.1585,$$

se concluye⁷ que no se rechazar la hipótesis nula ya que no se tienen suficientes evidencias para considerar que el promedio de los efectos de los fertilizantes $N = 75$, $P_2O_5 = 0kg/ha$ y $N = 150$, $P_2O_5 = 0kg/ha$ sea diferente al efecto del fertilizante $N = 150$, $P_2O_5 = 40kg/ha$, sobre la altura de las plantas de colza, con un nivel de significancia de 0.05. $\checkmark$

Nota 3.13. Cuando cada una de las diferentes pruebas de hipótesis de la forma $H_0 : \mu_i = \mu_{i'}, i \neq i'$, en que $(\mu_i = \mu + \tau_i)$, son realizadas al mismo nivel de significancia α , a α se la llama **nivel de significancia por comparación** [51].

Nota 3.14. La dificultad con hacer cada comparación $H_0 : \mu_i = \mu_{i'}, i \neq i'$, a un nivel de significancia α es la posible inflación del error tipo I (o de deflación del nivel de confianza). Para un conjunto de tratamientos, el número de diferencias $\mu_i - \mu_{i'}, i \neq i'$, diferentes es $\frac{a(a-1)}{2}$. La probabilidad de que al menos con una se cometa error tipo I es

$$\alpha_c = 1 - (1 - \alpha)^c,$$

con $c = \frac{a(a-1)}{2}$. Así, para un número grande de comparaciones $\mu_i = \mu_{i'}, i \neq i'$, se concluirá que una parte de tratamientos son diferentes cuando en realidad **no** lo son.

Nota 3.15. Una **familia** es una colección de “diferencias” (comparaciones a realizar) para la cual es importante tener en cuenta alguna medida de error (por ejemplo, el error tipo I) [51].

Nota 3.16. La **tasa de error** en pruebas de hipótesis representa la probabilidad de error tipo I. Así, una tasa de error más pertinente es la tasa de error para family wise (FWER)⁸ [51].

⁷0.02 = $|\hat{\Gamma}| < \sqrt{S_{cu}^2 (6 - 1) F_{5;12;0.95}} = 0.1585$

⁸Es la probabilidad de rechazar de forma incorrecta al menos una de las hipótesis nulas que componen el grupo de hipótesis (familia).

6.3.2. Método LSD de Fisher

Sea el contraste $\Gamma = \sum_{i=1}^a c_i \tau_i$ y su estimador $\hat{\Gamma} = \sum_{i=1}^a c_i \bar{Y}_{i*}$. De (3.17) y (3.18) se tiene⁹:

$$\hat{\Gamma} \sim N \left(\sum_{i=1}^a c_i \tau_i, \frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2 \right)$$

$$\frac{\hat{\Gamma}}{\sqrt{\frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2}} \sim N \left(\frac{\Gamma}{\sqrt{\frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2}}, 1 \right).$$

Por otro lado, de (B.12) se tiene que:

$$\frac{SCE}{\sigma_e^2} \sim \chi_{a(r-1)}^2,$$

entonces:

$$\frac{\frac{\hat{\Gamma}}{\sqrt{\frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2}}}{\sqrt{\frac{SCE}{\sigma_e^2 a(r-1)}}} = \frac{\hat{\Gamma}}{\sqrt{\frac{CME \sum_{i=1}^a c_i^2}{r}}} \sim t_{a(r-1), \delta} \quad (3.25)$$

con

$$\delta = \frac{\Gamma}{\sqrt{\frac{\sigma_e^2}{r} \sum_{i=1}^a c_i^2}}.$$

Si $\Gamma = \tau_i - \tau_{i'}, i \neq i'$, (3.25) se puede escribir como:

$$\frac{\bar{Y}_{i*} - \bar{Y}_{i'*}}{\sqrt{\frac{2CME}{r}}} \sim t_{a(r-1), \delta},$$

con

$$\delta = \frac{\tau_i - \tau_{i'}}{\sqrt{\frac{2\sigma_e^2}{r}}},$$

⁹Cabe aclarar que $c_0 = 0$ porque Γ es un contraste.

y si, adicionalmente, $\tau_i - \tau_{i'} = 0$,

$$\frac{\bar{Y}_{i*} - \bar{Y}_{i'*}}{\sqrt{\frac{2CME}{r}}} \sim t_{a(r-1)}.$$

Por tanto, se rechaza $H_0 : \Gamma = 0$ si

$$\begin{aligned} \left| \frac{\bar{Y}_{i*} - \bar{Y}_{i'*}}{\sqrt{\frac{2CME}{r}}} \right| &> t_{a(r-1), 1-\frac{\alpha}{2}} \\ |\bar{Y}_{i*} - \bar{Y}_{i'*}| &> \sqrt{\frac{2CME}{r}} \times \left(t_{a(r-1), 1-\frac{\alpha}{2}} \right) \end{aligned} \quad (3.26)$$

al lado derecho de la desigualdad en (3.26), es decir, $\sqrt{\frac{2CME}{r}} \left(t_{a(r-1), 1-\frac{\alpha}{2}} \right)$, se le llama **diferencia mínima significativa o LSD**.

6.3.3. Método de Bonferroni

Fisher inicialmente propone el test LSD de Fisher¹⁰ que no protege *FWER*. Posteriormente, propone el **test de Bonferroni**, que consiste en usar el *t*-test (LSD de Fisher) para cada par de medias con un nivel de significancia por comparación de $\frac{\alpha}{c}$, con $c = \frac{a(a-1)}{2}$. Así, la *FWER* es a lo más α . Este test protege el error por familia (*PFER*) (no es una probabilidad) que representa el número esperado de errores en la familia. El test de Bonferroni consiste en rechazar $H_0 : \mu_i = \mu_{i'}, i \neq i'$, si

$$|\bar{y}_{i*} - \bar{y}_{i'*}| > t_{a(r-1), 1-\frac{\alpha}{2c}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)}.$$

Vale resaltar que al usar $\frac{\alpha}{c}$, la región de rechazo es menor (en área) que si se hubiera usado α , por lo que es más exigente (evidencia más fuerte) para rechazar la hipótesis nula.

¹⁰Se rechaza $H_0 : \mu_i = \mu_{i'}, i \neq i'$ si

$$|\bar{Y}_{i*} - \bar{Y}_{i'*}| > t_{a(r-1), 1-\frac{\alpha}{2}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)}.$$

6.3.4. Agrupación de tratamientos

Para realizar la agrupación de tratamientos se debe tener en cuenta:

1. Ordenar las medias de los tratamientos de menor a mayor sin olvidar a qué tratamiento pertenecen.
2. Calcular el punto crítico según el test y según el nivel de significancia a partir del cual se considera que dos efectos de tratamientos son estadísticamente diferentes.
3. Comparar $|\bar{y}_{i*} - \bar{y}_{i'*}|$ con el punto crítico,
 - Si $|\bar{y}_{i*} - \bar{y}_{i'*}|$ es menor que el punto crítico se asocia el tratamiento i y el tratamiento i' al mismo grupo y llevan la misma letra.
 - Si $|\bar{y}_{i*} - \bar{y}_{i'*}|$ es mayor que el punto crítico se asocia el tratamiento i a un grupo distinto al del tratamiento i' y deben llevar letra diferente.
 - El tratamiento con menor media inicia con una letra, por ejemplo la letra a. Al final todos los tratamientos deben tener al menos una letra asociada, de forma tal que tratamientos con la misma letra se considera que tienen efectos estadísticamente iguales. Es importante resaltar que:
 - Un tratamiento puede pertenecer a más de una agrupación.
 - Una agrupación puede estar conformada por uno o por varios tratamientos.
 - Las agrupaciones no necesariamente son disjuntas.
 - Una agrupación no debe contener completamente a otra agrupación.
 - No hay transitividad, es decir, si $A = B$ y $B = C$ NO implica que $A = C$.

Ejemplo 3.23. Continuando con el ejemplo sobre los fertilizantes, ejemplo 3.3, realizar pruebas de comparación múltiple usando el test de Bonferroni y un nivel de significancia de 10 %.

Solución

1. Ordenación de las medias.

Fertilizante (kg/ha)		
N	P_2O_5	Promedio
0	40	1.76
0	0	1.82
75	40	1.92
75	0	1.957
150	40	2.003
150	0	2.09

2. El punto crítico. Para calcular el punto crítico se necesita $c = \frac{a(a-1)}{2}$
 $= \frac{30}{2} = 15$.

$$\begin{aligned}
 t_{a(r-1), 1-\frac{\alpha}{2c}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)} &= t_{6(3-1), 1-\frac{\alpha}{2 \times 15}} \sqrt{CME \left(\frac{2}{3} \right)} \\
 &= t_{12; 1-\frac{0.1}{30}} \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= t_{12; 0.997} \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= 3.278 \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= 0.1521.
 \end{aligned}$$

3. Agrupación de efectos de tratamiento según el punto crítico.

Fertilizante (kg/ha)			
N	P_2O_5	Media	Agrupación
0	40	1.76	a
0	0	1.82	ab
75	40	1.92	bc
75	0	1.957	bcd
150	40	2.003	cd
150	0	2.09	d

Se concluye que hay cuatro agrupaciones de tratamientos con un nivel de significancia del 10 %. Un gráfico con las agrupaciones y el rango por tratamiento se puede ver en la figura 3.6.

Las instrucciones en R para obtener las agrupaciones y la figura 3.6 del ejemplo 3.23 son:

```
### correr instrucciones página 67 hasta obtener ###
### el modelo ajuste1 ###
library(agricolae)
gle=df.residual(ajuste1)#extrae los grados de libertad del error
SCE = deviance(ajuste1) # extrae la suma de cuadrados del error
CME = SCE/gle # calcula el cuadrado medio del error
t=6 # número de tratamientos
c=t*(t-1)/2 # número de diferencias
alpha=0.10 # nivel de significancia
alphaaju=alpha/c # nivel de significancia ajustado
LSD=LSD.test(Altura,fertilizante,gle,CME,alphaaju)
LSD # test de bonferroni
old.par <- par(mai=c(2,1,1,0.1), no.readonly = TRUE)
plot(LSD,main=" ",ylab="Altura (metros)", xlab=" ",
      cex.lab=1.1,las=2)
par(old.par)
```



Diseño completamente aleatorizado

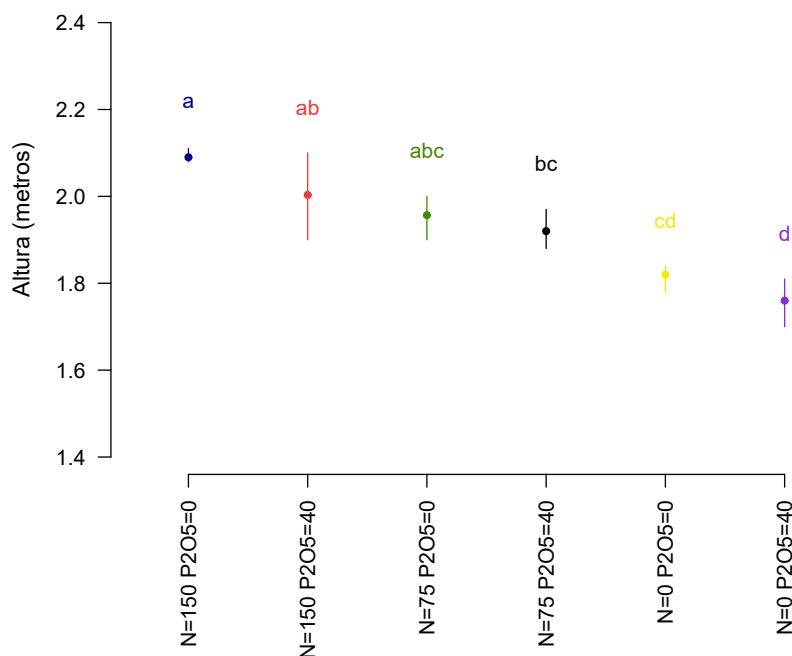


Figura 3.6. Agrupaciones y rango (mínimo y máximo) obtenidas en el ejemplo 3.23

Nota: En el eje de las abscisas, los labels corresponden a los fertilizantes

Ejemplo 3.24. [47] realizó un experimento en la Universidad Nacional de Colombia sobre almidones, en que se estudió, entre otras variables, la velocidad de flujo (g/seg) por el método dinámico que determina la velocidad a través de un orificio. Estudió cuatro almidones: almidón pregelatinizado, almidón de maíz, almidón de ñame criollo y almidón de ñame espino. Se realizaron 10 réplicas por tipo de almidón. Las medias para los almidones son: almidón pregelatinizado: 4.4466, almidón de maíz: 3.369, almidón de ñame criollo: 5.1997, y almidón de ñame espino: 4.827. Un gráfico de puntos por almidón se presenta en la figura 3.7. Realizar pruebas de comparación múltiple usando el test de Bonferroni y un nivel de significancia de 5 %. Se realizó un análisis de varianza en que se obtuvo que $CME = 0.1244$.

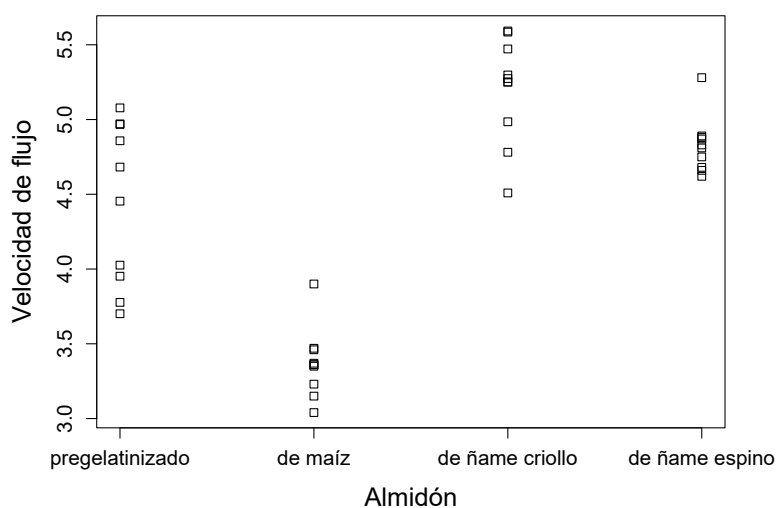


Figura 3.7. Gráfico de puntos para el ejemplo sobre los almidones y la velocidad de flujo (g/seg)

Solución

1. Ordenación de las medias.

Almidón	media
Almidón de maíz	3.369
Almidón pregelatinizado	4.4466
Almidón de ñame espinoso	4.827
Almidón de ñame criollo	5.1997

2. El punto crítico. Para calcular el punto crítico se necesita:

$$c = \frac{a(a-1)}{2} = \frac{4 \times 3}{2} = 6.$$

$$\begin{aligned}
 t_{a(r-1);1-\frac{\alpha}{2c}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)} &= t_{4(10-1);1-\frac{\alpha}{2 \times 6}} \sqrt{CME \left(\frac{2}{10} \right)} \\
 &= t_{36;1-\frac{0.05}{12}} \sqrt{\frac{0.1244}{5}} \\
 &= t_{36;0.9958} \sqrt{\frac{0.1244}{5}} \\
 &= 2.789 \sqrt{\frac{0.1244}{5}} \\
 &= 0.4399.
 \end{aligned}$$

3. Agrupación de efectos de tratamiento según el punto crítico.

Almidón	Media	Agrupación
Almidón de maíz	3.369	a
Almidón pregelatinizado	4.4466	b
Almidón de ñame espino	4.827	bc
Almidón de ñame criollo	5.1997	c

Se concluye que hay tres agrupaciones de tratamientos, en que los almidones tienen estadísticamente el mismo efecto sobre la velocidad de flujo. En particular se destaca el almidón de maíz por su menor velocidad de flujo. El nivel de significancia usado es del 5 %. Un gráfico con las agrupaciones y el rango por tratamiento se puede ver en la figura 3.8.

Las instrucciones en R para obtener las agrupaciones y las figuras 3.7 y 3.8 del ejemplo 3.24 son:

```

library(agricolae)
remove(list=ls())
almidon=gl(4,10)
velocidad=c(4.026,4.454,3.777,3.952,5.078,4.971,4.967,4.682,
            3.701,4.858,3.9,3.47,3.46,3.37,3.15,3.04,3.35,
            3.36,3.36,3.23,4.781,5.251,5.592,5.274,5.25,
            5.472,5.585,4.509,4.985,5.298,4.62,4.68,4.66,
            4.83,4.81,4.89,4.88,4.87,5.28,4.75)

```

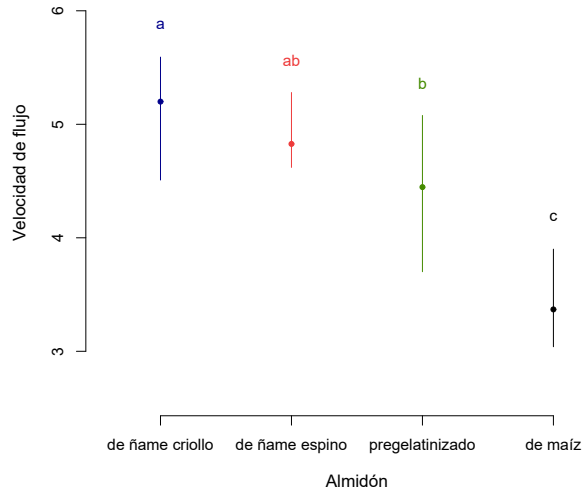


Figura 3.8. Agrupaciones y rango (mínimo y máximo) obtenidas en el ejemplo 3.24

```
almidon=as.factor(almidon)
almidon = factor(almidon,levels = c(1,2,3,4),
                 labels = c("pregelatinizado","de maíz",
                           "de ñame criollo","de ñame espino"))
datos=data.frame(almidon,velocidad)
ajuste1 = aov(velocidad ~ almidon, data=datos)
stripchart(velocidad~almidon,vertical=TRUE,xlab="Almidón",
           ylab="Velocidad de flujo",cex.axis=1.2,cex.lab=1.5)
gle = df.residual(ajuste1); SCE = deviance(ajuste1)
CME = SCE/gle; t=4; c=t*(t-1)/2; alpha=0.05
alphaaju=alpha/c # nivel de significancia ajustado
LSD=LSD.test(velocidad,almidon,gle,CME,alphaaju) # test de
# bonferroni

LSD
plot(LSD,main=" ",xlab="Almidón",ylab="Velocidad de flujo",
     cex.lab=1.1)
```



Nota 3.17. No es necesario pasar primero por el análisis de varianza para hacer comparaciones múltiples.

Otros procedimientos que protejen el $FWER$ son los métodos de Tukey-Kramer, Dunnett y Scheffe, entre otros.

6.3.5. Método de Tukey-Kramer

El método es similar al método LSD de Fisher, cambiando la distribución con la cual se establece el punto crítico para la región de rechazo. El test de Tukey [36], usa la distribución del rango estudentizado y se rechaza $H_0 : \mu_i = \mu_{i'}$, $i \neq i'$ si

$$|\bar{Y}_{i*} - \bar{Y}_{i'*}| > \frac{q_{\alpha; a; a(r-1)}}{\sqrt{2}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)}.$$

Ejemplo 3.25. Continuando con el ejemplo 3.23, sobre los fertilizantes, realizar pruebas de comparación múltiple usando el test de Tukey y un nivel de significancia de 5 %.

Solución

1. El punto crítico es dado por:

$$\begin{aligned} \frac{q_{\alpha; a; a(r-1)}}{\sqrt{2}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)} &= \frac{q_{\alpha; 6; 6(3-1)}}{\sqrt{2}} \sqrt{\frac{0.00323 \times 2}{3}} \\ &= \frac{q_{0.05; 6; 12}}{\sqrt{2}} \sqrt{\frac{0.00323 \times 2}{3}} \\ &= \frac{4.7502}{\sqrt{2}} \sqrt{\frac{0.00323 \times 2}{3}} \\ &= 0.1559. \end{aligned}$$

2. Agrupación de efectos de tratamiento según el punto crítico.

Fertilizante (kg/ha)			
N	P_2O_5	Media	Agrupación
0	40	1.76	a
0	0	1.82	ab
75	40	1.92	bc
75	0	1.957	bcd
150	40	2.003	cd
150	0	2.09	d

Se concluye que hay cuatro agrupaciones de tratamientos con un nivel de significancia del 5 %.

La instrucción en R para el test de Tukey-Kramer es:

```
### correr instrucciones página 67 hasta obtener ###
### el modelo ajuste1 ###
gle=df.residual(ajuste1)
SCE = deviance(ajuste1)
CME = SCE/gle
alpha=0.05
qtukey((1-alpha), 6, df=12, lower.tail=T) # cuantil de la
# distribución del rango estudentizado
HSD=HSD.test(Altura,fertilizante,gle,CME,alpha) # test de tukey
HSD
plot(HSD)
```

Ejemplo 3.26. Continuando con el ejemplo 3.24 sobre comparación de velocidad de flujo al considerar cuatro diferentes almidones. Realizar pruebas de comparación múltiple usando el test de Tukey con un nivel de significancia de 5 %.

Solución

1. El punto crítico es dado por:

$$\begin{aligned} \frac{q_{\alpha;a;(r-1)}}{\sqrt{2}} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)} &= \frac{q_{\alpha;4;4(10-1)}}{\sqrt{2}} \sqrt{\frac{0.1244 \times 2}{10}} \\ &= \frac{q_{0.05;4;36}}{\sqrt{2}} \sqrt{\frac{0.1244}{5}} \\ &= \frac{3.8088}{\sqrt{2}} \sqrt{\frac{0.1244}{5}} \\ &= 0.4248. \end{aligned}$$

2. Agrupación de efectos de tratamiento según el punto crítico.

Almidón	Media	Agrupación
Almidón de maíz	3.369	a
Almidón pregelatinizado	4.4466	b
Almidón de ñame espinoso	4.827	bc
Almidón de ñame criollo	5.1997	c

Se concluye que hay tres agrupaciones de tratamientos, en que los tipos de almidones tienen estadísticamente el mismo efecto sobre la velocidad de flujo. En particular, se destaca el almidón de maíz por su menor velocidad de flujo. El nivel de significancia usado es del 5 %.

6.3.6. Método de Dunnett

Este método es útil cuando el interés es comparar los tratamientos contra uno en particular, como por ejemplo un tratamiento control. Su cálculo es similar al método de LSD, cambiando la distribución con la cual se establece el punto crítico para la región de rechazo. El test de Dunnett [23], usa la tabla de Dunnett y se rechaza $H_0 : \mu_{\text{control}} = \mu_i, i \neq \text{control}$, si:

$$|\bar{Y}_{i*} - \bar{Y}_{i' *} | > d_{\alpha; a-1; a(r-1)} \sqrt{CME \left(\frac{1}{r_{\text{control}}} + \frac{1}{r_i} \right)}.$$

Ejemplo 3.27. Continuando con el ejemplo 3.23 sobre los fertilizantes. Realizar pruebas de comparación múltiple usando el test de Dunnett en que el fertilizante $N = 0$, $P_2O_5 = 0 \text{ kg/ha}$ es el tratamiento control, con un nivel de significancia de 5 %.

Solución

1. El punto crítico es dado por:

$$\begin{aligned}
 d_{\alpha;a-1;a(r-1)} \sqrt{CME \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right)} &= d_{\alpha;5;6(3-1)} \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= d_{0.05;5;12} \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= 2.90 \sqrt{\frac{0.00323 \times 2}{3}} \\
 &= 0.1346.
 \end{aligned}$$

2. Agrupación de efectos de tratamiento según el punto crítico.

Fertilizantes	Diferencia entre medias	Diferentes?
$N = 75, P_2O_5 = 0$ versus control	0.1367	Sí
$N = 150, P_2O_5 = 0$ versus control	0.27	Sí
$N = 0, P_2O_5 = 40$ versus control	-0.06	No
$N = 75, P_2O_5 = 40$ versus control	0.1	No
$N = 150, P_2O_5 = 40$ versus control	0.1833	Sí

Se concluye que los fertilizantes $N = 0, P_2O_5 = 40$ y $N = 75, P_2O_5 = 40$ tienen un efecto similar al “control” en la altura de la planta de colza, con un nivel de significancia del 5 %. Un gráfico con los intervalos de confianza para las comparaciones realizadas se presenta en la figura 3.9.

Diseño completamente aleatorizado

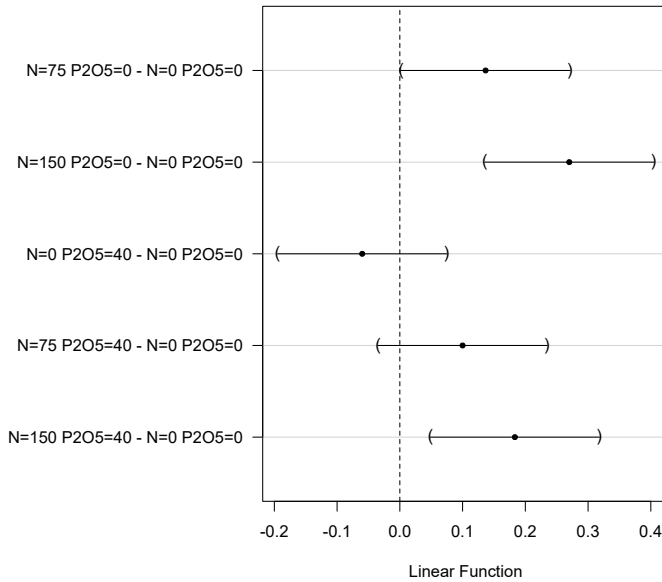


Figura 3.9. Intervalos de confianza. Método Dunnett - ejemplo 3.27

Las instrucciones en R para obtener las comparaciones de tratamientos y la figura 3.9 del ejemplo 3.27 son:

```
### correr instrucciones página 67 ###  
library(multcomp)  
Dun = glht(ajuste1, linfct=mcp(fertilizante="Dunnett"),  
           alternative="two.side")  
summary(Dun)  
old.par <- par(mai=c(1,2.7,1,0.2), no.readonly = TRUE)  
plot(Dun)  
par(old.par)
```

Nota 3.18. Para comparar con un tratamiento control, se podría usar el test de Bonferroni con $c = a - 1$.

7. Polinomios ortogonales

Cuando los a niveles de un factor son cantidades $x_1 = 0, x_2 = d, x_3 = 2d, \dots, x_a = (a-1)d$ (equidistantes), puede ser más útil encontrar tendencias que hacer comparaciones pareadas.

Sin pérdida de generalidad se puede tomar a $d = 1$. Sean:

$$z_i = x_i - \frac{a-1}{2},$$

los niveles centrados. Note que $\sum_{i=1}^a z_i = 0$.

Veamos:

$$\begin{aligned} \sum_{i=1}^a z_i &= \sum_{i=1}^a \left(x_i - \frac{a-1}{2} \right) = \sum_{i=1}^a x_i - \sum_{i=1}^a \frac{a-1}{2} \\ &= \sum_{i=1}^a (i-1)d - \frac{1(a-1)}{2} = d \sum_{i=1}^a i - da - \frac{a(a-1)}{2} \\ &= \frac{da(a+1)}{2} - da - \frac{a(a-1)}{2} = \underbrace{\frac{da(a+1) - 2da - a(a-1)}{2}}_{d=1} = 0. \end{aligned}$$

✓

Ahora, suponga el modelo lineal simple:

$$Y_{ij} = \mu + \alpha z_i + e_{ij} \quad i = 1, \dots, a \text{ y } j = 1, \dots, r, \quad (3.27)$$

con Y_{ij} siendo la respuesta de la unidad experimental j en el tratamiento i , μ la media general y α el coeficiente de regresión. El estimador de α es:

$$\hat{\alpha} = \frac{\sum_{i=1}^a \sum_{j=1}^r Y_{ij} z_i - ar \bar{Y}_{**}}{\sum_{i=1}^a \sum_{j=1}^r z_i^2 - nr^2} = \frac{\sum_{i=1}^a \sum_{j=1}^r Y_{ij} z_i}{\sum_{i=1}^a \sum_{j=1}^r z_i^2} = \sum_{i=1}^a \left[r \bar{Y}_{i*} \frac{z_i}{\left(\sum_{i=1}^a \sum_{j=1}^r z_i^2 \right)} \right],$$

con $\bar{Y}_{**} = \frac{1}{ar} \sum_{i=1}^a \sum_{j=1}^r Y_{ij}$ y $\bar{Y}_{i*} = \frac{1}{r} \sum_{j=1}^r Y_{ij}$. Note que $\hat{\alpha}$ corresponde a una medida del crecimiento (o decrecimiento) debido al aumento en el nivel del factor en d unidades.

Sin embargo, si se requiere ajustar un polinomio de mayor grado, en vez de ajustar un polinomio de grado dos y luego uno de grado tres, etc., una alternativa es usar polinomios ortogonales que permite probar el grado del polinomio en un solo ajuste. En este caso $\bar{Y}_{i*}$ se puede escribir como [24] y [52]:

$$\bar{Y}_{i*} = \sum_{l=0}^{a-1} \gamma_l P_l(z_i) + \text{error}, \quad (3.28)$$

donde $P_l(z_i)$ son polinomios con l ($l = 0, 1, \dots, a-1$) el grado del polinomio, tal que:

$$\begin{aligned} P_0(z_i) &= 1 \\ \sum_{i=1}^a P_l(z_i) &= 0, \quad l = 1, 2, a-1. \\ \sum_{i=1}^a P_l(z_i) P_{l'}(z_i) &= 0. \end{aligned}$$

En particular, si los polinomios están normalizados se tiene $\sum_{i=1}^a [P_l(z_i)]^2 = 1$, $l = 1, 2, \dots, a-1$. Los coeficientes de estos polinomios se pueden obtener en R usando, por ejemplo:

```
options(scipen=999)
a=3 # número de tratamientos
g=2 # grado del polinomio
polinom=round(poly(1:a,g),4) # da los polinomios ortogonales
polinom
norm(polinom[,1], type="2")
norm(polinom[,2], type="2")
#####
      i      P_1      P_2
      1      2
[1,] -0.7071  0.4082
[2,]  0.0000 -0.8165
[3,]  0.7071  0.4082
#####
polinom=round(poly(1:5,4),4)
```

#####

i	P_1	P_2	P_3	P_4
	1	2	3	4
[1,]	-0.6325	0.5345	-0.3162	0.1195
[2,]	-0.3162	-0.2673	0.6325	-0.4781
[3,]	0.0000	-0.5345	0.0000	0.7171
[4,]	0.3162	-0.2673	-0.6325	-0.4781
[5,]	0.6325	0.5345	0.3162	0.1195

Ejemplo 3.28. Sea $a = 3$, presente $\bar{Y}_{i*}$ según (3.28).

Solución Se tiene que:

$$\begin{aligned}\bar{Y}_{i*} &= \sum_{l=0}^{a-1} \gamma_l P_l(z_i) + e_i \\ \bar{Y}_{1*} &= \gamma_0 P_0(z_1) + \gamma_1 P_1(z_1) + \gamma_2 P_2(z_1) + e_1 \\ &= \gamma_0 - 0.7071\gamma_1 + 0.4082\gamma_2 + e_1 \\ \bar{Y}_{2*} &= \gamma_0 P_0(z_2) + \gamma_1 P_1(z_2) + \gamma_2 P_2(z_2) + e_2 \\ &= \gamma_0 - 0.8165\gamma_2 + e_2 \\ \bar{Y}_{3*} &= \gamma_0 P_0(z_3) + \gamma_1 P_1(z_3) + \gamma_2 P_2(z_3) + e_3 \\ &= \gamma_0 + 0.7071\gamma_1 + 0.4082\gamma_2 + e_3,\end{aligned}$$

en que:

γ_1 : es el coeficiente asociado a la medida de tendencia lineal.

γ_2 : es el coeficiente asociado a la medida de tendencia cuadrática.

Nota 3.19. La estimación de $\hat{\gamma}_l$ es **independiente** del número de polinomios incluidos.

Matricialmente, el modelo (3.28) se puede escribir como $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$ con:

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1r} \\ Y_{2r} \\ \vdots \\ Y_{a1} \\ Y_{a2} \\ \vdots \\ Y_{ar} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} P_0(z_1) & P_1(z_1) & \dots & P_{a-1}(z_1) \\ P_0(z_1) & P_1(z_1) & \dots & P_{a-1}(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(z_1) & P_1(z_1) & \dots & P_{a-1}(z_1) \\ P_0(z_2) & P_1(z_2) & \dots & P_{a-1}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(z_a) & P_1(z_a) & \dots & P_{a-1}(z_a) \\ P_0(z_a) & P_1(z_a) & \dots & P_{a-1}(z_a) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(z_a) & P_1(z_a) & \dots & P_{a-1}(z_a) \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{a-1} \end{pmatrix}, \quad (3.29)$$

en que $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$. Note que:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} r \sum_{i=1}^a [P_0(z_i)]^2 & r \sum_{i=1}^a P_0(z_i) P_1(z_i) & \dots & r \sum_{i=1}^a P_0(z_i) P_{a-1}(z_i) \\ r \sum_{i=1}^a P_1(z_i) P_0(z_i) & r \sum_{i=1}^a [P_1(z_i)]^2 & \dots & r \sum_{i=1}^a P_1(z_i) P_{a-1}(z_i) \\ \vdots & \vdots & \ddots & \vdots \\ r \sum_{i=1}^a P_{a-1}(z_i) P_0(z_i) & r \sum_{i=1}^a P_{a-1}(z_i) P_1(z_i) & \dots & r \sum_{i=1}^a [P_{a-1}(z_i)]^2 \end{pmatrix} \\ &= r \begin{pmatrix} a & \mathbf{0}_{1 \times (a-1)} \\ \mathbf{0}_{(a-1) \times 1} & \mathbf{I}_{a-1} \end{pmatrix} \end{aligned}$$

y

$$\mathbf{X}^\top \mathbf{Y} = \begin{pmatrix} \sum_{i=1}^a \sum_{j=1}^r P_0(z_i) Y_{ij} \\ \sum_{i=1}^a \sum_{j=1}^r P_1(z_i) Y_{ij} \\ \vdots \\ \sum_{i=1}^a \sum_{j=1}^r P_{a-1}(z_i) Y_{ij} \end{pmatrix} = r \begin{pmatrix} a \bar{Y}_{**} \\ \sum_{i=1}^a P_1(z_i) \bar{Y}_{i*} \\ \vdots \\ \sum_{i=1}^a P_{a-1}(z_i) \bar{Y}_{i*} \end{pmatrix}.$$

Entonces:

$$\begin{aligned}\hat{\beta} &= (\hat{\gamma}_0 \quad \hat{\gamma}_1 \quad \cdots \quad \hat{\gamma}_{a-1})^\top \\ &= \begin{matrix} \text{\textit{r}}' \\ \text{\textit{r}}' \end{matrix} \begin{pmatrix} \frac{1}{a} & \mathbf{0}_{1 \times (a-1)} \\ \mathbf{0}_{(a-1) \times 1} & \mathbf{I}_{a-1} \end{pmatrix} \begin{pmatrix} a\bar{Y}_{**} \\ \sum_{i=1}^a P_1(z_i)\bar{Y}_{i*} \\ \vdots \\ \sum_{i=1}^a P_{a-1}(z_i)\bar{Y}_{i*} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{**} \\ \sum_{i=1}^a P_1(z_i)\bar{Y}_{i*} \\ \vdots \\ \sum_{i=1}^a P_{a-1}(z_i)\bar{Y}_{i*} \end{pmatrix}.\end{aligned}$$

Por tanto,

$$\begin{aligned}\hat{\gamma}_0 &= \bar{Y}_{**} \\ \hat{\gamma}_l &= \sum_{i=1}^a P_l(z_i)\bar{Y}_{i*}, \quad l = 1, 2, \dots, a-1.\end{aligned}$$

Como $\mathbb{V}(\bar{Y}_{i*}) = \frac{1}{r^2} \sum_{j=1}^r \sigma_e^2 = \frac{\sigma_e^2}{r}$, entonces:

$$\begin{aligned}\mathbb{V}(\hat{\gamma}_0) &= \mathbb{V}(\bar{Y}_{**}) = \mathbb{V}\left(\frac{1}{a} \sum_{i=1}^a \bar{Y}_{i*}\right) = \frac{1}{a^2} \sum_{i=1}^a \mathbb{V}(\bar{Y}_{i*}) = \frac{a\sigma_e^2}{ra^2} = \frac{\sigma_e^2}{ar} \\ \mathbb{V}(\hat{\gamma}_l) &= \mathbb{V}\left(\sum_{i=1}^a P_l(z_i)\bar{Y}_{i*}\right) = \sum_{i=1}^a [P_l(z_i)]^2 \mathbb{V}(\bar{Y}_{i*}) = \frac{\sigma_e^2}{r} \sum_{i=1}^a [P_l(z_i)]^2 \xrightarrow{1} \\ &= \frac{\sigma_e^2}{r}, \quad l = 1, 2, \dots, a-1.\end{aligned}$$

La respectiva suma de cuadrados asociada a la hipótesis:

$$H_0 : \gamma_l = 0, \quad l = 1, 2, \dots, a-1,$$

que matricialmente se puede escribir como:

$$\begin{aligned}H_0 : \gamma_l &= \underbrace{\begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}}_{\lambda_l^\top} \underbrace{\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{a-1} \end{pmatrix}}_{\beta} \\ &= 0, \quad l = 1, 2, \dots, a-1,\end{aligned}$$

en que λ_l es un vector con entradas cero excepto en la posición $l + 1$, es dada por, ver nota 2.6:

$$\begin{aligned} SC(\hat{\gamma}_0) &= (\lambda_0^\top \hat{\beta})^\top \left(\lambda_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} \lambda_0 \right)^{-1} (\lambda_0^\top \hat{\beta}) = \hat{\gamma}_0 \left(\frac{1}{ar} \right)^{-1} \hat{\gamma}_0 = ar \hat{\gamma}_0^2 \\ &= ar \bar{Y}_{**}^2 = n \bar{Y}_{**}^2 \\ SC(\hat{\gamma}_l) &= (\lambda_l^\top \hat{\beta})^\top \left(\lambda_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \lambda_l \right)^{-1} (\lambda_l^\top \hat{\beta}) = \hat{\gamma}_l \left(\frac{1}{r} \right)^{-1} (\hat{\gamma}_l) = r \hat{\gamma}_l^2 \\ &= r \left(\sum_{i=1}^a P_l(z_i) \bar{Y}_{i*} \right)^2, \quad l = 1, 2, \dots, a-1. \end{aligned} \quad (3.30)$$

La distribución de las sumas de cuadrados y cuadrados medios se puede ver en el anexo 8.

Para encontrar la tendencia del efecto de tratamientos se ajusta inicialmente un modelo de orden bajo y se evalúa su falta de ajuste. Si es necesario, se añaden términos de orden superior. Suponga que se ajusta el modelo:

$$\bar{Y}_{i*} = \sum_{l=0}^q \gamma_l P_l(z_i) + \text{error},$$

con $q < a - 1$. En la tabla 3.19 se presenta el análisis de varianza.

Para probar la falta de ajuste se usa el estadístico:

$$F_c = \frac{CM(FAj)}{CME} \sim F(a - q - 1, a(r - 1)). \quad (3.31)$$

Si el valor de F es significativo según un nivel de significancia dado, implica que hay falta de ajuste y se deben adicionar términos al modelo hasta que el término asociado a la falta de ajuste no sea significativo.

Ejemplo 3.29. [28] realizaron un experimento en la Universidad Nacional de Colombia sobre la utilización de probióticos en la fase de levante de la mojarra roja. En el experimento se consideraron cuatro tratamientos incluido el control, cada uno con cinco réplicas que corresponden a cinco acuarios con densidad de cinco peces por acuario. Se medían, entre otros, los pesos de los peces a los 45 días. En la tabla 3.20 se presentan los pesos promedio por acuario y en la figura 3.10, un gráfico de puntos por nivel de probiótico.

Tabla 3.19. Análisis de varianza para un diseño completamente aleatorizado usando polinomios ortogonales

Causa de variación	Grados de libertad	Suma de cuadrados
Tratamientos	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{y}^2$
γ_1	1	$SC(\gamma_1)$
γ_2	1	$SC(\gamma_2)$
$\vdots$	$\vdots$	$\vdots$
γ_q	1	$SC(\gamma_q)$
Falta Ajuste (FAj)	$a - q - 1$	$SCTr - \sum_{i=1}^q SC(\gamma_i) = SC(FAj)$
Error	$a(r - 1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$
Total	$ar - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$

Tabla 3.20. Pesos (g) promedio de las mojaras por acuario a los 45 días

Alimento concentrado con probiótico (gramos de probiótico /kg)	Acuario				
	1	2	3	4	5
0	19.32	21.86	25.80	22.18	18.08
2	19.64	24.36	25.10	22.98	21.60
4	26.78	30.66	29.48	33.14	26.00
6	32.40	33.80	30.96	25.76	37.14

Fuente: [28].

Teniendo en cuenta que el diseño tiene submuestreo, ya que los tratamientos se asignaban a los acuarios, los acuarios son unidades experimentales, y los peces corresponden a unidades observacionales.

1. *Utilice los promedios de los pesos de los peces en cada acuario como el dato del acuario, ignorando el submuestreo. Realice el análisis de varianza sin considerar los polinomios ortogonales y luego considerando los polinomios ortogonales. ¿un*

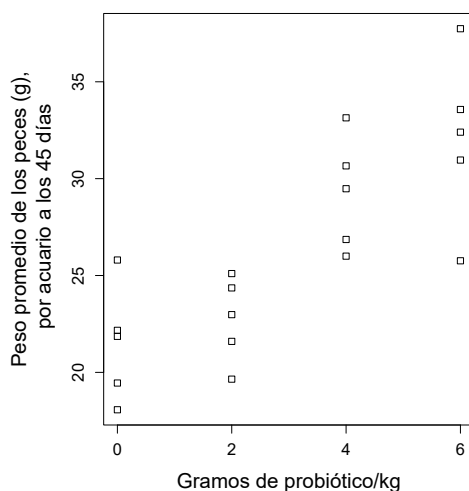


Figura 3.10. Gráfico de puntos para el ejemplo sobre los concentrados con probiótico y peso promedio de los peces por acuario a los 45 días

ajuste lineal es adecuado?, ¿es necesario hacer un ajuste de un polinomio de grado superior?

2. Realice el punto anterior, teniendo en cuenta que el diseño tiene submuestreo.

Solución

1. En la tabla 3.21 se presenta el análisis de varianza para un diseño completamente aleatorizado, en que se concluye que hay efecto de tratamiento con un P-valor igual a 0.00015. Al realizar el test de Tukey-Kramer se encuentran dos agrupaciones de tratamientos, una correspondiente a los dos niveles más bajos de probiótico (0 y 2 g de probiótico/kg) y al otra con los dos niveles más altos de probiótico (4 y 6 g de probiótico/kg). De otro lado, se va a ajustar el modelo lineal simple dado en (3.27) para el cual en la tabla 3.22 se presentan los promedios de los pesos (en gramos) a los 45 días de los tratamiento, junto con sus respectivos x_i y z_i .

Tabla 3.21. Análisis de varianza para los datos del ejemplo 3.29, sin considerar el submuestreo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Probióticos	3	386.8	128.9	13.03	0.000146
Error	16	158.4	9.9		
Total	19	545.2			

Tabla 3.22. Promedio de los pesos (g) de mojarra roja a los 45 días (fase de levante), en el experimento de los concentrados con probióticos

Cantidad Probiótico	x_i	z_i	y_{i*}	r_i	$\bar{y}_{i*}$
0	0	-1.5	107.24	5	21.448
2	1	-0.5	113.68	5	22.736
4	2	0.5	146.06	5	29.212
6	3	1.5	160.06	5	32.012

$$\sum_{i=1}^a z_i^2 = \sum_{i=1}^4 z_i^2 = 5.$$

Se tiene que:

$$a = 4,$$

$$z_i = x_i - \frac{4-1}{2} = x_i - \frac{3}{2}$$

$$\hat{\alpha} = \sum_{i=1}^4 \left[r \bar{Y}_{i*} \frac{z_i}{\left(\sum_{i=1}^a \sum_{j=1}^r z_i^2 \right)} \right] = \frac{5 \times 19.084}{5 \times 5} = 3.8168.$$

Por tanto, cuando la cantidad de probiótico aumentó en 2 gramos por kilogramo, el peso promedio de los peces en los acuarios aumentó a los

45 días en 3,8168 gramos. La pregunta es si es suficiente realizar un ajuste lineal o si es necesario considerar un ajuste de un polinomio de grado mayor. Para esto se pueden usar los polinomios ortogonales.

Sea el modelo:

$$\begin{aligned}\bar{Y}_{i*} &= \sum_{l=0}^3 \gamma_l P_l(z_i) + e_i = \gamma_0 P_0(z_i) + \gamma_1 P_1(z_i) + \gamma_2 P_2(z_i) + \\ &\quad \gamma_3 P_3(z_i) + e_i \\ \bar{Y}_{1*} &= \gamma_0 - 0.6708\gamma_1 + 0.5\gamma_2 - 0.2236\gamma_3 + e_1 \\ \bar{Y}_{2*} &= \gamma_0 - 0.2236\gamma_1 - 0.5\gamma_2 + 0.6708\gamma_3 + e_2 \\ \bar{Y}_{3*} &= \gamma_0 + 0.2236\gamma_1 - 0.5\gamma_2 - 0.6708\gamma_3 + e_3 \\ \bar{Y}_{4*} &= \gamma_0 + 0.6708\gamma_1 + 0.5\gamma_2 + 0.2236\gamma_3 + e_4.\end{aligned}$$

Las respectivas estimativas de γ_0 , γ_1 , γ_2 y γ_3 , junto con sus sumas de cuadrados son:

$$\begin{aligned} \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \end{pmatrix} &= \begin{pmatrix} \bar{y}_{**} \\ \sum_{i=1}^4 P_1(z_i) \bar{Y}_{i*} \\ \sum_{i=1}^4 P_2(z_i) \bar{Y}_{i*} \\ \sum_{i=1}^4 P_3(z_i) \bar{Y}_{i*} \end{pmatrix} \\ &= \begin{pmatrix} \bar{y}_{**} \\ -0.6708\bar{y}_{1*} - 0.2236\bar{y}_{2*} + 0.2236\bar{y}_{3*} + 0.6708\bar{y}_{4*} \\ 0.5\bar{y}_{1*} - 0.5\bar{y}_{2*} - 0.5\bar{y}_{3*} + 0.5\bar{y}_{4*} \\ -0.2236\bar{y}_{1*} + 0.6708\bar{y}_{2*} - 0.6708\bar{y}_{3*} + 0.2236\bar{y}_{4*} \end{pmatrix} \\ &= \begin{pmatrix} 26.350 \\ 8.534 \\ 0.756 \\ -1.982 \end{pmatrix} \end{aligned}$$

$$SC(\hat{\gamma}_0) = ra\bar{y}_{**}^2 = 13\,888.558$$

$$SC(\hat{\gamma}_1) = r(\hat{\gamma}_1)^2 = 5(8.572)^2 = 364.177$$

$$SC(\hat{\gamma}_2) = r(\hat{\gamma}_2)^2 = 5(0.795)^2 = 2.858$$

$$SC(\hat{\gamma}_3) = r(\hat{\gamma}_3)^2 = 5(-1.980)^2 = 19.641$$

$$F_{\hat{\gamma}_1} = \frac{CM(\hat{\gamma}_1)}{CME} = \frac{SC(\hat{\gamma}_1)}{CME} = \frac{364.177}{9.9} = 36.7855$$

$$\text{signo}(\hat{\gamma}_1)\sqrt{F_{\hat{\gamma}_1}} = 6.065 \quad \text{t-valor en la salida de R.}$$

$$F_{\hat{\gamma}_2} = \frac{CM(\hat{\gamma}_2)}{CME} = \frac{SC(\hat{\gamma}_2)}{CME} = \frac{2.858}{9.9} = 0.289$$

$$\text{signo}(\hat{\gamma}_2)\sqrt{F_{\hat{\gamma}_2}} = 0.537 \quad \text{t-valor en la salida de R.}$$

$$F_{\hat{\gamma}_3} = \frac{CM(\hat{\gamma}_3)}{CME} = \frac{SC(\hat{\gamma}_3)}{CME} = \frac{19.641}{9.9} = 1.984$$

$$\text{signo}(\hat{\gamma}_3)\sqrt{F_{\hat{\gamma}_3}} = -1.408 \quad \text{t-valor en la salida de R.}$$

Diseño completamente aleatorizado

En la tabla 3.23 se presenta el análisis de varianza, en que se observa que el término cuadrático y cúbico no son significativos. Por tanto, el ajuste lineal es adecuado, ver figura 3.11.

Tabla 3.23. Análisis de varianza para los datos del ejemplo 3.29 usando polinomios ortogonales

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	3	386.8	128.9	13.03	0.000146
Término lineal ($\hat{\gamma}_1$)	1	364.177	364.177	36.7855	0.00002
Término cuadrático ($\hat{\gamma}_2$)	1	2.858	2.858	0.289	0.5983
Término cúbico ($\hat{\gamma}_3$)	1	19.641	19.641	1.984	0.1781
Error	16	158.4	9.9		
Total	19	545.2			

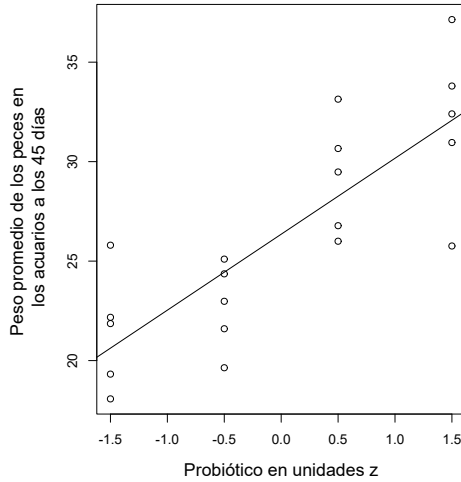


Figura 3.11. Gráfico de dispersión, nivel de probiótico en unidades z_i versus peso promedio de los acuarios a los 45 días, y línea ajustada bajo el modelo de regresión lineal $\hat{y} = 26.380 + 3.834z$ para los datos del ejemplo 3.29

Las instrucciones en R para el modelo ajustado en el ejemplo 3.29, sin considerar el submuestreo son:

```
library(agricolae); library(readxl); library(varhandle);
library(dplyr)
remove(list=ls())
mojarra = read_excel("BasesDatosDE.xlsx",sheet="mojarra")
attach(mojarra)
names(mojarra)=make.names(names(mojarra))
mojarra$probiotico = factor(mojarra$probiotico,
                           levels = c(1,2,3,4),labels = c(0,2,4,6))
mojarra$Acuario=factor(Acuario)
m.Acu = mojarra %>% group_by(probiotico,Acuario) %>%
  summarise(peso45 = mean(Peso.45,na.rm=T))
m.Acu
options(pillar.sigfig = 4); m.Acu
old.par <- par(mai=c(1,1.5,1,1), no.readonly = TRUE)

stripchart(m.Acu$peso45~m.Acu$probiotico,vertical=T,
```

Diseño completamente aleatorizado

```
      xlab="Gramos de probiótico/kg",
      ylab="Peso promedio de los peces (g),
      \n por acuario a los 45 días",cex.lab=1.3)
par(old.par)
ajuste1 = aov(peso45 ~ probiotico,data=m.Acu)
summary(ajuste1)
gle = df.residual(ajuste1); SCE = deviance(ajuste1)
CME = SCE/gle
HSD=HSD.test(m.Acu$peso45,m.Acu$probiotico,gle,CME,0.05)
HSD
## ajustando la regresión lineal simple ###
probiotico.num <- unfactor(m.Acu$probiotico)/2 # para d=1
probiotico.num <- probiotico.num-mean(probiotico.num) # z_i
ajuste2=lm(m.Acu$peso45~probiotico.num); ajuste2
## gráfico
old.par <- par(mai=c(1,1.5,1,1), no.readonly = TRUE)
plot(m.Acu$peso45~probiotico.num,
      xlab="Probiótico en unidades z",
      ylab="Peso promedio de los peces en \n los acuarios
      a los 45 días",
      cex.lab=1.3); abline(coef(ajuste2))
par(old.par)
## polinomios ortogonales
contrasts(m.Acu$probiotico)=
      contr.poly(levels(m.Acu$probiotico))
      # se usa polinomios ortogonales normalizados
contrasts(m.Acu$probiotico) # imprime el conjunto de
# polinomios ortogonales normalizados usados en el análisis
ajuste3 = lm(peso45~probiotico,data=m.Acu) # ajusta un mo-
# delo de análisis de varianza para polimonios ortogonales
summary(ajuste3) # imprime en pantalla-análisis de varianza
```

2. Teniendo en cuenta el submuestreo, en la figura 3.12 se presenta un gráfico de puntos de los pesos de los peces por nivel de probiótico. En la tabla 3.24 se presenta el análisis de varianza para un diseño completamente aleatorizado con submuestreo, en que se concluye que hay efecto de tratamiento con un P-valor igual a 0.00014. Al realizar el test de

Tukey-Kramer se encuentran las mismas agrupaciones encontradas en el primer ítem.

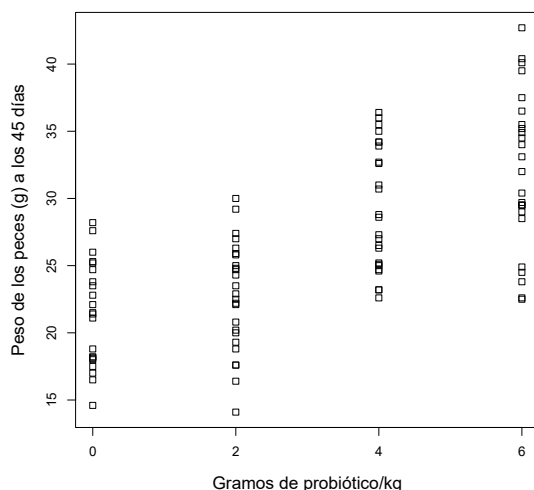


Figura 3.12. Gráfico de puntos para el ejemplo sobre los concentrados con probiótico y peso de los peces a los 45 días

Tabla 3.24. Análisis de varianza para los datos del ejemplo 3.29 considerando el submuestreo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Probióticos	3	1877.0	625.7	13.16	0.000138
Error Experimental (acuarios)	16	760.9	47.6		
Error Observ. (peces)	77	1254.0	16.28		
Total	96	3891.9			

De otro lado, se van a usar polinomios ortogonales para identificar si un ajuste lineal es suficiente para establecer una tendencia o si es necesario ajustar un polinomio de grado mayor. Pero antes de iniciar con

el ajuste, vale resaltar que hay 3 datos faltantes, por lo cual el modelo es desbalanceado y se perdería la ortogonalidad. Se va a usar el promedio dentro de cada acuario en que faltan datos, como una imputación de los datos faltantes. Vale resaltar que la imputación afecta la varianza, lo cual debería tenerse en cuenta en el análisis, sin embargo, en este ejemplo no se tendrá en cuenta dicha afectación.

Ya con los datos completos, el modelo es dado por:

$$\mathbf{Y}_{ar'n^* \times 1} = \mathbf{X}_{ar'n^* \times a} \boldsymbol{\beta}_{a \times 1} + \mathbf{Z}_{ar'n^* \times ar'} \boldsymbol{\varepsilon}_{ar' \times 1} + \boldsymbol{\eta}_{ar'n^* \times 1}, \quad (3.32)$$

con $\mathbf{X}$ y $\boldsymbol{\beta}$ similar a los dados en (3.29), teniendo en cuenta los respectivos cambios en el número de filas de $\mathbf{X}$; y $\mathbf{Z}$, $\boldsymbol{\varepsilon}$ y $\boldsymbol{\eta}$ dados en (3.11).

Para calcular el vector de parámetros estimados, se tiene que:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 25 \end{pmatrix} \\ (\mathbf{X}^\top \mathbf{X})^{-1} &= \begin{pmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0.04 \end{pmatrix} \\ \mathbf{X}^\top \mathbf{Y} &= \begin{pmatrix} 2635.1500 \\ 213.3991 \\ 18.8750 \\ -49.5401 \end{pmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \end{pmatrix} = \begin{pmatrix} 26.3515 \\ 8.5360 \\ 0.7550 \\ -1.9816 \end{pmatrix}. \end{aligned}$$

Note que $\mathbf{X}^\top \mathbf{X}$ es una matriz diagonal porque se imputaron los tres datos faltantes.

En cuanto a las sumas de cuadrados, se usa:

$$SC(\hat{\gamma}_l) = (\boldsymbol{\lambda}_l^\top \hat{\boldsymbol{\beta}})^\top \left(\boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \right)^{-1} (\boldsymbol{\lambda}_l^\top \hat{\boldsymbol{\beta}}).$$

En la tabla 3.25 se presenta el análisis de varianza, en el que se observa que el término cuadrático y cúbico no son significativos, siendo el ajuste lineal adecuado.

Tabla 3.25. Análisis de varianza para los datos del ejemplo 3.29 usando polinomios ortogonales y submuestreo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	3	1934	644.7	13.03	0.000146
Término lineal ($\hat{\gamma}_1$)	1	1821.568	1821.568	36.8058	0.000016
Término cuadrático ($\hat{\gamma}_2$)	1	14.251	14.251	0.2879	0.5989
Término cúbico ($\hat{\gamma}_3$)	1	98.169	98.169	1.9836	0.1782
Error Exper. (acuario)	16	791.9	49.5		
Error Observ.(peces)	80	1254	15.67		
Total	99	3979.9			

El modelo (3.32) corresponde a un modelo de *efectos mixtos* [57, p. 138], por ejemplo, con:

$$\begin{aligned}
 \mathbb{E}(\boldsymbol{\varepsilon}) &= \mathbf{0}_{ar' \times 1} & \mathbb{V}(\boldsymbol{\varepsilon}) &= \sigma_{\varepsilon}^2 \mathbf{I}_{ar'} \\
 \mathbb{E}(\boldsymbol{\eta}) &= \mathbf{0}_{ar'n^* \times 1} & \mathbb{V}(\boldsymbol{\eta}) &= \sigma_{\eta}^2 \mathbf{I}_{ar'n^*} \\
 \mathbb{E}(\mathbf{Y}|\boldsymbol{\varepsilon}) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} & \mathbb{V}(\mathbf{Y}|\boldsymbol{\varepsilon}) &= \sigma_{\eta}^2 \mathbf{I}_{ar'n^*} \\
 \mathbb{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) &= \mathbf{0} \\
 \mathbb{E}(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} & \mathbb{V}(\mathbf{Y}) &= \mathbb{V}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta}) \\
 & & &= \mathbb{V}(\mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta}) \\
 & & &= \sigma_{\varepsilon}^2 \mathbf{Z}\mathbf{Z}^{\top} + \sigma_{\eta}^2 \mathbf{I}_{ar'n^*}.
 \end{aligned}$$

Luego el ajuste del modelo se puede realizar vía máxima verosimilitud restringida [57, p. 249], sin necesidad de imputar los datos faltantes.

Diseño completamente aleatorizado

Las instrucciones en R usadas para el modelo ajustado en el ejemplo 3.29, considerando submuestreo, son:

```
stripchart(mojarra$Peso.45~mojarra$probiotico,vertical=T,
  xlab="Gramos de probiótico/kg",
  ylab="Peso de los peces en (g) a los 45 días",
  cex.lab=1.3)
ajuste4=aov(Peso.45~probiotico+Error(Acuario),data=mojarra)
summary(ajuste4)
gle = df.residual(ajuste4$Acuario)
SCE = deviance(ajuste4$Acuario); CME = SCE/gle
HSD=HSD.test(mojarra$Peso.45,mojarra$probiotico,gle,CME,
  0.05)
HSD # agrupaciones Tukey
##### usando polinomios ortogonales ###
mojarra %>% group_by(probiotico,Acuario)%>%
  summarise(peces = sum(!is.na(Peso.45))) # observar
  # desbalanceamiento imputando los datos faltantes
  # con el promedio del acuario
moj.s <- mojarra %>% group_by(probiotico,Acuario) %>%
  mutate(Peso.45=ifelse(is.na(Peso.45),mean(Peso.45,
    na.rm=TRUE),Peso.45))
contrasts(moj.s$probiotico)=
  contr.poly(levels(moj.s$probiotico))
X=model.matrix(~moj.s$probiotico); Y=moj.s$Peso.45
round(t(X)%*%X,4); round(solve(t(X)%*%X),4)
round(t(X)%*%Y,4)
beta=solve(t(X)%*%X)%*%t(X)%*%Y; round(beta,4)
ajuste5 = aov(Peso.45 ~ probiotico+Error(Acuario),
  data=moj.s)
gle = df.residual(ajuste5$Acuario)
SCE = deviance(ajuste5$Acuario); CME = SCE/gle
c1=c(1,0,0,0);c2=c(0,1,0,0);c3=c(0,0,1,0);c4=c(0,0,0,1);
lambda=as.matrix(c4) # usar el respectivo c1 o c2 o
# c3 o c4
SC=((t(lambda)%*%beta)^2)/(t(lambda)%*%solve(t(X)%*%X)%*%
  lambda)
```

```
SC ; SC/CME; sqrt(SC/CME) # t-valor
# ajustando el modelo mixto
library(nlme)
ajuste6=lme(Peso.45 ~ probiotico, random=~1|Acuario,
            data=moj.s)
summary(ajuste6) # ajuste modelo mixto - datos
                # imputados con la media
moj.m=na.omit(mojarra)
contrasts(moj.m$probiotico)=
          contr.poly(levels(moj.m$probiotico))
ajuste7=lme(Peso.45 ~ probiotico, random=~1|Acuario,
            data=moj.m)
summary(ajuste7) #ajuste modelo mixto-sin imputar los datos
```

Observe que, en este caso, la estimación de β y los t -valores del ajuste6 coinciden con la estimación de β y los t -valores obtenidos a partir de los F -valores del ajuste5, esto es, la estimación de β vía mínimos cuadrados es igual a la estimación de β vía máxima verosimilitud restringida porque el diseño es balanceado.



8. Análisis de covarianza

El análisis de covarianza busca incorporar información auxiliar al modelo y así reducir la varianza asociada al error experimental. Se usa cuando la variable respuesta depende de covariables auxiliares de las cuales el investigador tiene información disponible. Al utilizar estas covariables auxiliares en el ajuste de los efectos del modelo sobre la variable respuesta, la comparación de los tratamientos, que es el principal objetivo, se puede realizar de manera más eficiente.

Nota 3.20. *La inclusión de una covariable es efectiva si esta permite explicar en forma substancial una parte de la variación de la variable respuesta, es decir, si es significativa [61, p. 81].*

Nota 3.21. Al usar análisis de covarianza se asume que [33, p. 252]:

- La relación funcional entre la variable respuesta, Y , y la covariable, x , es lineal.
- La relación entre la variable respuesta, Y , y la covariable, x , es la misma para cada tratamiento.
- La covariable no es afectada por los tratamientos.
- Las observaciones son realizaciones de una variable aleatoria con distribución normal.

Nota 3.22. [33, p. 253] concluyen que hay ganancia en el análisis de covarianza si la correlación entre la variable respuesta, y , y la covariable, x , es mayor a 0.3, es decir:

$$|\rho| > 0.3.$$

Ejemplo 3.30. Retomando el experimento sobre la utilización de probióticos en la fase de levante de la mojarra roja, ejemplo 3.29, los experimentadores tomaron los pesos iniciales de los alevinos, dato que puede afectar el peso a los 45 días, por lo que se usó esta información como covariable. En la tabla 3.26 se presentan los promedios en los acuarios de los pesos iniciales y los promedio de los pesos a los 45 días, y en la figura 3.13 un gráfico de dispersión de los promedios por acuario.

Tabla 3.26. Peso promedio inicial y peso promedio a los 45 días de las mojarras, por acuario, según el nivel de probiótico

Alimento con probiótico	Acuario									
	1		2		3		4		5	
	inic.	45 d.	inic.	45 d.	inic.	45 d.	inic.	45 d.	inic.	45 d.
0	8.78	19.32	8.82	21.86	8.90	25.80	8.80	22.18	8.78	18.08
2	9.18	19.64	8.82	24.36	8.92	25.10	8.66	22.98	8.74	21.60
4	8.90	26.78	8.72	30.66	8.78	29.48	9.12	33.14	8.68	26.00
6	8.86	32.40	9.04	33.80	9.04	30.96	8.98	25.76	8.82	37.14

Gramos de prob./kg

Fuente: [28].

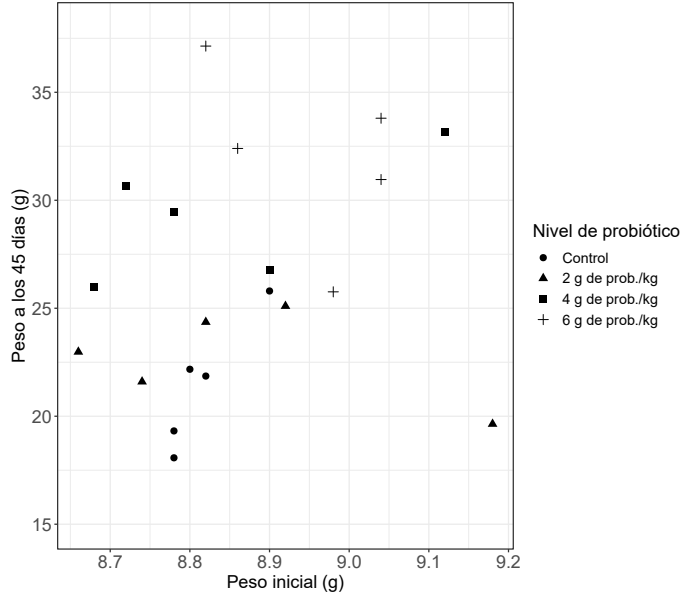


Figura 3.13. Gráfico de dispersión para el ejemplo sobre los niveles de probióticos teniendo en cuenta el peso promedio inicial y el peso promedio a los 45 días por acuario

8.1. El modelo

Sea

$$Y_{ij} = \mu + \tau_i + \gamma(x_{ij} - \bar{x}) + e_{ij} \quad i = 1, 2, \dots, a; j = 1, 2, \dots, r, \quad (3.33)$$

en que:

- x_{ij} es el valor de la covariable para la unidad ij .
- γ es el coeficiente de regresión lineal de y_{ij} sobre los x_{ij} .
- e_{ij} es el componente asociado con el error aleatorio que se asume $e_{ij} \sim N(0, \sigma_e^2)$.

Diseño completamente aleatorizado

$$\blacksquare \bar{x} = \frac{1}{ar} \sum_{i=1}^a \sum_{j=1}^r x_{ij}.$$

Asumiendo la restricción $\sum_{i=1}^a \tau_i = 0$,

- μ es la media general o global.
- τ_i es el efecto del tratamiento i con respecto a la media general, μ .

Vale resaltar que el modelo (3.33) asume que la relación entre la variable respuesta, y , y la covariable, x , es la misma para todos los tratamientos.

Matricialmente

Considerando la restricción sobre los τ en el modelo (3.33):

$$\tau_1 + \tau_2 + \tau_3 + \cdots + \tau_{a-1} + \tau_a = 0,$$

se tiene que:

$$\tau_a = -\tau_2 - \tau_3 - \cdots - \tau_{a-1}.$$

Así, en el vector de parámetros β no es necesario incluir a τ_a puesto que τ_a se puede escribir y obtener en términos de los otros τ . El modelo escrito en forma matricial es dado por:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}, \tag{3.34}$$

con

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1r} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2r} \\ \vdots \\ Y_{a1} \\ Y_{a2} \\ \vdots \\ Y_{ar} \end{pmatrix}; \quad \mathbf{X}_{n \times (a+1)} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & x_{11} - \bar{x} \\ 1 & 1 & 0 & \dots & 0 & x_{12} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 0 & \dots & 0 & x_{1r} - \bar{x} \\ 1 & 0 & 1 & \dots & 0 & x_{21} - \bar{x} \\ 1 & 0 & 1 & \dots & 0 & x_{22} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 1 & \dots & 0 & x_{2r} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & -1 & -1 & \dots & -1 & x_{a1} - \bar{x} \\ 1 & -1 & -1 & \dots & -1 & x_{a2} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & -1 & -1 & \dots & -1 & x_{ar} - \bar{x} \end{pmatrix};$$

$$\boldsymbol{\beta}_{a \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{a-1} \\ \gamma \end{pmatrix}; \quad \mathbf{E}_{n \times 1} = \begin{pmatrix} e_{11} \\ \vdots \\ e_{1r} \\ e_{21} \\ \vdots \\ e_{2r} \\ \vdots \\ e_{a1} \\ \vdots \\ e_{ar} \end{pmatrix}.$$

El estimador mínimos cuadrados de $\boldsymbol{\beta}$, teniendo en cuenta la restricción de la suma, es:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

con

$$\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma_e^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Ejemplo 3.31. Continuando con el ejemplo 3.30 sobre los probióticos, aunque el experimento tiene submuestreo en que los acuarios corresponden a las unidades experimentales y los peces a las unidades observacionales, en la tesis no se menciona que los alevinos hayan sido marcados por lo que no se puede considerar que el peso inicial de los alevinos corresponda con el peso a los 45 días de los peces, por lo cual se utilizarán los pesos promedios tanto inicial como a los 45 días de cada acuario, ignorando el submuestreo.

Encuentre un estimador de β e interprete.

Solución. El modelo de análisis de covarianza es dado por:

$$\begin{array}{c}
 \mathbf{Y} \\
 \left(\begin{array}{c} 19.32 \\ 21.86 \\ 25.80 \\ 22.18 \\ 18.08 \\ 19.64 \\ 24.36 \\ 25.10 \\ 22.98 \\ 21.60 \\ 26.78 \\ 30.66 \\ 29.48 \\ 33.14 \\ 26.00 \\ 32.40 \\ 33.80 \\ 30.96 \\ 25.76 \\ 37.14 \end{array} \right)
 \end{array}
 =
 \begin{array}{c}
 \mathbf{X} \\
 \left(\begin{array}{cccccc} 1 & 1 & 0 & 0 & 8.78 - 8.867 \\ 1 & 1 & 0 & 0 & 8.82 - 8.867 \\ 1 & 1 & 0 & 0 & 8.90 - 8.867 \\ 1 & 1 & 0 & 0 & 8.80 - 8.867 \\ 1 & 1 & 0 & 0 & 8.78 - 8.867 \\ 1 & 0 & 1 & 0 & 9.18 - 8.867 \\ 1 & 0 & 1 & 0 & 8.82 - 8.867 \\ 1 & 0 & 1 & 0 & 8.92 - 8.867 \\ 1 & 0 & 1 & 0 & 8.66 - 8.867 \\ 1 & 0 & 1 & 0 & 8.74 - 8.867 \\ 1 & 0 & 0 & 1 & 8.90 - 8.867 \\ 1 & 0 & 0 & 1 & 8.72 - 8.867 \\ 1 & 0 & 0 & 1 & 8.78 - 8.867 \\ 1 & 0 & 0 & 1 & 9.12 - 8.867 \\ 1 & 0 & 0 & 1 & 8.68 - 8.867 \\ 1 & -1 & -1 & -1 & 8.86 - 8.867 \\ 1 & -1 & -1 & -1 & 9.04 - 8.867 \\ 1 & -1 & -1 & -1 & 9.04 - 8.867 \\ 1 & -1 & -1 & -1 & 8.98 - 8.867 \\ 1 & -1 & -1 & -1 & 8.82 - 8.867 \end{array} \right)
 \end{array}
 \begin{array}{c}
 \beta \\
 \left(\begin{array}{c} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \gamma \end{array} \right)
 \end{array}
 +
 \begin{array}{c}
 \mathbf{E} \\
 \left(\begin{array}{c} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \\ e_{25} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{34} \\ e_{35} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{44} \\ e_{45} \end{array} \right)
 \end{array}
 .$$

Para calcular el vector de parámetros estimado:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 20 & 0 & 0 & 0 & 0 \\ 0 & 10 & 5 & 5 & 0 \\ 0 & 5 & 10 & 5 & 0 \\ 0 & 5 & 5 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0.3395 \end{pmatrix}; \mathbf{X}^\top \mathbf{Y} = \begin{pmatrix} 527.03 \\ -52.83 \\ -46.38 \\ -14.00 \\ 0.1824 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.05 & 0 & 0 & 0 & 0 \\ 0 & 0.15 & -0.05 & -0.05 & 0 \\ 0 & -0.05 & 0.15 & -0.05 & 0 \\ 0 & -0.05 & -0.05 & 0.15 & 0 \\ 0 & 0 & 0 & 0 & 2.9453 \end{pmatrix}; \hat{\boldsymbol{\beta}} = \begin{pmatrix} 26.351 \\ -4.906 \\ -3.616 \\ 2.861 \\ 0.537 \end{pmatrix}.$$

De $\hat{\boldsymbol{\beta}}$ se puede decir que:

- $\mu = 26.351$, es decir, que el peso de las mojaras a los 45 días en media es 26.35 gramos.
- $\hat{\tau}_1 = -4.906$, luego el efecto medio del tratamiento control, esto es, sin probiótico, sobre el peso a los 45 días es -4.9 gramos con respecto a la media general. Así, el tratamiento control disminuye en media el peso de las mojaras a los 45 días en 4.9 gramos.
- $\hat{\tau}_2 = -3.616$, es decir, que el efecto medio del alimento concentrado con 2 g de probiótico/kg sobre el peso a los 45 días es -3.6 gramos con respecto a la media general. Así, el alimento concentrado con 2 g de probiótico/kg disminuye en media el peso de las mojaras a los 45 días en 3.6 gramos.
- $\hat{\tau}_3 = 2.861$, por tanto el efecto medio del alimento concentrado con 4 g de probiótico/kg sobre el peso a los 45 días es 2.86 gramos con respecto a la media general. Así, el alimento concentrado con 4 g de probiótico/kg aumentó en media el peso de las mojaras a los 45 días en 2.86 gramos.
- $\hat{\tau}_4 = 4.906 + 3.616 - 2.861 = 5.661$, es decir, que el efecto medio del alimento concentrado con 6 g de probiótico/kg sobre el peso a los 45 días es 5.7 gramos con respecto a la media general. Así, el alimento

concentrado con 6 g de probiótico/kg aumentó en media el peso de las mojaras a los 45 días en 5.7 gramos.

- $\hat{\gamma} = 0.537$, luego por cada gramo de diferencia en el peso inicial de los alevinos, el peso promedio a los 45 días aumenta 0.54 gramos.



8.2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que ahora hay una fuente de variación adicional en el modelo, que consiste en el ajuste por la covariable. Lo anterior implica separar la varianza debida al efecto de la covariable de la varianza causada por los tratamientos. Para esto se va a particionar la matriz diseño $\mathbf{X}$ en dos partes, en que una parte corresponde a las columnas asociadas con la media general y los tratamientos ($\mathbf{X}_1$), y la otra parte corresponde a la columna asociada con la covariable ($\mathbf{X}_2$), es decir,

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \end{pmatrix} \quad \text{y} \quad \mathbf{X}_2 = \begin{pmatrix} x_{11} - \bar{x} \\ x_{12} - \bar{x} \\ \vdots \\ x_{1r} - \bar{x} \\ x_{21} - \bar{x} \\ x_{22} - \bar{x} \\ \vdots \\ x_{2r} - \bar{x} \\ \vdots \\ x_{a1} - \bar{x} \\ x_{a2} - \bar{x} \\ \vdots \\ x_{ar} - \bar{x} \end{pmatrix}.$$

La hipótesis de interés es:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0 \quad \text{versus} \quad H_a : \text{Al menos un } \tau_i \neq 0,$$

$i = 1, 2, \dots, a$. H_0 se rechazará si el P-valor asociado a los tratamientos es menor a α en la tabla del análisis de covarianza, tabla 3.27, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los tratamientos en consideración. Para la deducción de las sumas de cuadrados, esperanza de cuadrados medio y distribución de cuadrados, ver anexo 9.

Tabla 3.27. Tabla de análisis de covarianza para un diseño completamente aleatorizado

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Regresión	1	$\mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Y}^\top \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1$	SC_{Regr}	$\frac{CM_{Regr}}{CME}$	$P(F_{1,gle} > F_c)$
Tratam.	$a - 1$	$\mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Y}^\top \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 - n\bar{y}^2$	$\frac{SC_{Trat}}{a - 1}$	$\frac{CM_{Trat}}{CME}$	$P(F_{a-1,gle} > F_c)$
Error	$n - a - 1$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}}$	$\frac{SCE}{gle}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			
$n = r * a$ y gle =grados de libertad del error. $\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^r y_{ij}$ $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad \hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{Y} \quad \hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{Y}$					

Nota 3.23. En la tabla 3.27

$$SCTotal \neq SCTrat + SCRegre + SCError.$$

Ejemplo 3.32. Continuando con el ejemplo 3.30, realice el análisis de varianza y concluya.

Solución Se tiene que:

$$\begin{aligned} \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} &= 14\,274.93, & \mathbf{Y}^\top \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 &= 14\,274.83, & \mathbf{Y}^\top \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 &= 0.098, \\ \mathbf{Y}^\top \mathbf{Y} &= 14\,433.2, & y \quad n\bar{y}^2 &= 13\,888.03. \end{aligned}$$

Diseño completamente aleatorizado

$$X_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 8,78 - 8.867 \\ 8.82 - 8.867 \\ 8.90 - 8.867 \\ 8.80 - 8.867 \\ 8.78 - 8.867 \\ 9.18 - 8.867 \\ 8.82 - 8.867 \\ 8.92 - 8.867 \\ 8.66 - 8.867 \\ 8.74 - 8.867 \\ 8.90 - 8.867 \\ 8.72 - 8.867 \\ 8.78 - 8.867 \\ 9.12 - 8.867 \\ 8.68 - 8.867 \\ 8.86 - 8.867 \\ 9.04 - 8.867 \\ 9.04 - 8.867 \\ 8.98 - 8.867 \\ 8.82 - 8.867 \end{pmatrix},$$

Tabla 3.28. Tabla de análisis de covarianza para el ejemplo 3.30

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Regresión	1	0.1	0.1	0.0093	0.9245
Tratam.	3	386.8	128.9333	12.2192	0.0003
Error	15	158.3	10.5533		
Total	19	545.2			

Si los supuestos se cumplen, se concluye de la tabla 3.28 que el peso inicial de las mojaras no es significativo¹¹ y que hay efecto del nivel de probiótico sobre el peso promedio de las mojaras a los 45 días.

Las instrucciones en R para análisis de covarianza realizado al ejemplo 3.30 son:

¹¹Vale aclarar que [28] controlaron el peso inicial de los alevinos tal que fuera alrededor de 8.8 gramos.

Diseño y análisis de experimentos

```
library(car) # Para calcular la suma de cuadrados tipo III
library(ggplot2) # Para hacer el gráfico
### correr instrucciones página 135 ###
m1.Acu = mojarra %>% group_by(probiotico,Acuario) %>%
summarise(pesoI=mean(Peso.Inicial,na.rm=T),
           peso45=mean(Peso.45,na.rm=T))
m1.Acu
ggplot(m1.Acu, aes(x=pesoI, y=peso45,
  shape=factor(probiotico),
  labels=c("Control","2 gr de prob./kg","4 gr de prob./kg",
  "6 gr de prob./kg")))) + geom_point(size=2.5) + theme_bw()+
ylim(c(15, 38)) + labs(y="Peso a los 45 días (g)",
x="Peso inicial (g)") +
theme(axis.text = element_text(size=15),
axis.title = element_text(size=15),
legend.title=element_text(size=15),
legend.text=element_text(size=12))+
guides(shape=guide_legend("Nivel de probiótico"))
mean(m1.Acu$pesoI)
contrasts(m1.Acu$probiotico)=
  contr.sum(levels(m1.Acu$probiotico))
m1.Acu <- m1.Acu %>% mutate(PI.c=pesoI-mean(pesoI))
# centra la covariable alrededor de la media
ajuste.cov1 = aov(peso45 ~ probiotico + PI.c,data=m1.Acu)
Anova(ajuste.cov1,type=3) # ajusta el modelo de análisis de
# covarianza usando la suma de cuadrados tipo III
```



9. Prueba Kruskal-Wallis para k muestras aleatorias independientes

Es una prueba no paramétrica usada cuando el supuesto de normalidad no está justificado. Su hipótesis nula es

$$H_0 : f_1(y) = f_2(y) = \cdots = f_k(y),$$

en que $f_i(y)$ indica la función de densidad de la población i . Así, la hipótesis nula significa que *las k muestras aleatorias provienen de poblaciones con distribuciones idénticas* [12, p. 582]. El estadístico de prueba es dado por:

$$H = \frac{1}{s^2} \left[\sum_{i=1}^k \frac{R_{i*}^2}{r_i} - \frac{n(n+1)^2}{4} \right] \quad (3.35)$$

con:

- R_{ij} , el rango de las observación ij .
- s^2 , la varianza de los rangos.
- r_i , el número de unidades en la subpoblación i .
- $R_{i*} = \sum_{j=1}^{r_i} R_{ij}$.

En caso de empates, se asigna el rango promedio a cada una de las observaciones empatadas. Si no hay empates, H se puede escribir como [35, p. 126]:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k r_i \left[\bar{R}_{i*} - \frac{n+1}{2} \right]^2.$$

Si $r_i \geq 5$ para todo $i = 1, 2, \dots, k$, bajo H_0 verdadera, H se distribuye aproximadamente χ_{k-1}^2 , y se rechaza H_0 si:

$$H > \chi_{k-1, 1-\alpha}^2.$$

Ejemplo 3.33. [29] realizó un estudio en la Universidad Nacional de Colombia sobre algunos aspectos tróficos y reproductivos de peces de un caño de aguas negras amazónicas. En el estudio, entre otros, se recolectaron diferentes especies de peces y se les midió el número de ciegos pilóricos. Para este ejemplo solo se van a considerar cuatro de las especies recolectadas en abril de 1999 con 5 o más individuos. Se desea determinar cómo varía el número de ciegos pilóricos entre las especies observadas. En la tabla 3.29 se presentan los datos. Usar $\alpha = 0.05$.

Tabla 3.29. Ejemplo del número de ciegos pilóricos en peces de un caño de aguas negras amazónicas

Especie			
<i>A.microlepis</i>	<i>H.analis</i>	<i>H.schmardae</i>	<i>K.moenkhausii</i>
17	7	9	6
16	8	10	7
15	6	11	7
15	6	8	8
20	6	9	7
	6	8	6
	6	10	6
	7	10	9
	8	10	7
	6	9	7
$r_1 = 5$	$r_2 = 10$	$r_3 = 10$	$r_4 = 10$

Fuente: [29][p.130].

Solución Para el cálculo del estadístico de prueba es necesario calcular los rangos. En la tabla 3.30 se presentan los datos en rangos. Note que los rangos se asignaron sin tener en cuenta la especie.

El estadístico de prueba calculado es:

$$\begin{aligned}
 H &= \frac{1}{101.896} \left[\frac{165^2}{5} + \frac{94^2}{10} + \frac{248.5^2}{10} + \frac{122.5^2}{10} - \frac{35(35+1)^2}{4} \right] \\
 &= 26.1711.
 \end{aligned}$$

Tabla 3.30. Ejemplo del número de ciegos pilóricos en peces de un caño de aguas negras amazónicas. Rangos

Especie			
<i>A.microlepis</i>	<i>H.analis</i>	<i>H.schmardae</i>	<i>K.moenkhausii</i>
34	13	23.5	5
33	19	27.5	13
31.5	5	30	13
31.5	5	19	19
35	5	23.5	13
	5	19	5
	5	27.5	5
	13	27.5	23.5
	19	27.5	13
	5	23.5	13
$R_{1*} = 165$	$R_{2*} = 94$	$R_{3*} = 248.5$	$R_{4*} = 122.5$

y

$$\chi^2_{3;0.95} = 7.82.$$

Por tanto, se concluirá que se rechaza la hipótesis nula, esto es, hay suficientes evidencias para considerar que hay diferencias significativas entre el número de ciegos pilóricos de las cuatro especies encontradas en el caño estudiado en abril de 1999, con un nivel de significancia del 5 %. Para el P-valor se tiene que $\chi^2_{3;0.9999} = 26.3341$, concluyendo que P-valor < 0.00001.

Las instrucciones en R para evaluar el test de Kruskal-Wallis son:

```
remove(list=ls())
especie=c(rep("A.microlepis",5),rep("H.analis",10),
          rep("K.schmardae",10),rep("H.moenkhausii",10))
# Variable independiente
ciegos=c(17,16,15,15,20,7,8,6,6,6,6,6,7,8,6,9,10,11,8,9,8,10,10,
        10,9,6,7,7,8,7,6,6,9,7,7) # Variable dependiente
especie=as.factor(especie)
ciegospil=data.frame(especie,ciegos) # une en un solo data.frame
# las variables independiente y dependiente
```



```
kruskal.test(ciegos~especie,data=ciegospil) # ejecuta el test de
                                             # Kruskal-Wallis
library(nparcomp) # Llama la libreria nparcomp para
                  # comparaciones no paramétricas múltiples
ajuste=mctp(ciegos~especie,data=ciegospil,asy.method="fisher",
            type="Tukey",alternative="two.sided") # realiza las
          # comparaciones múltiples basadas en rangos
summary(ajuste) # muestra las comparaciones múltiples
plot(ajuste)# realiza un gráfico de intervalos de confianza
```



10. Ejercicios

- 10.1 Considere los datos proporcionados en el ejemplo 3.9. Complete la tabla 3.31 sobre el número de réplicas para un experimento similar suponiendo una potencia de $1 - \beta$ y un nivel de significancia de α .

Tabla 3.31. Tamaños de muestra para el ejemplo 3.9

	$1 - \alpha$		
$1 - \beta$	0.90	0.95	0.99
0.70			
0.80			
0.85			
0.90			

- 10.2 En un estudio sobre cantidad de radón 1 en duchas [41, p. 120], se probaron seis diámetros diferentes de los orificios de las regaderas. Los datos se presentan en la tabla 3.32.

Tabla 3.32. Radón liberado versus diámetro de los orificios de las regaderas

Díámetro de los orificios	Radón liberado (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

Fuente: [41, p. 121].

Proponer un conjunto de 5 contrastes ortogonales, encontrar su varianza estimada y realizar la respectiva tabla de análisis de varianza, incluyendo causa de variación, grados de libertad, suma de cuadrados, cuadrado medio, F calculado y P-valor.

- 10.3 Para el diseño completamente aleatorizado con t tratamientos, r' replicas por tratamiento y n^* unidades de submuestreo por unidad experimental con $\mathbb{V}(Y_{ijk}) = \sigma_\eta^2 + \sigma_\varepsilon^2$ [33, p. 210], muestre que:

$$\mathbb{V}\left(\sum_{i=1}^a c_i \bar{Y}_{i**}\right) = \frac{\sigma_\eta^2 + n^* \sigma_\varepsilon^2}{r' n^*} \sum_{i=1}^a c_i^2$$

- 10.4 Sea $\mathbf{Y}$ un vector aleatorio de orden $abr \times 1$ con distribución $N(\mathbf{0}, \sigma_e^2 \mathbf{I})$.

a) Encuentre las matrices $\mathbf{A}$ y $\mathbf{B}$ tal que:

$$q_1 = \frac{1}{r} \sum_{i=1}^a \sum_{j=1}^b Y_{ij*}^2 - \frac{1}{br} \sum_{i=1}^a Y_{i**}^2$$

y

$$q_2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r Y_{ijk}^2 - \frac{1}{r} \sum_{i=1}^a \sum_{j=1}^b Y_{ij*}^2$$

se puedan escribir como las formas cuadráticas $q_1 = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ y $q_2 = \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$.

- b) Encuentre la distribución de q_1 .
 - c) Encuentre la distribución de q_2 .
 - d) Muestre que q_1 y q_2 son independientes.
 - e) Encuentre la distribución de $\frac{q_1}{q_2}$.
 - f) Encuentre $E(q_1)$.
 - g) Encuentre $E(q_2)$.
- 10.5 Con respecto a la sección 7 sobre polinomios ortogonales, muestre que al sumar las $a - 1$ sumas de cuadrados, en que a corresponde al número de niveles del factor, para los $a - 1$ coeficientes de regresión, se tiene la suma de cuadrados de los tratamientos ($SCTr$).
- 10.6 Suponga el modelo $y_{ij} = \mu + \alpha_i + e_{ij}$ con $i = 1, \dots, a$ y $j = 1, \dots, r_i$. Si los y_{ij} son rangos, derive las expresiones explícitas en la tabla de análisis de varianza para causa de variación, grados de libertad y sumas de cuadrados. Asuma que no hay empates y compare el resultado con H dado en la expresión (3.35).
- 10.7 Para los datos del ejercicio 10.2, usando el método de polinomios ortogonales, investigue si los datos exhiben una tendencia lineal o cuadrática. Como los niveles no son igualmente espaciados, use el método propuesto por [52] para la construcción de los polinomios ortogonales.



Capítulo 4

Diseños con una o más fuentes de bloqueo

El tercer aspecto del diseño experimental es el control local o bloqueo, ver la sección 3. En este capítulo se presentan los diseños en bloques completos aleatorizados, incluida su eficiencia frente a los diseños completamente aleatorizados y su análisis cuando es necesario usar un test no paramétrico. En la sección 3 se dan ideas generales sobre los diseños en bloques incompletos balanceados. En las secciones 4 y 5 se presentan los diseños en cuadro latino que consideran dos fuentes de bloqueo y los diseños grecolatinos que consideran tres fuentes de bloqueo, respectivamente. Vale resaltar que estos diseños pueden tener las variantes presentadas en el capítulo 3, esto es, submuestreo, variables independientes regresoras, efectos aleatorios, etc.

1. Diseño en bloques completos aleatorizados

Cuando se sospecha o se sabe qué fuentes de variación no deseables pueden interferir en la respuesta de un experimento, es posible reducir y/o eliminar este efecto usando **bloqueo**.

Un diseño en bloques completos aleatorizados tiene un factor experimental con a tratamientos y b bloques de tamaño a , ver tabla 4.1. Se llama **completo** porque los a tratamientos están asignados dentro de cada bloque y el proceso de aleatorización es tal que los a tratamientos son aleatoriamente asignados a las a unidades experimentales **dentro de cada bloque**.

Para que el bloqueo sea efectivo, las unidades experimentales dentro de un bloque deben ser más homogéneas que las unidades experimentales entre los bloques.

Nota 4.1. *Se supone que dos unidades experimentales en el mismo bloque son similares y pueden ser medidas bajo similares condiciones de experimentación.*

Nota 4.2. *Se supone que dos unidades experimentales en diferentes bloques pueden dar “diferentes respuestas” aún cuando sean asignados al mismo tratamiento.*

Tabla 4.1. Esquema general - diseño en bloques

Tratamiento	Bloque					
	1	2	...	j	...	b
1	Y_{11}	Y_{12}	...	Y_{1j}	...	Y_{1b}
2	Y_{21}	Y_{22}	...	Y_{2j}	...	Y_{2b}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	Y_{i1}	Y_{i2}	...	Y_{ij}	...	Y_{ib}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a	Y_{a1}	Y_{a2}	...	Y_{aj}	...	Y_{ab}

Nota 4.3. [33, p. 314] señalan que los **diseños en bloques completos aleatorizados generalizados** son contruidos tal que el número de unidades experimentales por bloque es $k = va$, es decir, el número de unidades experimentales por bloque es un múltiplo de a (número de tratamientos). El caso particular es cuando $v = 1$ (diseños en bloques completos aleatorizados) y por tanto cada bloque contiene a unidades experimentales, $k = a$.

Ejemplo 4.1. Considere un experimento realizado en la Universidad Nacional de Colombia sobre rendimiento de variedades de cebada [53]. En el experimento consideraron 20 variedades, pero para efectos **ilustrativos**, en este ejemplo se van a considerar solo 5 variedades. Los investigadores realizaron un diseño en bloques completos al azar considerando tres bloques correspondientes a las replicaciones en uno de sus experimentos. El objetivo es comparar el rendimiento de las 5 variedades consideradas teniendo en cuenta que el diseño se planeo en bloques. En la tabla 4.2 se presentan los datos, y en la figura 4.1 un gráfico de puntos.

Tabla 4.2. Ejemplo de las variedades de cebada y el rendimiento en gramos por parcela de $7.5m^2$

Variedad	Replicaciones		
	I	II	III
1	207	57	73
2	481	372	624
3	63	206	94
4	155	29	257
5	228	218	174

Fuente: [53, p. 73].

1.1. El modelo

De acuerdo con [41, p. 128], sea:

$$Y_{ij} = \mu + \tau_i + \delta_j + e_{ij} \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b, \quad (4.1)$$

en que:

- Y_{ij} es la respuesta de la unidad experimental que fue asignada al tratamiento i en el bloque j .
- e_{ij} es el componente asociado con el error aleatorio que se asume $e_{ij} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones $\sum_{i=1}^a \tau_i = 0$ y $\sum_{j=1}^b \delta_j = 0$ en el modelo (4.1),

- μ es la media general o global.
- τ_i es el efecto del tratamiento i con respecto a la media general, μ .
- δ_j es el efecto del bloque j con respecto a la media general, μ .

Matricialmente

Considerando la restricción sobre los τ y los δ en el modelo (4.1), se tiene que:

$$\tau_1 + \tau_2 + \tau_3 + \dots + \tau_{a-1} + \tau_a = 0,$$

Diseños con una o más fuentes de bloqueo

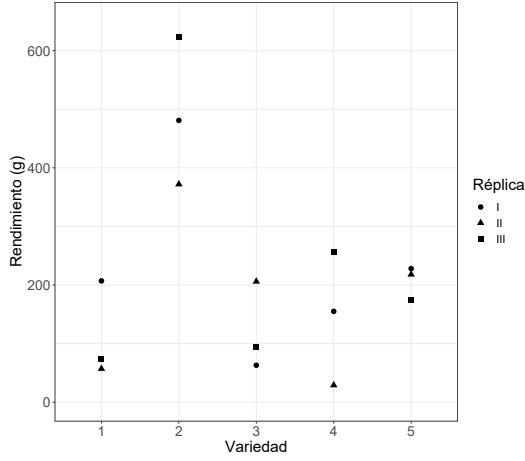


Figura 4.1. Gráfico de dispersión para el ejemplo sobre el rendimiento de 5 variedades de cebada

entonces

$$\tau_a = -\tau_1 - \tau_2 - \tau_3 - \cdots - \tau_{a-1}.$$

Igualmente,

$$\delta_1 + \delta_2 + \delta_3 + \cdots + \delta_{b-1} + \delta_b = 0,$$

entonces

$$\delta_b = -\delta_1 - \delta_2 - \delta_3 - \cdots - \delta_{b-1}.$$

Así, en el vector de parámetros β no es necesario incluir a τ_a ni a δ_b puesto que τ_a y δ_b se pueden escribir y obtener en términos de los otros τ y δ , respectivamente.

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E},$$

con

$$\underbrace{\mathbf{Y}}_{n \times 1} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1b} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2b} \\ \vdots \\ Y_{a1} \\ Y_{a2} \\ \vdots \\ Y_{ab} \end{pmatrix}; \quad \underbrace{\mathbf{X}}_{n \times (a+b-1)} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & -1 & -1 & \dots & -1 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 & 1 & 0 & \dots & 0 \\ 1 & -1 & -1 & \dots & -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 \end{pmatrix},$$

$$\underbrace{\boldsymbol{\beta}}_{(a+b-1) \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{a-1} \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{b-1} \end{pmatrix}; \quad \mathbf{E}_{n \times 1} = \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1b} \\ e_{21} \\ \vdots \\ e_{2b} \\ \vdots \\ e_{a1} \\ e_{a2} \\ \vdots \\ e_{ab} \end{pmatrix}.$$

El estimador mínimos cuadrados de $\boldsymbol{\beta}$ es dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

con

$$\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma_e^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Ejemplo 4.2. Continuando con el ejemplo 4.1, estime $\boldsymbol{\beta}$ e interprete.

Solución. El modelo, matricialmente escrito, es:

$$\begin{pmatrix} 207 \\ 57 \\ 73 \\ 481 \\ 372 \\ 624 \\ 63 \\ 206 \\ 94 \\ 155 \\ 29 \\ 257 \\ 228 \\ 218 \\ 174 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & -1 & -1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{51} \\ e_{52} \\ e_{53} \end{pmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 15 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 3 & 3 & 3 & 0 & 0 \\ 0 & 3 & 6 & 3 & 3 & 0 & 0 \\ 0 & 3 & 3 & 6 & 3 & 0 & 0 \\ 0 & 3 & 3 & 3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 & 5 \\ 0 & 0 & 0 & 0 & 0 & 5 & 10 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.067 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.267 & -0.067 & -0.067 & -0.067 & 0 & 0 \\ 0 & -0.067 & 0.267 & -0.067 & -0.067 & 0 & 0 \\ 0 & -0.067 & -0.067 & 0.267 & -0.067 & 0 & 0 \\ 0 & -0.067 & -0.067 & -0.067 & 0.267 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.133 & -0.067 \\ 0 & 0 & 0 & 0 & 0 & -0.067 & 0.133 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 3238 \\ -283 \\ 857 \\ -257 \\ -179 \\ -88 \\ -340 \end{pmatrix} ; \quad \mathbf{y} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 215.9 \\ -103.5 \\ 276.5 \\ -94.9 \\ -68.9 \\ 10.9 \\ -39.5 \end{pmatrix} .$$

De $\hat{\boldsymbol{\beta}}$ se puede concluir que:

- $\hat{\mu} = 215.9$, es decir, el rendimiento medio es 215.9 gramos.
- $\hat{\tau}_1 = -103.5$, esto es, el efecto medio de la variedad 1 sobre el rendimiento es -103.5 con respecto a la media general. Así, la variedad 1 disminuye el rendimiento en 103.5 gramos.
- $\hat{\tau}_2 = 276.5$, es decir, el efecto medio de la variedad 2 sobre el rendimiento es 276.5 con respecto a la media general. Por tanto, la variedad 2 aumenta el rendimiento en 276.5 gramos.
- $\hat{\tau}_3 = -94.9$, es decir, el efecto medio de la variedad 3 sobre el rendimiento es -94.9 con respecto a la media general. Entonces, la variedad 3 disminuye el rendimiento en 94.9 gramos.
- $\hat{\tau}_4 = -68.9$, esto es, el efecto medio de la variedad 4 sobre el rendimiento es -68.9 con respecto a la media general. Así, la variedad 4 disminuye el rendimiento en 68.9 gramos.
- $\hat{\tau}_5 = 103.5333 - 276.4667 + 94.8667 + 68.8667 = -9.2$, es decir, el efecto medio de la variedad 5 sobre el rendimiento es -9.2 con respecto a la media general. Luego, la variedad 5 disminuye el rendimiento en 9.2 gramos.
- $\hat{\delta}_1 = 10.9$, es decir, el efecto medio en la réplica I sobre el rendimiento es 10.9 con respecto a la media general. Por tanto, en la réplica I aumentó el rendimiento en 10.9 gramos.

- $\hat{\delta}_2 = -39.5$, en otras palabras, el efecto medio en la réplica II sobre el rendimiento es -39.5 con respecto a la media general. De esta forma, en la réplica II disminuyó el rendimiento en 39.5 gramos.
- $\hat{\delta}_3 = -10.9333 + 39.4667 = -28.5$, es decir, que el efecto medio en la réplica III sobre el rendimiento es -28.5 con respecto a la media general. Entonces, en la réplica III disminuyó el rendimiento en 28.5 gramos.



1.2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay una fuente de variación adicional en el modelo que consiste en el ajuste por el bloque. Lo anterior implica separar la varianza debida al efecto del factor de bloqueo de la varianza debida a los tratamientos. Para esto se va a subdividir la matriz diseño $\mathbf{X}$ en dos partes: una parte corresponde a las columnas asociadas con los tratamientos ($\mathbf{X}_\tau$), y la otra parte corresponde a las columnas asociadas con el factor de bloqueo ($\mathbf{X}_\delta$), es decir,

$$\mathbf{X}_\tau = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \end{pmatrix} \quad \text{y} \quad \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix}.$$

La hipótesis de interés es:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_{a-1} = \tau_a = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \tau_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado a los tratamientos es menor a α , un nivel de significancia dado, en la tabla del análisis de varianza, tabla 4.3, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los tratamientos en consideración.

Tabla 4.3. Análisis de varianza para un diseño en bloques completamente aleatorizados

Causa de var.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$\text{rango}(\mathbf{P}_X - \mathbf{P}_\delta) - 1 = a - 1$	$\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_\delta \mathbf{Y} - n\bar{y}^2$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CME}$	$P(F_{a-1, gle} > F_c)$
Bloques	$\text{rango}(\mathbf{P}_X - \mathbf{P}_\tau) = b - 1$	$\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_\tau \mathbf{Y}$	$\frac{SCBloq}{b - 1}$		
Error	$\text{rango}(\mathbf{I} - \mathbf{P}_X)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}}$	$\frac{SCE}{gle}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			
$gle = \text{grados de libertad del error.}$ $n = \sum_{i=1}^a r_i \quad y \quad \bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} y_{ij}$ $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \quad \mathbf{P}_\tau = \mathbf{X}_\tau(\mathbf{X}_\tau^\top \mathbf{X}_\tau)^{-1} \mathbf{X}_\tau^\top \quad \mathbf{P}_\delta = (\mathbf{X}_\delta \mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top$					

Si el diseño es balanceado, es decir, no hay datos faltantes, se tiene que:

$$\mathbf{X}_\tau^\top \mathbf{X}_\delta = \mathbf{0},$$

y entonces el diseño es ortogonal, ver sección 5, es decir,

$$\begin{aligned}
 & \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Y}^\top \mathbf{X}_\delta \hat{\boldsymbol{\beta}}_\delta - n\bar{y}^2 = \\
 & \mathbf{Y}^\top \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}_{P_{12}} - \mathbf{Y}^\top \underbrace{\mathbf{X}_\delta(\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top \mathbf{Y}}_{P_1=P_\delta} - n\bar{y}^2 \\
 & = \mathbf{Y}^\top \underbrace{\mathbf{X}_\tau(\mathbf{X}_\tau^\top \mathbf{X}_\tau)^{-1} \mathbf{X}_\tau^\top \mathbf{Y}}_{P_2=P_\tau} - n\bar{y}^2 = \mathbf{Y}^\top \mathbf{P}_\tau \mathbf{Y} - n\bar{y}^2
 \end{aligned}$$

y

$$\begin{aligned}
 & \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} - \mathbf{Y}^\top \mathbf{X}_\tau \hat{\boldsymbol{\beta}}_\tau = \\
 & \mathbf{Y}^\top \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}_{P_{12}} - \mathbf{Y}^\top \underbrace{\mathbf{X}_\tau(\mathbf{X}_\tau^\top \mathbf{X}_\tau)^{-1} \mathbf{X}_\tau^\top \mathbf{Y}}_{P_2=P_\tau} \\
 & = \mathbf{Y}^\top \underbrace{\mathbf{X}_\delta(\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top \mathbf{Y}}_{P_1=P_\delta} = \mathbf{Y}^\top \mathbf{P}_\delta \mathbf{Y}.
 \end{aligned}$$

Luego la tabla 4.3 se puede escribir de forma reducida, ver tabla 4.4.

Tabla 4.4. Análisis de varianza para un diseño en bloques completamente aleatorizado (balanceado)

Causa de var.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$a - 1$	$\mathbf{Y}^\top \mathbf{X}_\tau \hat{\boldsymbol{\beta}}_\tau - n\bar{y}^2$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CME}$	$P(F_{a-1, gle} > F_c)$
Bloques	$b - 1$	$\mathbf{Y}^\top \mathbf{X}_\delta \hat{\boldsymbol{\beta}}_\delta$	$\frac{SCBbloq}{b - 1}$		
Error	$(a - 1)(b - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}}$	$\frac{SCE}{gle}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			
$n = b \times a, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij}$					

Ejemplo 4.3. Continuando con el ejemplo 4.1, realizar el análisis de varianza y concluir.

Solución.

$$\mathbf{X}_\tau = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad \text{y} \quad \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}$$

entonces:

$$\begin{aligned} \mathbf{y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} &= 1\,014\,372.5, & \mathbf{y}^\top \mathbf{X}_\tau \hat{\boldsymbol{\beta}}_\tau &= 1\,001\,916, & \mathbf{y}^\top \mathbf{X}_\delta \hat{\boldsymbol{\beta}}_\delta &= 12\,456.53, \\ \mathbf{y}^\top \mathbf{y} &= 1\,086\,488, & n\bar{y}^2 &= 698\,976.27, & \mathbf{y}^\top \mathbf{X}_\tau \hat{\boldsymbol{\beta}}_\tau - n\bar{y}^2 &= 302\,939.7. \end{aligned}$$

Tabla 4.5. Análisis de varianza para los datos del ejemplo 4.1

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Variedad	4	302 939.7	75 734.9	8.402	0.006
Réplicas	2	12 456.53	6228.3	0.691	
Error	8	72 115.467	9014.43		
Total	14	387 511.73			

Se concluye de la tabla 4.5, siempre y cuando los supuestos se cumplan, que hay efecto significativo de la variedad de cebada sobre el rendimiento.

Las intrucciones en R para el ejemplo de bloques, ejemplo 4.1, son:

```
library(ggplot2)
remove(list=ls())
variedad <- gl(5,3,15)
replica <- gl(3,1,15,labels = c("I", "II","III"))
rendimiento <- c(207,57,73,481,372,624,63,206,94,155,29,257,228,
                218,174)
replica=as.factor(replica)
variedad=as.factor(variedad)
contrasts(variedad)=contr.sum(levels(variedad))
contrasts(replica)=contr.sum(levels(replica))
cebada=data.frame(replica,variedad,rendimiento)
ggplot(cebada, aes(x=variedad, y=rendimiento)) +
  geom_point(aes(shape = replica),size=2.5) + theme_bw()+
  ylim(c(0, 650)) + labs(y="Rendimiento (gr)", x="Variedad") +
  guides(shape=guide_legend("Réplica"))+
  theme(axis.text = element_text(size=15),
        axis.title = element_text(size=17),
        legend.text=element_text(size=12),
        legend.title=element_text(size=17))
ajuste = aov(rendimiento ~ replica + variedad,data=cebada)
summary(ajuste); ajuste$model; ajuste$residuals
ajuste$fitted.values
library(agricolae)
gle = df.residual(ajuste); SCE = deviance(ajuste)
CME = SCE/gle
HSD.test(rendimiento,variedad,gle,CME,0.05)
TukeyHSD(ajuste,which="variedad",ordered=TRUE)
plot(TukeyHSD(ajuste, "variedad",ordered=TRUE))
```



Nota 4.4. Según [41, p. 130], debido a la aleatorización dentro de los bloques, hay autores que consideran que el estadístico:

$$F_c = \frac{CMBloq}{CME},$$

no es adecuado para ver el efecto de bloques (no es un test) [33, p. 291], puesto que influye el efecto de la restricción sobre la aleatorización. Sin embargo, se puede examinar el cociente:

$$\frac{CMBloq}{CME}.$$

- Si $\frac{CMBloq}{CME} \sim 1$ entonces los bloques fueron **útiles** para reducir el tamaño de CME .
- Si $\frac{CMBloq}{CME} \neq 1$ entonces los bloques disminuyen la potencia del test de hipótesis e incrementarían la longitud de los intervalos de confianza para los contrastes de tratamientos.

Ejemplo 4.4. Continuando con el ejemplo 4.1, se observa que $\frac{CMBloq}{CME} < 1$ y por tanto parecería que no fue útil haber utilizado en el diseño el bloqueo por réplicas.

Nota 4.5. Una vez se ha realizado un diseño en bloques **NO** se puede remover como causa de variación del análisis de varianza el bloqueo, puesto que la aleatorización no es la de un diseño completamente aleatorizado.

Nota 4.6. Para tamaño de muestra y comparaciones múltiples se procede de forma similar a los cálculos en un diseño completamente aleatorizado haciendo los respectivos cambios de r por b y de los grados de libertad.

1.3. Datos faltantes

La ortogonalidad del diseño en bloques completos aleatorizados depende del balanceamiento del diseño, luego en caso de presentarse datos faltantes, se pierde la ortogonalidad. Una solución exacta para hacer estimación cuando hay presencia de datos faltantes es realizar los cálculos desde la matriz diseño, sin el dato faltante, teniendo en cuenta que el diseño **NO** será ortogonal, es decir,

$$\mathbf{X}_\tau^\top \mathbf{X}_\delta \neq \mathbf{0}.$$

Los grados de libertad del error corresponden al número de datos menos el rango de la matriz diseño $\mathbf{X}$, que en este caso corresponde a n menos el número de columnas de la matriz diseño $\mathbf{X}$.

Ejemplo 4.5. Continuando con el ejemplo 4.1, suponga que se pierde la observación asociada a la variedad 2 de la réplica 3. ¿cómo queda el análisis de varianza?

Solución. Teniendo en cuenta el desbalanceamiento, las restricciones serían:

$$\begin{aligned} 3\tau_1 + 2\tau_2 + 3\tau_3 + 3\tau_4 + 3\tau_5 &= 0 & 5\delta_1 + 5\delta_2 + 4\delta_3 &= 0 \\ \tau_5 &= -\tau_1 - (2/3)\tau_2 - \tau_3 - \tau_4 & \delta_3 &= (-5/4)\delta_1 + (-5/4)\delta_2, \end{aligned}$$

y

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$$

$$\begin{bmatrix} 207 \\ 57 \\ 73 \\ 481 \\ 372 \\ 63 \\ 206 \\ 94 \\ 155 \\ 29 \\ 257 \\ 228 \\ 218 \\ 174 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & -5/4 & -5/4 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & -5/4 & -5/4 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & -5/4 & -5/4 \\ 1 & -1 & -2/3 & -1 & -1 & 1 & 0 \\ 1 & -1 & -2/3 & -1 & -1 & 0 & 1 \\ 1 & -1 & -2/3 & -1 & -1 & -5/4 & -5/4 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{51} \\ e_{52} \\ e_{53} \end{bmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 14 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 2 & 3 & 3 & 0 & 0 \\ 0 & 2 & 3.33 & 2 & 2 & 1.17 & 1.17 \\ 0 & 3 & 2 & 6 & 3 & 0 & 0 \\ 0 & 3 & 2 & 3 & 6 & 0 & 0 \\ 0 & 0 & 1.17 & 0 & 0 & 11.25 & 6.25 \\ 0 & 0 & 1.17 & 0 & 0 & 6.25 & 11.25 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.0714 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2628 & -0.0765 & -0.0706 & -0.0706 & 0.0051 & 0.0051 \\ 0 & -0.0765 & 0.4592 & -0.0765 & -0.0765 & -0.0306 & -0.0306 \\ 0 & -0.0706 & -0.0765 & 0.2628 & -0.0706 & 0.0051 & 0.0051 \\ 0 & -0.0706 & -0.0765 & -0.0706 & 0.2628 & 0.0051 & 0.0051 \\ 0 & 0.0051 & -0.0306 & 0.0051 & 0.0051 & 0.1306 & -0.0694 \\ 0 & 0.0051 & -0.0306 & 0.0051 & 0.0051 & -0.0694 & 0.1306 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 2614 \\ -283 \\ 439.67 \\ -257 \\ -179 \\ 386.5 \\ 134.5 \end{pmatrix}; \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 186.7 \\ -74.6 \\ 241.0 \\ -65.9 \\ -39.9 \\ 24.0 \\ -26.4 \end{pmatrix}$$

$$\mathbf{X}_\tau = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -2/3 & -1 & -1 \\ 1 & -1 & -2/3 & -1 & -1 \\ 1 & -1 & -2/3 & -1 & -1 \\ 1 & -1 & -2/3 & -1 & -1 \end{pmatrix} \quad \mathbf{y} \quad \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -5/4 & -5/4 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -5/4 & -5/4 \\ 1 & 0 \\ 0 & 1 \\ -5/4 & -5/4 \\ 1 & 0 \\ 0 & 1 \\ -5/4 & -5/4 \end{pmatrix}$$

$$\mathbf{X}_\tau^\top \mathbf{X}_\delta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1.167 & 1.167 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \mathbf{0}.$$

Entonces:

$$\begin{aligned} \mathbf{y}^\top \mathbf{X} \hat{\boldsymbol{\beta}} &= 644\,939.94, & \mathbf{y}^\top \mathbf{X}_\tau \hat{\boldsymbol{\beta}}_\tau &= 638\,544.1667, \\ \mathbf{y}^\top \mathbf{X}_\delta \hat{\boldsymbol{\beta}}_\delta &= 14\,105.8571, & \mathbf{y}^\top \mathbf{y} &= 697\,112, \\ n\bar{y}^2 &= 488\,071.1429. \end{aligned}$$

Tabla 4.6. Análisis de varianza para los datos del ejemplo 4.1, considerando la pérdida del dato asociado con la variedad 2 y la réplica III

Causa de variación	Grad. de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Variedad	4	142 762.942	35 690.7354	4.789	0.0353
Réplicas	2	6395.775	3197.8875	0.429	
Error	7	52 172.06	7453.15		
Total	13	209 040.86			

Se concluye de la tabla 4.6 que hay efecto de la variedad de cebada sobre el rendimiento. Note que $SCTrat + SCbloq + SCE \neq SCT$.

Las instrucciones en R son:

```
### correr instrucciones página 172 ###
library(car) # Para calcular la suma de cuadrados tipo III
cebadaInc=cebada[-6,]
contrasts(cebadaInc$variedad)=matrix(c(1,0,0,0, 0,1,0,0, 0,0,1,0,
    0,0,0,1, -1,-2/3,-1,-1),nrow=5,byrow=TRUE)
contrasts(cebadaInc$replica)=matrix(c(1,0, 0,1,-5/4,-5/4),
    nrow=3,byrow=TRUE)
ajuste1 = aov(rendimiento ~ replica + variedad,data=cebadaInc)
Anova(ajuste1,type=3)
```



Nota 4.7. Para generar un diseño en bloques completos aleatorizados con $a = 5$ y $b = 6$ en R, se puede usar:

```
remove(list=ls())
library(agricolaeplotr); library(agricolae)
Tratamiento<-c("A","B","C","D","E")
```

```
dis <- design.rcbd(Tratamiento,r=6,serie=2,seed=111) #seed es
#una semilla
dis$book
plot_rcdb(dis,y="block",factor_name="Tratamiento",
labels="plots")
```

1.4. Eficiencia relativa de un DBCA frente a DCA

[33, p. 291] señalan que la eficiencia relativa estimada del diseño completamente aleatorizado (DCA) con $r = b$, con respecto al diseño en bloques completos aleatorizados (DBCA) con b bloques, es dada por:

$$\begin{aligned} ERE &= \frac{\hat{\sigma}_{e,DCA}^2}{\hat{\sigma}_{e,DBCA}^2} = \frac{CME_{DCA}}{CME_{DBCA}} = \frac{\frac{SCE_{DCA}}{gle_{DCA}}}{CME_{DBCA}} = \frac{\frac{SCE_{DCA}}{a(b-1)}}{CME_{DBCA}} \\ &= \frac{\frac{SCE_{DBCA}}{a(b-1)} + \frac{SCBloq}{a(b-1)}}{CME_{DBCA}} = \frac{\frac{(a-1)}{(a-1)} \frac{SCE_{DBCA}}{a(b-1)} + \frac{1}{a} CMBloq}{CME_{DBCA}} \\ &= \frac{\frac{(a-1)}{a} CME_{DBCA} + \frac{1}{a} CMBloq}{CME_{DBCA}} = \frac{(a-1)CME_{DBCA} + CMBloq}{aCME_{DBCA}} \\ &= 1 - \frac{1}{a} + \frac{1}{a} \frac{CMBloq}{CME_{DBCA}} = 1 + \frac{1}{a} \left(\frac{CMBloq}{CME_{DBCA}} - 1 \right). \end{aligned}$$

- Si $ERE > 1$, entonces es mejor usar DBCA.
- Si $ERE \leq 1$, entonces es mejor usar DCA.

Una interpretación práctica para usar ERE es que se requieran:

$$r = b \times ERE,$$

réplicas por tratamiento para que un DCA sea tan efectivo como un DBCA con b réplicas (bloques)¹.

Ejemplo 4.6. Continuando con el ejemplo 4.1 sobre las variedades de cebada y el rendimiento, calcule la eficiencia relativa.

¹El ERE se refiere solo a la precisión de las estimaciones y no a la potencia de la prueba.

Solución. De la tabla 4.5, se sabe que:

$$CMBloq = 6228.3 \text{ y } CME_{DBCA} = 9014.43,$$

luego

$$ERE = 1 - \frac{1}{a} + \frac{1}{a} \frac{CMBloq}{CME} = 0.9382.$$

Por tanto, en este caso, es mejor usar un diseño completamente aleatorizado, y serían necesarias $r = b \times ERE = 3 \times 0.9382 = 2.8146 \approx 3$ réplicas por variedad de cebada. $\checkmark$

2. Prueba de Friedman para k muestras

Es una prueba alternativa cuando en un diseño en bloques completos aleatorizados (o en cluster) el supuesto de normalidad no se cumple o cuando las mediciones no son realizadas en escala al menos de intervalo, es decir, para la prueba de Friedman se asume que las mediciones son realizadas en al menos escala ordinal.

Supuestos del test [17, p. 299] o [35, p. 228]:

1. Los resultados dentro de un bloque no influyen los resultados dentro de otros bloques.
2. Dentro de cada bloque, las observaciones pueden ser ordenadas según algún criterio de interés.

La hipótesis nula corresponde a que los efectos de los tratamientos son los mismos, es decir, las poblaciones de interés tienen distribuciones idénticas. La hipótesis alterna es que existe una diferencia entre los efectos de tratamientos:

$$H_0 : f_1(y) = f_2(y) = \dots = f_k(y).$$

Se pueden considerar dos procedimientos para el cálculo del estadístico de prueba:

1. Según [12, p. 584], se puede utilizar cuando $b \geq 10$ y $a \geq 4$ y consiste en asignar rangos de 1 a a dentro de cada bloque, y luego calcular el estadístico de prueba:

$$G = \frac{12}{ba(a+1)} \left[\sum_{i=1}^a R_{i*}^2 \right] - 3b(a+1),$$

siendo R_{ij} el rango de la observación ij , b el número de bloques, a el número de tratamientos y

$$R_{i*} = \sum_{j=1}^b R_{ij}.$$

En este caso, la distribución del estadístico G se puede aproximar a través de la distribución chi-cuadrado,

$$G \sim \chi_{a-1}^2,$$

y se rechaza H_0 si $G > \chi_{a-1, 1-\alpha}^2$.

2. Según [17, p. 299], consiste en asignar rangos de 1 a a dentro de cada bloque, luego calcular el estadístico de prueba:

$$T = \frac{(b-1) \left[B_2 - \frac{ba(a+1)^2}{4} \right]}{A_2 - B_2}, \quad A_2 \neq B_2, \quad (4.2)$$

con:

$$A_2 = \sum_{i=1}^a \sum_{j=1}^b R_{ij}^2 \quad \text{y}$$

$$B_2 = \frac{1}{b} \sum_{i=1}^a R_{i*}^2.$$

Se rechaza H_0 si:

$$T > F_{1-\alpha, a-1, (b-1)(a-1)}.$$

Nota 4.8. ■ Si $A_2 = B_2$, no se puede usar el estadístico T dado en (4.2). Para este caso el P -valor es dado por:

$$\hat{\alpha} = \left(\frac{1}{a!} \right)^{b-1}.$$

- **Sobre comparaciones múltiples.** Para comparaciones múltiples entre efectos de tratamientos, se considerará que dos efectos de tratamiento son diferentes si:

$$|R_{i*} - R_{i' *}| > t_{1-\frac{\alpha}{2}, (b-1)(a-1)} \sqrt{\frac{2b(A_2 - B_2)}{(b-1)(a-1)}}.$$

- En caso de empates se asigna el rango promedio a cada una de las observaciones empatadas.
- Si no hay empates, A_2 se simplifica a $A_2 = \frac{ba(a+1)(2a+1)}{6}$.

Ejemplo 4.7. Continuando con el ejemplo 4.1, sobre las variedades de cebada y su rendimiento, al realizar la verificación de supuestos de normalidad y de homogeneidad de varianza, se obtiene que los P-valor son adecuados para homogeneidad no así para normalidad, lo que se evidencia en el envelope, ver figura 4.2. Por tanto, realice una prueba no paramétrica para analizar el efecto de las variedades en el rendimiento.

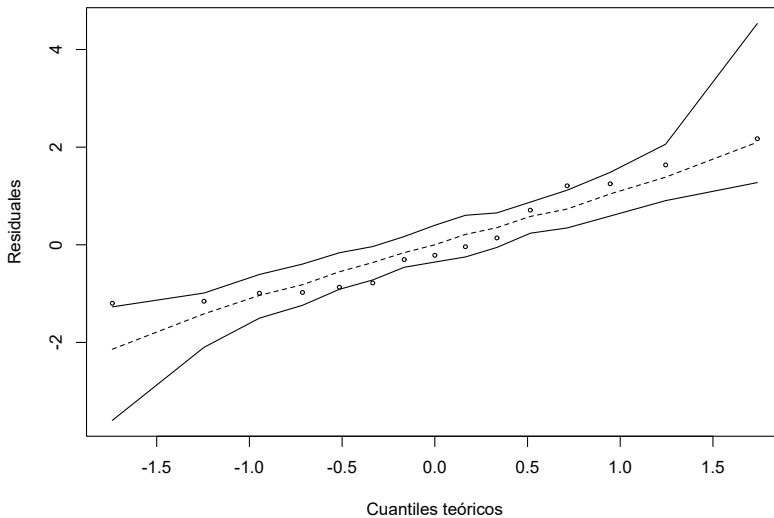


Figura 4.2. Envelope para evaluar el cumplimiento de los supuestos del modelo para los datos del ejemplo 4.1

Solución. En la tabla 4.7 se muestra los rangos. A partir de ellos se tiene que:

Tabla 4.7. Ejemplo 4.1 sobre las variedades de cebada y el rendimiento en gramos. Rangos

	Replicaciones			
Variedad	I	II	III	Suma rangos
1	3	2	1	6
2	5	5	5	15
3	1	3	2	6
4	2	1	4	7
5	4	4	3	11

$$A_2 = \frac{3 \times 5(5+1)(2 \times 5+1)}{6} = 165,$$

$$B_2 = \frac{1}{3}(6^2 + 15^2 + 6^2 + 7^2 + 11^2) = 155.667.$$

Por tanto:

$$T = \frac{(2) \left[155.667 - \frac{3 \times 5 \times 6^2}{4} \right]}{165 - 155.667} = 4.4285.$$

Como $F_{0.95;4;8} = 3.8379$, se rechaza la hipótesis que los efectos de la variedades de cebadas son iguales sobre el rendimiento.

Para comparar las cinco variedades, se usa el valor crítico:

$$t_{1-\frac{\alpha}{2},(b-1)(a-1)} \sqrt{\frac{2b(A_2 - B_2)}{(b-1)(a-1)}} = t_{0.975;8} \sqrt{\frac{2 \times 3 \times (165 - 155.667)}{8}}$$

$$= 2.306 \times \sqrt{7} = 6.101,$$

luego

Variedad	R_{i*}	Agrupación
1	6	a
3	6	a
4	7	a
5	11	ab
2	15	b

Por tanto, hay dos agrupaciones no excluyentes de las variedades de cebada. Se puede considerar que las variedades 1, 3, 4 y 5 tienen el mismo efecto medio sobre el rendimiento, mientras que las variedades 2 y 5 tiene el mismo efecto medio, siendo esta agrupación la que presentó mayor rendimiento.

La intrucción en R para el test de Friedman es:

```
### correr instrucciones página 172 ###  
friedman.test(rendimiento ~ variedad | replica)
```



Nota 4.9. Para comparaciones múltiples no paramétricas en R se puede usar la librería: *nparcomp*.

3. Ideas sobre bloques incompletos balanceados

3.1. Características

1. Cada bloque contiene k unidades experimentales con $k < a$, por tanto hay más tratamientos que unidades experimentales en un bloque.
2. Cada tratamiento aparece exactamente en r bloques.
3. Cada tratamiento aparece o una vez o no aparece en un bloque.
4. Cada par de tratamientos ocurre junto en el mismo número de bloques λ veces. Esta propiedad hace posible que cualquier par de tratamientos sean comparados con el mismo error estándar.

Entre sus propiedades están que todos los contrastes de tratamientos son estimables, y todas las comparaciones dos a dos entre efectos de tratamientos son estimadas con la misma varianza, y por tanto sus intervalos de confianza tienen la misma longitud [21, p. 353].

Si

a : Número de tratamientos

b : Número de bloques

k : Número de unidades experimentales por bloque

r : Número de réplicas por tratamiento

λ : Número de veces que aparecen un par de tratamientos

entonces un diseño en bloques incompletos balanceados cumple con las siguientes tres condiciones:

$$\begin{aligned} r &= \frac{bk}{a}, \quad b \geq a, \\ \lambda &= \frac{r(k-1)}{(a-1)}, \\ b &\geq a. \end{aligned} \tag{4.3}$$

Nota 4.10. Como λ debe ser entero, no existen bloques incompletos para cualquier r , k , a y b . [14] presentan una variedad de arreglos en bloques incompletos balanceados. En R se pueden obtener usando la instrucción:

```
library(agricolae); library(agricolaeplotr)
remove(list=ls())
Tratamiento<-c("A","B","C","D","E","F")
dis<-design.bib(Tratamiento,k=2,r=5,serie=2,seed=222)
dis$book
plot_bib(dis,y = "block", factor_name = "Tratamiento",
         labels = "plots")
```

3.1.1. Tipos de bloques incompletos balanceados:

- Tipo I: Experimentos en que los bloques pueden ser agrupados en repeticiones. Ver figura 4.3 y figura 4.4.

Diseños con una o más fuentes de bloqueo

Bloque	Rep. I	
1	T1	T2
2	T3	T4
3	T5	T6

Bloque	Rep. II	
4	T1	T3
5	T2	T5
6	T4	T6

Bloque	Rep. III	
7	T1	T4
8	T2	T6
9	T3	T5

Bloque	Rep. IV	
10	T1	T5
11	T2	T4
12	T3	T6

Bloque	Rep. V	
13	T1	T6
14	T2	T3
15	T4	T5

Figura 4.3. Diseño tipo I

Fuente: [14, p. 487].

Bloque	Rep. I	
1	T1	T2
2	T3	T4

Bloque	Rep. II	
3	T1	T3
4	T2	T4

Bloque	Rep. III	
5	T1	T4
6	T2	T3

Figura 4.4. Diseño tipo I

Fuente: [14, p. 516].

Bloque	Rep. I	
1	T1	T2
2	T2	T6
3	T3	T4
4	T4	T7
5	T1	T5
6	T5	T6
7	T3	T7

Bloque	Rep. II	
8	T1	T3
9	T2	T4
10	T3	T5
11	T4	T6
12	T5	T7
13	T1	T6
14	T2	T7

Bloque	Rep. III	
15	T1	T4
16	T2	T3
17	T3	T6
18	T4	T5
19	T2	T5
20	T6	T7
21	T1	T7

Figura 4.5. Diseño tipo II

Fuente: [14, p. 517].

- Tipo II: Son experimentos en que los bloques pueden ser agrupados en grupos de repeticiones. Ver figura 4.5.
- Tipo III: Diseños en que los bloques no pueden ser agrupados en repeticiones o grupos de repeticiones. Ver figura 4.6 y figura 4.7.

Bloque			
1	T1	T2	T5
2	T1	T2	T6
3	T1	T3	T4
4	T1	T3	T6
5	T1	T4	T5
6	T2	T3	T4
7	T2	T3	T5
8	T2	T4	T6
9	T3	T5	T6
10	T4	T5	T6

Figura 4.6. Diseño tipo III

Fuente: [14, p. 516].

Bloque						
1	T4	T6	T7	T9	T10	T11
2	T1	T5	T7	T8	T10	T11
3	T1	T2	T6	T8	T9	T11
4	T1	T2	T3	T7	T9	T10
5	T2	T3	T4	T8	T10	T11
6	T1	T3	T4	T5	T9	T11
7	T1	T2	T4	T5	T6	T10
8	T2	T3	T5	T6	T7	T11
9	T1	T3	T4	T6	T7	T8
10	T2	T4	T5	T7	T8	T9
11	T3	T5	T6	T8	T9	T10

Figura 4.7. Diseño tipo III

Fuente: [14, p. 521].

DEFINICIÓN 4.1. Según [32, p. 73], un diseño en bloques incompletos con $a = b$ se llama **un diseño en bloques incompletos simétrico**.

Nota 4.11. [21, p. 359] presentan la estimación de los efectos de tratamiento,

$$\hat{\tau}_i = \frac{k}{\lambda a} Q_i,$$

con

$$Q_i = T_i - \frac{1}{k} \sum_{h=1}^b n_{h_i} B_h,$$

donde:

T_i : Suma de todas las observaciones sobre el tratamiento i .

$$n_{h_i} = \begin{cases} 1, & \text{Si el tratamiento } i \text{ es observado en el bloque } h. \\ 0, & \text{En otro caso.} \end{cases}$$

B_h : Suma de todas las observaciones en el bloque h .

b : Número de bloques.

Así, el estimador mínimos cuadrados del contraste $\sum_{i=1}^a c_i \hat{\tau}_i$ es dado por:

$$\begin{aligned} \sum_{i=1}^a c_i \hat{\tau}_i &= \sum_{i=1}^a \frac{k}{\lambda a} Q_i c_i \\ &= \frac{k}{\lambda a} \sum_{i=1}^a \left[T_i - \frac{1}{k} \sum_{h=1}^b n_{h_i} B_h \right] c_i \\ &= \frac{k}{\lambda a} \sum_{i=1}^a T_i c_i - \frac{1}{\lambda a} \sum_{i=1}^a \sum_{h=1}^b n_{h_i} B_h c_i. \end{aligned}$$

Nota 4.12.

$$\mathbb{V} \left(\sum_{i=1}^a c_i \hat{\tau}_i \right) = \frac{k}{\lambda a} \sigma_e^2 \sum_{i=1}^a c_i^2.$$

Los test de comparaciones múltiples entre efectos de tratamientos se pueden usar teniendo en cuenta que los grados de libertad del error son: $bk - b - a + 1$.

Nota 4.13. Para detalles sobre análisis de diseños en bloques incompletos en general, sean balanceados o no, se puede adicionalmente consultar en [32, p. 1].

Nota 4.14 (Tamaño de muestra). Partiendo de:

$$\mathbb{V} (\tau_{\text{máx}} - \tau_{\text{mín}}) = \frac{k}{\lambda a} \sigma_e^2 \times 2,$$

y reemplazando λ dada en (4.3) se tiene que:

$$\mathbb{V} (\tau_{\text{máx}} - \tau_{\text{mín}}) = 2 \frac{k(a-1)}{ra(k-1)} \sigma_e^2 = \frac{2\sigma_e^2}{\frac{ra(k-1)}{k(a-1)}} = \frac{2\sigma_e^2}{r^*}, \quad (4.4)$$

$$\text{con } r^* = \frac{ra(k-1)}{k(a-1)}.$$

De (3.9) se obtiene:

$$r = \frac{2a\sigma_e^2\phi^2}{\Delta^2}, \quad (4.5)$$

con

$$\mathbb{V}(\tau_{\text{máx}} - \tau_{\text{mín}}) = \frac{2\sigma_e^2}{r}.$$

Cambiando r en (4.5) por r^* dada (4.4), se obtiene:

$$r = \frac{2\phi^2\sigma_e^2}{\Delta^2} \frac{k(a-1)}{(k-1)}.$$

El resto del procedimiento es similar al visto en la sección 2.4, teniendo en cuenta al correr el código R del anexo 4.1, cambiar los grados de libertad del error, $v2$, por $v2=bk-b-a+1^2$, esto es, $v2=ra - \frac{ra}{k} - a + 1$.

4. Diseño en cuadro latino

Cuando hay dos fuentes de bloqueo es posible organizar el diseño de forma que el número de unidades experimentales necesarias sea mínimo. Una alternativa es el diseño en cuadro latino. Un **diseño en cuadro latino** tiene un factor experimental con a tratamientos y tiene *dos fuentes de bloqueo* en que cada una tiene a niveles. Esta es una restricción ya que el número de niveles de las dos fuentes de bloqueo debe ser igual al número de tratamientos. Un esquema de cuadro latino se presenta en la figura 4.8, en que las filas están asociadas a los niveles del factor de bloqueo G , las columnas están asociadas a los niveles del factor de bloqueo F y las letras latinas dentro del cuadro están asociadas a los niveles del factor tratamiento. Es importante resaltar que cada letra latina debe aparecer solo una vez en cada columna y en cada fila del cuadrado.

$$^2v2=gle = \underbrace{b}_{\frac{ra}{k}}k - \underbrace{b}_{\frac{ra}{k}} - a + 1 = \frac{ra}{k}k - \frac{ra}{k} - a + 1 = ra - \frac{ra}{k} - a + 1 \text{ ya que } bk = ra,$$

ver (4.3).

Diseños con una o más fuentes de bloqueo

Fuente de bloqueo G	Fuente de bloqueo F			
	F1	F2	F3	F4
G1	C	D	A	B
G2	B	C	D	A
G3	A	B	C	D
G4	D	A	B	C

Figura 4.8. Ejemplo de un diseño en cuadro latino para $a = 4$

Nota 4.15. Para generar un diseño en cuadro latino con $a = 4$ en R, se puede usar:

```
library(agricolae); library(agricolaeplotr)
Tratamientos<-c("A","B","C","D")
dis <-design.lsd(Tratamientos,serie=2,seed=121212) # seed es
                                                # una semilla
dis$book; print(dis$sketch)
plot_latin_square(dis,x = "col",y = "row",
                  factor_name = "Tratamientos",labels = "plots")
```

Ejemplo 4.8. Considere un experimento realizado en la Universidad Nacional del Altiplano - Puno, Perú, sobre la determinación de nitrógeno fecal, entre otros, en llamas hembras de dos años de edad [16]. Los investigadores realizaron un diseño en cuadro latino en que los factores fueron etapas (etapa I, etapa II y etapa III), llamas (1, 2, y 3) y dietas (4 % de proteína cruda, 6 % de proteína cruda, y 8 % de proteína cruda). Este diseño implica que debe haber un adecuado periodo entre las aplicaciones de tratamiento para que sea despreciable la correlación de los datos tomados a una misma llama. La variable respuesta a considerar es pérdida de nitrógeno fecal (gramos por día, g/d). En la tabla 4.8 se presentan datos **ilustrativos**, ya que la tesis en mención no presenta los datos en forma individual, sino en estadísticas resumen.

Tabla 4.8. Ejemplo de las dietas y las llamas en cuadro latino

Etapa	Llama		
	1	2	3
1	4 % (6.97 g/d)	6 % (6.94 g/d)	8 % (7.85 g/d)
2	6 % (6.51 g/d)	8 % (7.72 g/d)	4 % (6.50 g/d)
3	8 % (7.91 g/d)	4 % (7.22 g/d)	6 % (7.73 g/d)

Entre paréntesis la pérdida de nitrógeno fecal, datos ilustrativos.

4.1. El modelo

Según [41, p. 145], sea

$$y_{ijk} = \mu + \tau_i + \delta_j + \gamma_k + e_{ijk} \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, a; \quad k = 1, 2, \dots, a,$$

en que:

- y_{ijk} es la respuesta de la unidad experimental que fue asignada al tratamiento i , en el bloque fila j , en el bloque columna k .
- e_{ijk} es el componente asociado con el error aleatorio que se asume $e_{ijk} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones $\sum_{i=1}^a \tau_i = 0$, $\sum_{j=1}^a \delta_j = 0$ y $\sum_{k=1}^a \gamma_k = 0$:

- μ es la media general o global.
- τ_i es el efecto del tratamiento i con respecto a la media general, μ .
- δ_j es el efecto del bloque fila j con respecto a la media general, μ .
- γ_k es el efecto del bloque columna k con respecto a la media general, μ .

Ejemplo 4.9. Continuando con el ejemplo 4.8, estime β .

Solución. El modelo, matricialmente escrito, es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$$

$$\begin{pmatrix} 6.97 \\ 6.51 \\ 7.91 \\ 6.94 \\ 7.72 \\ 7.22 \\ 7.85 \\ 6.5 \\ 7.73 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \delta_1 \\ \delta_2 \\ \gamma_1 \\ \gamma_2 \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} e_{111} \\ e_{221} \\ e_{331} \\ e_{212} \\ e_{322} \\ e_{132} \\ e_{313} \\ e_{123} \\ e_{233} \end{pmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 3 \\ 0 & 0 & 0 & 0 & 0 & 3 & 6 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 65.35 \\ -1.1 \\ -2.13 \\ -0.69 \\ -0.2 \\ -2.79 \\ -2.3 \end{pmatrix}; \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 7.261 \\ -0.008 \\ -0.351 \\ -0.131 \\ 0.032 \\ -0.364 \\ -0.201 \end{pmatrix}.$$



4.2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay dos fuentes de variación adicionales en el modelo, que consisten en el ajuste por el bloque fila y el ajuste por el bloque columna. Lo anterior implica separar la varianza debida al efecto del factor de bloqueo fila y al efecto del factor de bloqueo columna de la varianza debida a los tratamientos. Para esto se va a subdividir la matriz diseño, $\mathbf{X}$, en tres partes, en que la primera parte corresponde a las columnas asociadas con los tratamientos ($\mathbf{X}_\tau$), la segunda parte corresponde a las columnas asociadas al factor de bloque fila ($\mathbf{X}_\delta$), y la tercera parte corresponde a las columnas asociadas con el factor de bloqueo columna ($\mathbf{X}_\gamma$).

La hipótesis de interés es:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_{a-1} = \tau_a = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \tau_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado a los tratamientos es menor a α , el nivel de significancia establecido en la tabla del análisis de varianza, tabla 4.9, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los tratamientos en consideración.

Ejemplo 4.10. Continuando con el ejemplo 4.8, realice el análisis de varianza y concluya.

Solución

Sean las matrices:

$$\mathbf{X}_\tau = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \quad \text{y} \quad \mathbf{X}_\gamma = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Tabla 4.9. Análisis de varianza para un diseño en cuadro latino (balanceado)

Causa de var.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\tau \mathbf{Y}$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CME}$	$P(F_{a-1, gle} > F_c)$
Bloqueo fila	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\delta \mathbf{Y}$	$\frac{SCFila}{a - 1}$		
Bloqueo columna	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\gamma \mathbf{Y}$	$\frac{SCColumna}{a - 1}$		
Error	$(a - 2)(a - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$	$\frac{SCE}{gle}$		
Total	$a^2 - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			
$n = a^2 \text{ y } gle = \text{grados de libertad del error.}$ $\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^a \sum_{k=1}^a y_{ijk} \quad \mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ $\mathbf{P}_\tau = \mathbf{X}_\tau(\mathbf{X}_\tau^\top \mathbf{X}_\tau)^{-1} \mathbf{X}_\tau^\top \quad \mathbf{P}_\delta = \mathbf{X}_\delta(\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top$ $\mathbf{P}_\gamma = \mathbf{X}_\gamma(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^\top$ No se incluye el vector de unos asociado a μ en ninguna submatriz, esto es, ni en $\mathbf{X}_\tau$, ni en $\mathbf{X}_\delta$, ni en $\mathbf{X}_\gamma$.					

Asumiendo que se cumplen los supuestos del modelo, se concluye de la tabla 4.10 que no hay efecto de la dieta proteína cruda en la pérdida de nitrógeno fecal.

Tabla 4.10. Análisis de varianza para los datos del ejemplo 4.8

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Dieta	2	1.4794	0.7397	6.995	0.125
Etapas	2	0.0840	0.0420	0.397	
Llama	2	0.7564	0.3782	3.577	
Error	2	0.2115	0.1057		
Total	8				

Las instrucciones en R para obtener la tabla de análisis de varianza del ejemplo 4.8 son:

```
remove(list=ls())
etapa <- gl(3,3,labels = c("I", "II","III"))
llama <- gl(3,1,9)
dieta <- c(1,2,3,2,3,1,3,1,2)
perdidaN <- c(6.97,6.51,7.91,6.94,7.72,7.22,7.85,6.5,7.73)
etapa=as.factor(etapa);llama=as.factor(llama)
dieta=as.factor(dieta)
contrasts(etapa)=contr.sum(levels(etapa))
contrasts(llama)=contr.sum(levels(llama))
contrasts(dieta)=contr.sum(levels(dieta))
ajuste = aov(perdidaN ~ dieta + etapa+llama)
anova(ajuste)
```



Nota 4.16. Según [33, p. 376], sea T_a el número total de cuadros latinos de orden a en **forma reducida**, es decir, en la cual la primera fila y la primera columna son arregladas en orden alfabético.

El número de cuadros latinos que puede ser generado de una forma reducida por permutación de las filas y columnas es $a!(a-1)!$

Ejemplo 4.11. Ejemplo de algunos cuadros latinos en forma reducida y permutaciones del mismo:

$$a = 2 \longrightarrow 2!1! = 2$$

$$\begin{array}{cc} A & B \\ B & A \end{array} \quad y \quad \begin{array}{cc} B & A \\ A & B \end{array}$$

$$a = 3 \longrightarrow 3!2! = 12$$

$$\begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \end{array} \quad \begin{array}{ccc} A & B & C \\ C & A & B \\ B & C & A \end{array} \quad \begin{array}{ccc} A & C & B \\ C & B & A \\ B & A & C \end{array} \quad \dots$$

$$a = 4 \longrightarrow 4!3! = 144$$

A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
B	A	D	C	B	A	D	C	B	C	D	A	B	D	A	C	...
C	D	A	B	C	D	B	A	C	D	A	B	C	A	D	B	
D	C	B	A	D	C	A	B	D	A	B	C	D	C	B	A	

Nota 4.17. Para seleccionar un cuadro latino a aplicar se puede:

1. Contruir el cuadro latino estándar.
2. Aleatorizar el orden de las filas.
3. Aleatorizar el orden de las columnas.
4. Asignar los tratamientos en forma aleatoria a las letras latinas.
5. Correr el experimento por fila o por columna.

Nota 4.18 (Cuadro latino replicado). Para aumentar los grados de libertad, una alternativa es replicar el cuadro latino. Las situaciones más simples son:

1. Replicar el cuadro latino r veces. El modelo es dado por:

$$y_{ijkl} = \mu + \tau_i + \delta_j + \gamma_k + \theta_l + e_{ijkl},$$

con $i = 1, 2, \dots, a$; $j = 1, 2, \dots, a$; $k = 1, 2, \dots, a$, $l = 1, 2, \dots, r$ y θ_l el efecto de la réplica l .

Un ejemplo de un cuadro latino replicado se presenta en la figura 4.9.

1	2	3	...	r
AB	AB	AB	...	AB
BA	BA	BA	...	BA

Figura 4.9. Ejemplo de un diseño en cuadro latino replicado

La tabla de análisis de varianza con causa de variación y grados de libertad se presenta en la tabla 4.11.

Tabla 4.11. Análisis de varianza para un diseño en cuadro latino replicado

Causa de variación	Grados de libertad
Réplicas	$r - 1$
Tratamientos	$a - 1$
Filas	$a - 1$
Columnas	$a - 1$
Error	$(a - 1)[r(a + 1) - 3]$
Total	$ra^2 - 1$

Fuente:[33, p. 391].

2. *Diseño en rectángulo latino. El modelo es dado por:*

$$y_{ijk} = \mu + \tau_i + \delta_j + \gamma_k + e_{ijk},$$

con $i = 1, 2, \dots, a$; $j = 1, 2, \dots, ra$; $k = 1, 2, \dots, a$.

Un ejemplo de un diseño en rectángulo latino se presenta en la figura 4.10.

A	B	C
C	A	B
B	C	A
A	B	C
C	A	B
B	C	A

Figura 4.10. Ejemplo de un diseño en rectángulo latino

En este caso, la tabla de análisis de varianza con causa de variación y grados de libertad se presenta en la tabla 4.12.

Tabla 4.12. Análisis de varianza para un diseño en rectángulo latino

Causa de variación	Grados de libertad
Filas	$ra - 1$
Columnas	$a - 1$
Tratamientos	$a - 1$
Error	$(ra - 2)(a - 1)$
Total	$ra^2 - 1$

Fuente: [33, p. 395].

5. Diseño en cuadro grecolatino

Cuando hay tres fuentes de bloqueo y un factor de tratamientos o cuando hay dos fuentes de bloqueo y dos factores de tratamientos que se considera no tienen interacción, es posible organizar el diseño de forma que el número de unidades experimentales necesarias sea el mismo que en un diseño en cuadro latino, es decir, $n = a^2$.

La alternativa es el diseño en cuadro grecolatino. Un **diseño en cuadro grecolatino** tiene un factor experimental con a tratamientos y tres fuentes de bloqueo en que cada una tiene a niveles, o tiene dos factores experimentales cada uno con a niveles, pero que se sabe no interactúan, y dos fuentes de bloqueo cada una con a niveles. Esta es una restricción ya que el número de niveles de las fuentes de bloqueo y de los factores de tratamientos debe ser igual al número de tratamientos.

Un esquema de cuadro grecolatino se presenta en la figura 4.11 en que las filas están asociadas a los niveles del factor de bloqueo G , las columnas están asociadas a los niveles del factor de bloqueo F , las letras griegas dentro del cuadro están asociadas al factor H , que puede ser un factor de bloqueo o un factor de tratamiento, y las letras latinas dentro del cuadro están asociadas a los niveles del factor tratamiento (generalmente). Es importante resaltar que cada pareja de letras latina y griega debe aparecer solo una vez en el cuadro. Adicionalmente, un cuadro grecolatino es el resultado de superponer dos cuadros latinos.

Fuente de bloqueo G	Fuente de bloqueo F			
	F1	F2	F3	F4
G1	C β	B γ	D δ	A α
G2	B α	C δ	A γ	D β
G3	A δ	D α	B β	C γ
G4	D γ	A β	C α	B δ

Figura 4.11. Ejemplo de un diseño en cuadro grecolatino para $a = 4$

Nota 4.19. Para generar un diseño en cuadro grecolatino con $a = 4$ en R, se puede usar:

```
library(agricolae)
Tratamientos <- c("A","B","C","D")
Letrasgriegas <- c("alpha","beta","gamma","delta")
dis <- design.graeco(Tratamientos,Letrasgriegas,
  serie=2,seed=121212)
dis$book; print(dis$sketch)
```

Ejemplo 4.12. Considere el experimento descrito en [19] sobre la extracción y fraccionamiento de la humina del suelo a dos profundidades, de un oxidol del Amazonas. Para este ejemplo solo se va a tener en cuenta una de las profundidades, la denotada como horizonte A en [19]. Los investigadores, entre otros, midieron el porcentaje de carbón orgánico extractable considerando peso del suelo (0.5 g, 1.0 g, 2.5 g y 4.0 g), mililitros de extractante (20 ml, 50 ml, 100 ml y 150 ml), número de extracciones (3, 4, 5 y 6) y tiempo de agitación (1/2 h, 1 h, 1 1/2 h y 2 h). El número de extracciones se asoció con las letras latinas (A, B, C y D) respectivamente y el tiempo de agitación con las letras griegas (α , β , γ y δ) respectivamente. En este ejemplo se va a considerar como tratamiento de interés el tiempo de agitación mientras que el número de extracciones será una fuente de bloqueo, luego las letras griegas estarán asociadas al factor de los tratamientos. En la tabla 4.13 se presentan los datos. Vale aclarar que hay modificaciones con fines **ilustrativos** con respecto al cuadro grecolatino presentado en el artículo de [19], ya que no cumple con la condición de que las letras griegas y latinas deben aparecer juntas solo una vez en el cuadro.

Tabla 4.13. Ejemplo del porcentaje de carbón orgánico extractable del suelo.

Cuadro grecolatino

Peso de la muestra (g)	Mililitros de extractante			
	20	50	100	150
0.5	$A\alpha=0.20$	$D\beta=0.27$	$B\gamma=0.28$	$C\delta=0.48$
1.0	$B\beta=1.61$	$C\alpha=1.16$	$A\delta=0.85$	$D\gamma=1.24$
2.5	$D\delta=2.74$	$A\gamma=1.23$	$C\beta=1.44$	$B\alpha=1.13$
4.0	$C\gamma=3.59$	$B\delta=1.98$	$D\alpha=2.74$	$A\beta=3.81$

Asignación de número de extracciones.

Modificada con respecto al cuadro grecolatino presentado en el artículo.

Fuente: [19].

5.1. El modelo

Según [41, p. 152], sea:

$$y_{ijk} = \mu + \tau_i + \delta_j + \gamma_k + \omega_l + e_{ijk},$$

con $i = 1, 2, \dots, a$; $j = 1, 2, \dots, a$; $k = 1, 2, \dots, a$; $l = 1, 2, \dots, a$, en que:

- y_{ijkl} es la respuesta de la unidad experimental que fue asignada a las letras latinas (generalmente asociadas al tratamiento) i , en el bloque fila j , en el bloque columna k , en la letra griega l .
- e_{ijkl} es el componente asociado con el error aleatorio que se asume $e_{ijkl} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones $\sum_{i=1}^a \tau_i = 0$, $\sum_{j=1}^a \delta_j = 0$, $\sum_{k=1}^a \gamma_k = 0$ y $\sum_{l=1}^a \omega_l = 0$,

- μ es la media general o global.
- τ_i es el efecto del factor asociado a la letra latina i con respecto a la media general, μ .
- δ_j es el efecto del bloque fila j con respecto a la media general, μ .
- γ_k es el efecto del bloque columna k con respecto a la media general, μ .

- ω_l es el efecto del factor asociado a la letra griega l con respecto a la media general, μ .

Ejemplo 4.13. Continuando con el ejemplo 4.12, estime β .

Solución. El modelo, matricialmente escrito, es dado por:

$$Y = X\beta + E,$$

en que:

$$y = \begin{pmatrix} 0.20 \\ 1.61 \\ 2.74 \\ 3.59 \\ 0.27 \\ 1.16 \\ 1.23 \\ 1.98 \\ 0.28 \\ 0.85 \\ 1.44 \\ 2.74 \\ 0.48 \\ 1.24 \\ 1.13 \\ 3.81 \end{pmatrix}, \quad \beta = \begin{pmatrix} \mu \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}, \quad E = \begin{pmatrix} e_{1111} \\ e_{2212} \\ e_{4314} \\ e_{3413} \\ e_{4122} \\ e_{3221} \\ e_{1323} \\ e_{2424} \\ e_{2133} \\ e_{1234} \\ e_{3332} \\ e_{4431} \\ e_{3144} \\ e_{4243} \\ e_{2341} \\ e_{1442} \end{pmatrix}$$

y

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & -1 & -1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$X^T X = \begin{pmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 8 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 4 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 8 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 4 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 8 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 4 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 8 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 4 & 8 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 24.75 \\ 1.48 \\ -2.02 \\ -1.35 \\ -10.89 \\ -7.26 \\ -5.58 \\ -0.90 \\ -1.99 \\ -0.32 \\ -0.82 \\ 1.08 \\ 0.29 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1.547 \\ 0.488 \\ -0.387 \\ -0.219 \\ -1.239 \\ -0.332 \\ 0.088 \\ -0.024 \\ -0.297 \\ 0.121 \\ -0.239 \\ 0.236 \\ 0.038 \end{pmatrix}.$$

✓

5.2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay cuatro fuentes de variación en el modelo que consisten en el ajuste por el bloque fila, el ajuste por el bloque columna, el ajuste por el factor asociado a las letras griegas y el ajuste por el factor asociado a las letras latinas. Lo anterior implica separar la varianza debida al efecto del factor de bloque fila, al efecto del factor de bloque columna, al efecto del factor asociado a las letras griegas y al efecto del factor asociado a las letras latinas. Para esto se va a subdividir la matriz diseño, $\mathbf{X}$, en cuatro partes, en que la primera parte corresponde a las columnas asociadas con el factor de las letras latinas ($\mathbf{X}_\tau$), la segunda parte corresponde a las columnas asociadas al factor de bloque fila ($\mathbf{X}_\delta$), la tercera parte corresponde a las columnas asociadas con el factor de bloque columna ($\mathbf{X}_\gamma$) y la cuarta parte corresponde a las columnas asociadas con el factor de las letras griegas ($\mathbf{X}_\omega$).

Assumiendo que solo hay un factor asociado a los tratamientos y que fue asignado al factor asociado con las letra latinas, la hipótesis de interés es:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_{a-1} = \tau_a = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \tau_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado a los tratamientos es menor a α , el nivel de significancia establecido, en la tabla del análisis de varianza, tabla 4.14, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los tratamientos en consideración.

En la tabla 4.14, $n = a^2$, gle corresponde a los grados de libertad del error, y no se incluye el vector de unos asociado a μ en ninguna submatriz, esto es, ni en $\mathbf{X}_\omega$, ni en $\mathbf{X}_\tau$, ni en $\mathbf{X}_\delta$, ni en $\mathbf{X}_\gamma$.

Tabla 4.14. Análisis de varianza para un diseño en cuadro grecolatino (balanceado)

Causa de var.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\tau \mathbf{Y}$	$\frac{SCTrat}{a - 1}$	$\frac{CMTrat}{CME}$	$P(F_{a-1, gle} > F_c)$
Bl. fila	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\delta \mathbf{Y}$	$\frac{SCFila}{a - 1}$		
Bl. columna	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\gamma \mathbf{Y}$	$\frac{SCCol}{a - 1}$		
Bl. griega	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\omega \mathbf{Y}$	$\frac{SCgriega}{a - 1}$		
Error	$(a-3)(a-1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$	$\frac{SCE}{gle}$		
Total	$a^2 - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			
$\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^a \sum_{k=1}^a y_{ijk} \quad \mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \quad \mathbf{P}_\omega = \mathbf{X}_\omega(\mathbf{X}_\omega^\top \mathbf{X}_\omega)^{-1} \mathbf{X}_\omega^\top$ $\mathbf{P}_\tau = \mathbf{X}_\tau(\mathbf{X}_\tau^\top \mathbf{X}_\tau)^{-1} \mathbf{X}_\tau^\top \quad \mathbf{P}_\delta = \mathbf{X}_\delta(\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top \quad \mathbf{P}_\gamma = \mathbf{X}_\gamma(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^\top$					

Ejemplo 4.14. Continuando con el ejemplo 4.12, realice el análisis de varianza y concluya.

Solución

Se asignó el factor agitación que corresponde al factor de tratamiento de interés con las letras griegas para mantener la notación de [19].

$$\begin{aligned}
 \mathbf{X}_\tau = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix} \\
 \mathbf{X}_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix}, \quad \mathbf{X}_\omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.
 \end{aligned}$$

Tabla 4.15. Análisis de varianza para los datos del ejemplo 4.12

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Peso	3	1.800	0.600	1.399	
Extractante	3	15.414	5.138	11.976	
Extracciones	3	0.574	0.191	0.446	
Agitación	3	0.462	0.154	0.359	0.789
Error	3	1.287	0.429		
Total	15	19.537			

Asumiendo que se cumplen los supuestos del modelo, se concluye de la tabla 4.15 que no hay efecto del tiempo de agitación sobre el porcentaje de carbón orgánico extractable.

El programa en R para el diseño en cuadro grecolatino que se usó con los datos ejemplo 4.12 es:

```
remove(list=ls())
Peso <- gl(4,4)
extractante <- gl(4,1,16)
extracciones <- c(1,2,4,3,4,3,1,2,2,1,3,4,3,4,2,1)
agitacion <- c(1,2,4,3,2,1,3,4,3,4,2,1,4,3,1,2)
porcentaje <- c(0.2,1.61,2.74,3.59,0.27,1.16,1.23,1.98,0.28,0.85,
                1.44,2.74,0.48,1.24,1.13,3.81)
Peso=as.factor(Peso)
extractante=as.factor(extractante)
extracciones=as.factor(extracciones)
agitacion=as.factor(agitacion)
contrasts(Peso)=contr.sum(levels(Peso))
contrasts(extractante)=contr.sum(levels(extractante))
contrasts(extracciones)=contr.sum(levels(extracciones))
contrasts(agitacion)=contr.sum(levels(agitacion))
ajuste = aov(porcentaje ~ Peso+extractante+extracciones+
            agitacion)
anova(ajuste)
```



6. Ejercicios

- 6.1 Aplicar el test de Bonferroni para los datos del ejemplo 4.1 con un nivel de significancia del 0.10.
- 6.2 Usando los datos del ejemplo 4.1 como prueba piloto, calcular un tamaño de muestra mínimo considerando 10 variedades de cebada, una potencia de 0.90 y un nivel de significancia de 0.05.
- 6.3 Suponga que en el ejemplo 4.1 se pierde la observación asociada a la variedad 2 de la réplica III y la observación asociada a la variedad 3 en la réplica II. Construya la matriz diseño $\mathbf{X}_\tau$ y $\mathbf{X}_\delta$, encuentre β_τ , β_δ y el análisis de varianza. Compare con la salida de R.
- 6.4 Con datos sobre un experimento con 6 dietas sobre la ganancia de peso en conejos usando 10 camadas de tres conejos cada una [32, p. 81]. Los datos se presentan en la tabla 4.16.

Tabla 4.16. Datos sobre 6 dietas y la ganancia de peso en conejos

Bloque	Tratamiento (Rta)	Tratamiento (Rta)	Tratamiento (Rta)
1	T6 (42.2)	T2 (32.6)	T3 (35.2)
2	T3 (40.9)	T1 (40.1)	T2 (38.1)
3	T3 (34.6)	T6 (34.3)	T4 (37.5)
4	T1 (44.9)	T5 (40.8)	T3 (43.9)
5	T5 (32.0)	T3 (40.9)	T4 (37.3)
6	T2 (37.3)	T6 (42.8)	T5 (40.5)
7	T4 (37.9)	T1 (45.2)	T2 (40.6)
8	T1 (44.0)	T5 (38.5)	T6 (51.9)
9	T4 (27.5)	T2 (30.6)	T5 (20.6)
10	T6 (41.7)	T4 (42.3)	T1 (37.3)

Fuente: [32, p. 82].

- a) Especificar λ .
- b) Obtener los $\hat{\tau}_i$.

- c) Obtener las estimaciones de $\hat{\tau}_i - \hat{\tau}_j$, para $i \neq j$, y sus respectivas varianzas.
- d) Usando los datos de la tabla 4.16 como piloto, obtener un tamaño de muestra para $\alpha = 0.05$ y $1 - \beta = 0.90$.

6.5 Suponga que el diseño en cuadro grecolatino dado en la tabla 4.13 es replicado r veces.

- a) Dé el modelo lineal y su correspondiente matriz diseño considerando las restricciones usuales para los parámetros del modelo.
- b) ¿Cómo queda la tabla de análisis de varianzas?

6.6 Considere el conjunto de datos (y, x) presentado en la tabla 4.17 para un diseño completamente aleatorizado, en que y representa la respuesta una vez aplicado el tratamiento y x representa la covariable.

Tabla 4.17. Datos ilustrativos considerando una variable respuesta y una covariable

Tratamiento	y	x	y	x	y	x
1	7.5	3.5	6.2	2.6	6.8	3.1
2	10	3	12.1	3.7	11.3	4.1
3	15.2	5.1	10.7	2.6	12.9	3.1
4	5.1	4.2	4.6	3.7	7.1	4.9
5	12.1	2.3	14.2	2.9	15	3.5
6	10	3.2	9.8	3	9.6	2.5

- a) Analice los datos, ignorando la covariable, es decir,
 - 1) Obtenga las medias de tratamiento.
 - 2) Obtenga el análisis de varianzas incluyendo causa de variación, grados de libertad, suma de cuadrados, cuadrado medio, valor F calculado y P-valor.
 - 3) Realice el test de Tukey con $\alpha = 0.05$.

4) Interprete los resultados.

- b) Usando el experimento como un estudio piloto e ignorando la covariable, determine el número de réplicas por tratamiento necesario para detectar una diferencia, entre el mejor y el peor tratamiento, de 5 unidades o más, con probabilidad de 0.8 usando $\alpha = 0.05$.
- c) Realice el mismo análisis del ítem 6.6a) usando la covariable.
- d) Realice el mismo análisis del ítem 6.6b) usando la covariable.
- e) Comente los resultados de 6.6a) y 6.6b) versus 6.6c) y 6.6d).

6.7 Suponga que un investigador viene a usted para que lo aconseje sobre el análisis de un experimento que él ha conducido. Ha usado 5 tratamientos y tiene 9 observaciones para cada tratamiento [33, p. 370]. Le muestra el ANOVA, ver tabla 4.18, de la salida de computador.

Tabla 4.18. Análisis de varianza mostrado por el investigador del ejercicio 6.7

Causa de variación	Grados de libertad	Suma de cuadrados	F-calculado	P-valor
Tratamientos	4	100	8.33	0.0001
Error	40	120		

El investigador desea conocer si el análisis es correcto, y además qué significa la salida. Para cada uno de los escenarios, responder las siguientes preguntas:

- a) ¿Qué nombre tiene el diseño y qué modelo es apropiado?
- b) Basado en el modelo anterior, cuál es el correspondiente ANOVA (incluyendo causa de variación, grados de libertad y forma del test para probar $H_0 : \tau_1 = \dots = \tau_5$).
- c) ¿Es el ANOVA impreso arriba apropiado para hacer probar la hipótesis $H_0 : \tau_1 = \dots = \tau_5$?

Diseños con una o más fuentes de bloqueo

- d) Usando los datos del ANOVA como información *a priori*, obtenga el tamaño de muestra apropiado si desea alcanzar una potencia de al menos 95 % con $\alpha = 0.01$; en caso necesario dé varios escenarios.

Escenarios:

- 1) 45 animales fueron usados para el estudio y cada tratamiento fue asignado a 9 animales aleatoriamente.
- 2) 45 animales fueron usados pero venían de 9 “clases” diferentes de tamaño, 5 cada una, y cada tratamiento fue asignado a un animal de cada “clase”.
- 3) 15 animales fueron usados, cada tratamiento fue asignado a 3 animales aleatoriamente, y 3 observaciones fueron obtenidas de cada animal.



Capítulo 5

Diseños con arreglos factoriales

En este tipo de diseño se tiene más de un factor (de tratamiento) de interés. Así, en este capítulo se hace una introducción a los diseño con arreglos factoriales iniciando por los diseños con arreglos factoriales $A \times B$ en la sección 2, los diseños con arreglos factoriales $A \times B$ en bloques completos en la sección 3, y los diseños con arreglos factoriales $A \times B \times C$ en la sección 4. También se presentan como casos particulares los diseños con arreglos factoriales 2^k y 3^k , en las secciones 5 y 6, respectivamente. Vale resaltar que los diseños con arreglos factoriales pueden tener las variantes presentadas en el capítulo 3, esto es, submuestreo, variables independientes regresoras, efectos aleatorios, etc., y considerar adicionalmente fuentes de bloqueo como las presentadas en el capítulo 4.

Los factores (de tratamiento) se denotan con letras latinas mayúsculas, A, B, C , etc. Al número de niveles de los factores se les denota con letras latinas minúsculas, respectivamente, a, b, c , etc. **Un tratamiento** consiste en la combinación de niveles, uno de cada factor, el cual se denota por $(a_i, b_j, c_k, \dots)$.

Ejemplo 5.1. Considere un estudio realizado en la Universidad Nacional de Colombia sobre la energía de corte bajo diferentes condiciones de ángulo de filo, porcentaje de humedad y velocidad de corte [27]. Los niveles de los factores a tener en cuenta fueron:

- Velocidad de corte (factor A): 2.3, 3.4 y 4.5 m/s, es decir, 3 niveles $a = 3$.
- Ángulo de filo (factor B): 20° , 40° y 60° , es decir, 3 niveles $b = 3$.
- Humedad (factor C): 10 %, 20 % y 30 %, es decir, 3 niveles $c = 3$.

El número de tratamientos es $3 \times 3 \times 3 = 27$: (a_1, b_1, c_1) , (a_1, b_1, c_2) , (a_1, b_1, c_3) , (a_1, b_2, c_1) , (a_1, b_2, c_2) , (a_1, b_2, c_3) , $\dots$, (a_3, b_3, c_3) .

Una de las características medidas fue, entre otras, la energía específica de corte en J/mm^2 . ✓

Ejemplo 5.2. Considere un experimento realizado en la Universidad Nacional de Colombia sobre un diseño del proceso de producción del biodisel considerando los aceites de fritura [10]. Los niveles de los factores tenidos en cuenta fueron:

- Relación molar aceite/alcohol mol/mol (factor A): 1:5 y 1:7, es decir, 2 niveles $a = 2$.

- *Porcentaje en peso de catalizador (factor B): 0.5 y 0.7, es decir, 2 niveles $b = 2$.*
- *Temperatura en °C (factor C): 50° y 60°, es decir, 2 niveles $c = 2$.*
- *Agente de lavado (factor D): Agua (40 °C) y ácido acético, es decir, 2 niveles $d = 2$.*

El número de tratamientos es $2 \times 2 \times 2 \times 2 = 16$: (a_1, b_1, c_1, d_1) , (a_1, b_1, c_1, d_2) , (a_1, b_1, c_2, d_1) , (a_1, b_1, c_2, d_2) , (a_1, b_2, c_1, d_1) , (a_1, b_2, c_1, d_2) , $\dots$, (a_2, b_2, c_2, d_2) .

Entre otros, para determinar el rendimiento del biodiésel, los investigadores midieron el tiempo de flujo (s), la viscosidad cinemática (mm^2/s), el contenido de cenizas (%), el índice de acidez (mg KOH/g de muestra), el contenido de agua (%), el peso del AUF (aceite de cocina usados de fritura) en g, el peso del biodiésel (g) y el rendimiento (%). ✓

Los tratamientos se pueden aplicar bajo cualquier diseño visto antes (diseño completamente aleatorizado, diseño en bloques completos aleatorizados, submuestreo, cuadro latino, cuadro greco-latino y diseño con covariable(s), al igual que, si los niveles de alguno de los factores son cantidades, se pueden usar polinomios ortogonales en el análisis). En este caso el objetivo es, más que comparar tratamientos, comparar el “comportamiento” de los factores (A, B, C , etc.) individual o conjuntamente.

Ejemplo 5.3. *Considere un estudio realizado en la Universidad Nacional de Colombia sobre el efecto de la temperatura y la presión sobre el rendimiento y la actividad citotóxica de los extractos de semillas de mangos de azúcar [4]. El diseño experimental consideró dos condiciones de temperatura (80 °C y 120 °C) y tres condiciones de presión (40, 80 y 120 bar). A pesar de que se recolectaron datos en diferentes tiempos (minutos), este factor no se va a considerar para este ejemplo. Se midió, entre otros, el contenido total de fenoles (CTF) expresado en $\text{mg AG (ácido gálico)}/100 \text{ g}$ en base seca. Datos **ilustrativos** se presentan en la tabla 5.1, teniendo en cuenta que no se cuenta en la tesis con los datos individuales sino con las estadísticas resumen, y gráficos de promedios en la figura 5.1.*

Tabla 5.1. Ejemplo de la actividad citotóxica, contenido de fenoles expresados en mg AG (ácido gálico)/100 g en base seca

Temperatura	Presión		
	40 bar	80 bar	120 bar
80 °C	974.56	971.61	889.92
	974.51	971.47	889.78
	974.50	971.94	889.51
120 °C	1310.73	1029.18	907.47
	1311.08	1030.12	907.17
	1310.81	1028.81	907.67

Fuente: datos simulados con base en los promedios y desviaciones estándar reportados en [4].

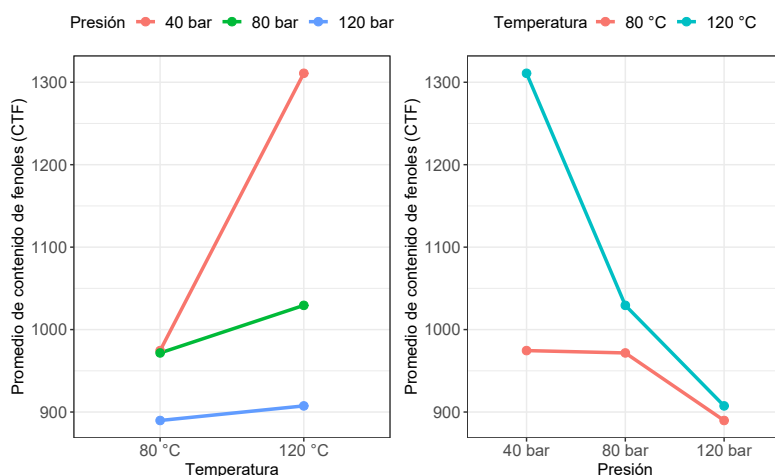


Figura 5.1. Promedios de contenido de fenoles por temperatura y presión - ejemplo 5.3

Los niveles de los factores son:

- *Temperatura (factor A): 2 niveles, $a = 2$.*
- *Presión (factor B): 3 niveles, $b = 3$.*

Por tanto se tienen $2 \times 3 = 6$ tratamientos.

Las instrucciones en R para los gráficos del ejemplo 5.3 son:

```
library(dplyr); library(ggplot2);library(cowplot)
remove(list=ls()) # Limpia R de bases anteriores
temperatura=gl(2,9,labels=c("80 °C","120 °C")) # Variable
                                     # independiente - Factor A
presion=gl(3,3,18,labels=c("40 bar","80 bar","120 bar"))
                                     # Variable independiente - Factor B
CTF=c(974.56,974.51,974.50,971.61,971.47,971.94,889.92,889.78,
      889.51,1310.73,1311.08,1310.81,1029.18,1030.12,1028.81,
      907.47,907.17,907.67)          # Variable dependiente
datos=data.frame(temperatura,presion,CTF)
datos.agr <- datos%>%mutate(
      Temperatura=temperatura%>%as.factor,
      Presión = presion %>% as.factor) %>%
      group_by(Temperatura, Presión) %>%
      summarise(Promedio = mean(CTF, na.rm = TRUE))
datos.agr <- data.frame(datos.agr)
# gráficos de perfiles
tema <- theme(axis.text=element_text(size=12),
      axis.title=element_text(size=12),
      legend.text=element_text(size=12),
      legend.title=element_text(size=12),legend.position="top")
p1 <- ggplot(datos.agr,aes(Temperatura, Promedio)) +
      geom_line(size = 1.2, aes(group = Presión, color = Presión)) +
      geom_point(size = 2.6, aes(color = Presión)) + theme_bw()+
      ylab("Promedio de contenido de fenoles (CTF)") + tema
p2 <- ggplot(datos.agr,aes(Presión, Promedio)) +
      geom_line(size = 1.2,aes(group = Temperatura,
      color = Temperatura)) +
      geom_point(size = 2.6, aes(color = Temperatura)) + theme_bw()+
```

```
ylab("Promedio de contenido de fenoles (CTF)") + tema  
plot_grid(p1,p2,nrow = 1)
```



Nota 5.1. *Cuando todos los niveles del factor A son considerados con todos los niveles del factor B, se dice que los factores A y B están cruzados.*

Nota 5.2. Efecto principal: *es el cambio en la respuesta media al cambiar de un nivel a otro en un mismo factor.*

1. Principios fundamentales para efectos de diseños con arreglos factoriales

Según [61, p. 111], los principios fundamentales para diseños con arreglos factoriales son:

- **Orden jerárquico.**
 - Efectos de orden inferior son más probables que sean más importantes que efectos de orden superior.
 - Efectos del mismo orden son más probables que sean igualmente importantes.
- **Principio de parsimonia de efectos.** El número de efectos relativamente importantes en un experimento con arreglos factoriales es pequeño.
- **Principio de heredabilidad de efectos.** Para que una interacción sea significativa, al menos uno de sus factores “padres” debe ser significativo.

2. El modelo para un diseño con arreglos factoriales $A \times B$

En un diseño con arreglos factoriales con dos factores, A y B, el modelo es dado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \quad (5.1)$$

en que:

- y_{ijk} es la respuesta de la unidad experimental k , asignada al nivel i del factor A y al nivel j del factor B .
- e_{ijk} son independientes e idénticamente distribuidos $N(0, \sigma_e^2)$.

Asumiendo las restricciones:

$$\begin{aligned} \sum_{i=1}^a \alpha_i &= 0, & \sum_{j=1}^b \beta_j &= 0, \\ \sum_{i=1}^a (\alpha\beta)_{ij} &= 0 \quad \text{para } j = 1, 2, \dots, b, \text{ y} & (5.2) \\ \sum_{j=1}^b (\alpha\beta)_{ij} &= 0 \quad \text{para } i = 1, 2, \dots, a, \end{aligned}$$

en que:

- μ es la media general o global.
- α_i es el efecto del nivel i del factor A con respecto a la media general, μ .
- β_j es el efecto del nivel j del factor B con respecto a la media general, μ .
- $(\alpha\beta)_{ij}$ es el efecto de la interacción de los niveles i de A y j de B con respecto a la media general, μ .

Nota 5.3. Si algunos de los elementos de la interacción $(\alpha\beta)_{ij}$ son significativos, se debe tener precaución en la interpretación de los parámetros. En este caso es más recomendable estimar los efectos de los tratamientos como tal, no de los parámetros $(\alpha\beta)_{ij}$, α_i y β_j .

Matricialmente

Considerando las restricciones sobre los α , los β y las interacciones $\alpha\beta$, se tiene que:

$$\begin{aligned}\alpha_a &= -\alpha_1 - \alpha_2 - \cdots - \alpha_{a-1} \quad y \\ \beta_b &= -\beta_1 - \beta_2 - \cdots - \beta_{b-1}.\end{aligned}$$

En cuanto a la columna correspondiente a la interacción, $(\alpha\beta)_{ij}$, en la matriz diseño $\mathbf{X}$, se pueden obtener multiplicando las respectivas columnas asociadas a α_i y β_j .

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E},$$

con:

$$\begin{aligned}\mathbf{Y}_{n \times 1} &= (y_{111} \ y_{112} \ \cdots \ y_{11k} \ y_{121} \ \cdots \ y_{12k} \ \cdots \ y_{1b1} \ \cdots \ y_{1bk} \ \cdots \ y_{2b1} \\ &\quad \cdots \ y_{2bk} \ \cdots \ y_{ab1} \ \cdots \ y_{abk})^\top, \\ \boldsymbol{\beta}_{ab \times 1} &= (\mu \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_{a-1} \ \beta_1 \ \beta_2 \ \cdots \ \beta_{b-1} \ (\alpha\beta)_{11} \ (\alpha\beta)_{12} \ \cdots \\ &\quad (\alpha\beta)_{1;b-1} \ (\alpha\beta)_{2;1} \ \cdots \ (\alpha\beta)_{2;b-1} \ \cdots \ (\alpha\beta)_{a-1;1} \ (\alpha\beta)_{a-1;2} \\ &\quad \cdots \ (\alpha\beta)_{a-1;b-1})^\top, \\ \mathbf{E}_{n \times 1} &= (e_{111} \ e_{112} \ \cdots \ e_{11k} \ e_{121} \ \cdots \ e_{12k} \ \cdots \ e_{1b1} \ \cdots \ e_{1bk} \ \cdots \ e_{2b1} \\ &\quad \cdots \ e_{2bk} \ \cdots \ e_{ab1} \ \cdots \ e_{abk})^\top,\end{aligned}$$

y

$$\mathbf{X}_{n \times ab} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & -1 & -1 & \dots & -1 & -1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & -1 & -1 & \dots & -1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 & -1 & -1 & \dots & -1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 & -1 & -1 & \dots & -1 & 0 & -1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & -1 & -1 & \dots & -1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 & -1 & -1 & \dots & -1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 & 1 & 1 & \dots & 1 \\ 1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

El estimador mínimos cuadrados de β es:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

con

$$\mathbb{V}(\hat{\beta}) = \sigma_e^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Ejemplo 5.4. Continuando con el ejemplo 5.3 sobre el contenido de fenoles, presente el modelo matricialmente y estime β .

Solución. El modelo, matricialmente escrito, es dado por:

$$Y = X\beta + E$$

$$\begin{bmatrix} 974.56 \\ 974.51 \\ 974.50 \\ 971.61 \\ 971.47 \\ 971.94 \\ 889.92 \\ 889.78 \\ 889.51 \\ 1310.73 \\ 1311.08 \\ 1310.81 \\ 1029.18 \\ 1030.12 \\ 1028.81 \\ 907.47 \\ 907.17 \\ 907.67 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ e_{121} \\ e_{122} \\ e_{123} \\ e_{131} \\ e_{132} \\ e_{133} \\ e_{211} \\ e_{212} \\ e_{213} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{231} \\ e_{232} \\ e_{233} \end{bmatrix}.$$

Para calcular el vector de parámetros estimado, se tiene que:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 6 & 0 & 0 \\ 0 & 0 & 6 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 12 & 6 \\ 0 & 0 & 0 & 0 & 6 & 12 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.0556 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0556 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1111 & -0.0556 & 0 & 0 \\ 0 & 0 & -0.0556 & 0.1111 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1111 & -0.0556 \\ 0 & 0 & 0 & 0 & -0.0556 & 0.1111 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 18\,250.85 \\ -1235.25 \\ 1464.66 \\ 611.61 \\ -955.94 \\ -119.98 \end{pmatrix}; \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1013.94 \\ -68.62 \\ 128.76 \\ -13.41 \\ -99.55 \\ 39.78 \end{pmatrix}.$$

De $\hat{\boldsymbol{\beta}}$ se puede concluir que $\hat{\mu} = 1013.94$, es decir, que el contenido de fenol medio es 1013.94mg/100g en base seca. En cuanto a los otros parámetros, teniendo en cuenta que está presente el parámetro de la interacción, lo mejor es interpretar usando los tratamientos, esto es,

$$\boldsymbol{\beta}^* = \left(\mu \quad \tau_1 \quad \tau_2 \quad \tau_3 \quad \tau_4 \quad \tau_5 \right)^\top,$$

con:

- τ_1 , el efecto medio de la temperatura a 80 °C y la presión a 40 bar sobre el contenido de fenol con respecto a la media general.
- τ_2 , el efecto medio de la temperatura a 120 °C y la presión a 40 bar sobre el contenido de fenol con respecto a la media general.
- τ_3 , el efecto medio de la temperatura a 80 °C y la presión a 80 bar sobre el contenido de fenol con respecto a la media general.
- τ_4 , el efecto medio de la temperatura a 120 °C y la presión a 80 bar sobre el contenido de fenol con respecto a la media general.
- τ_5 , el efecto medio de la temperatura a 80 °C y la presión a 120 bar sobre el contenido de fenol con respecto a la media general.
- $\tau_6 = -\tau_1 - \tau_2 - \tau_3 - \tau_4 - \tau_5$, el efecto medio de la temperatura a 120 °C y la presión a 120 bar sobre el contenido de fenol con respecto a la media general.

$$\hat{\beta}^* = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \\ \hat{\tau}_4 \\ \hat{\tau}_5 \end{pmatrix} = \begin{pmatrix} 1013.94 \\ -39.41 \\ 296.94 \\ -42.26 \\ 15.43 \\ -124.20 \end{pmatrix}.$$

Luego

- $\hat{\tau}_1 = -39.41$, es decir, la temperatura a 80 °C y la presión a 40 bar disminuye en promedio el contenido de fenol en 39.4 mg/100 g en base seca con respecto a la media general.
- $\hat{\tau}_2 = 296.94$, por tanto la temperatura a 120 °C y la presión a 40 bar aumenta en promedio el contenido de fenol en 296.9 mg/100 g en base seca con respecto a la media general.
- $\hat{\tau}_3 = -42.26$, entonces la temperatura a 80 °C y la presión a 80 bar disminuye en promedio el contenido de fenol en 42.3 mg/100 g en base seca con respecto a la media general.
- $\hat{\tau}_4 = 15.43$, por tanto la temperatura a 120 °C y la presión a 80 bar aumenta en promedio el contenido de fenol en 15.4 mg/100 g en base seca con respecto a la media general.
- $\hat{\tau}_5 = -124.20$, entonces la temperatura a 80 °C y la presión a 1200 bar disminuye en promedio el contenido de fenol en 124.2 mg/100 g en base seca con respecto a la media general.
- $\hat{\tau}_6 = -\tau_1 - \tau_2 - \tau_3 - \tau_4 - \tau_5 = -106.50$, esto es, la temperatura a 120 °C y la presión a 120 bar disminuye en promedio el contenido de fenol en 106.5 mg/100 g en base seca con respecto a la media general.

Si **NO hubiera interacción** entre el factor temperatura y presión, se puede decir que:

- $\hat{\alpha}_1 = -68.62$, esto es, que el efecto medio de la temperatura a 80 °C sobre el contenido de fenol es -68.62 con respecto a la media general. Así, la temperatura 80 °C disminuye el contenido de fenol en 68.62 mg AG (ácido gálico)/100 g en base seca.

- $\hat{\alpha}_2 = -\alpha_1 = 68.62$, es decir, que el efecto medio de la temperatura a 120 °C sobre el contenido de fenol es 68.62 con respecto a la media general. Así, la temperatura 120 °C aumenta el contenido de fenol en 68.62 mg AG (ácido gálico)/100 g en base seca.
- $\hat{\beta}_1 = 128.76$, es decir, que el efecto medio de la presión a 40 bar sobre el contenido de fenol es 128.76 con respecto a la media general. Así, la presión a 40 bar aumenta el contenido de fenol en 128.76 mg AG (ácido gálico)/100 g en base seca.
- $\hat{\beta}_2 = -13.41$, es decir, que el efecto medio de la presión a 80 bar sobre el contenido de fenol es -13.41 con respecto a la media general. Así, la presión a 80 bar disminuye el contenido de fenol en 13.41 mg AG (ácido gálico)/100 g en base seca.
- $\hat{\beta}_3 = -128.76 + 13.41 = -115.35$, es decir, que el efecto medio de la presión a 120 bar sobre el contenido de fenol es -115.35 con respecto a la media general. Así, la presión a 120 bar disminuye el contenido de fenol en 115.35 mg AG (ácido gálico)/100 g en base seca.



Nota 5.4. Si la interacción es significativa, la interpretación de α_i y β_j puede ser errónea. Por ejemplo, la temperatura a 120 °C podría aumentar el contenido de fenol si la presión es 40 bar, pero disminuir el contenido de fenol si la presión es 120 bar. Ver figura 5.2.

2.1. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay tres fuentes de variación en el modelo que consisten en el ajuste por el factor A , el ajuste por el factor B y el ajuste por la interacción de los factores A y B , ver tabla 5.2.

Lo anterior implica separar la varianza debida a las tres fuentes de variación. Para esto se va a subdividir la matriz diseño $\mathbf{X}$ en tres partes, en que una parte

Diseño y análisis de experimentos

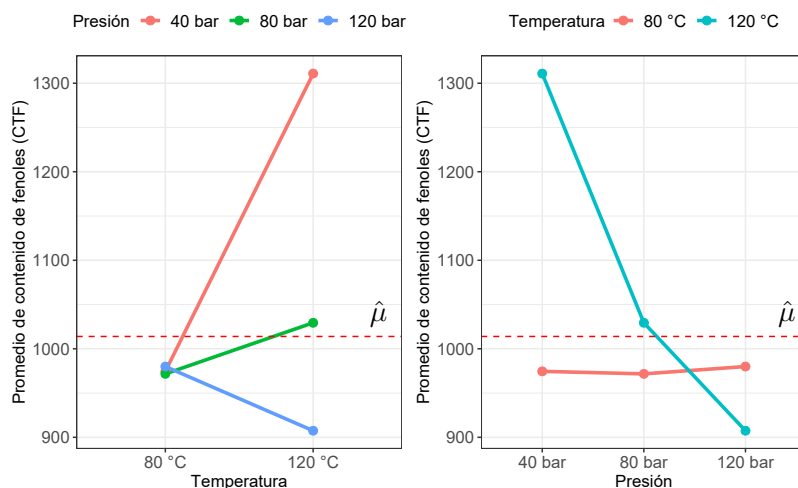


Figura 5.2. Promedios de contenido de fenoles por temperatura y presión .
Datos modificados para efecto ilustrativo

Tabla 5.2. Tabla de análisis de varianza para un diseño con arreglos factoriales $A \times B$. Diseño balanceado

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calc.	P-valor
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{SCFA}{a - 1}$	$\frac{CMFA}{CME}$	$P(F_{a-1, gle} > F_c)$
Factor B	$b - 1$	$\mathbf{Y}^\top \mathbf{P}_\beta \mathbf{Y}$	$\frac{SCFB}{b - 1}$	$\frac{CMFB}{CME}$	$P(F_{b-1, gle} > F_c)$
Fact. AB	$(a - 1)(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{(\alpha\beta)} \mathbf{Y}$	$\frac{SCFAB}{(a - 1)(b - 1)}$	$\frac{CMFAB}{CME}$	$P(F_{(a-1)(b-1), gle} > F_c)$
Error	$ab(r - 1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}$	$\frac{SCE}{gle}$		
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			

$n = a \times b \times r$ y gle =grados de libertad del error.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r y_{ijk}$$

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \quad \mathbf{P}_\alpha = \mathbf{X}_\alpha (\mathbf{X}_\alpha^\top \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^\top \quad \mathbf{P}_\beta = \mathbf{X}_\beta (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top$$

$$\mathbf{P}_{(\alpha\beta)} = \mathbf{X}_{(\alpha\beta)} (\mathbf{X}_{(\alpha\beta)}^\top \mathbf{X}_{(\alpha\beta)})^{-1} \mathbf{X}_{(\alpha\beta)}^\top$$

corresponde a las columnas asociadas al factor A , otra parte corresponde a las columnas asociadas al factor B y la tercera parte corresponde a las columnas asociadas a la interacción de los factores A y B , es decir,

$$\mathbf{X}_\alpha = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \end{pmatrix}; \mathbf{X}_\beta = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & -1 \end{pmatrix}; \mathbf{X}_{(\alpha\beta)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

Antes de considerar las hipótesis de interés, sea p_S el valor de α^1 , nivel de significancia, para retirar una variable del modelo, [45, p. 44] sugiere usar $p_S = 0.25$ o $p_S = 0.20$, esto es, si el P-valor asociado a la interacción es mayor de 0.20, por ejemplo, se decide retirar la variable del modelo. También se puede usar el criterio AIC para comparar entre dos modelos.

Nota 5.5. Criterio de información de Akaike (AIC). Es un índice usado en diferentes áreas para seleccionar entre dos modelos que compiten [25, p. 10]. Se define como:

$$AIC = -2l(\hat{\beta}) + 2q, \quad (5.3)$$

con $l(\hat{\beta}) = l(\beta)|_{\beta=\hat{\beta}}$, la función log-verosimilitud maximizada, y q el número de parámetros en el modelo. El calculo implica tanto el buen ajuste estadístico como el número de parámetros que tienen que ser estimados para obtener el modelo ajustado. Un

¹Cabe aclarar que α_i se refiere a un efecto, en cuanto α , sin el subíndice, se refiere al nivel de significancia.

valor menor de AIC indica que es un mejor modelo, ya que con menos parámetro da un buen ajuste.

Cabe resaltar que el criterio usando AIC es un procedimiento de minimización que no involucra pruebas estadísticas.

En orden, las hipótesis de interés son:

1.

$$H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{ab} = 0 \quad \text{versus} \\ H_a : \text{Al menos un } (\alpha\beta)_{ij} \neq 0, \quad i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, b.$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor al nivel de significancia preestablecido por el investigador, denotado por α , en la tabla del análisis de varianza, tabla 5.2, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de las interacciones sobre la respuesta.

En caso de rechazar la hipótesis nula sobre efectos de la interacción, **NO** se procede a analizar las hipótesis 2 y 3. Pero si se retira del modelo la interacción, entonces es válido considerar las hipótesis 2 y 3.

2.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{a-1} = \alpha_a = 0 \quad \text{versus} \\ H_a : \text{Al menos un } \alpha_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado al factor A es menor al nivel de significancia preestablecido por el investigador (α), en la tabla del análisis de varianza, tabla 5.2, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor A en consideración sobre la respuesta.

3.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{b-1} = \beta_b = 0 \quad \text{versus} \\ H_a : \text{Al menos un } \beta_j \neq 0, \quad j = 1, 2, \dots, b.$$

H_0 se rechazará si el P-valor asociado al factor B es menor al nivel de significancia preestablecido por el investigador (α), en la tabla del análisis de varianza, tabla 5.2, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor B en consideración sobre la respuesta.

En todos los casos se puede usar el criterio AIC para comparar entre los modelos, siempre y cuando se mantenga el principio de heredabilidad de efectos, en otras palabras, si la interacción entre A y B es significativa, se dejan en el modelo los factores A y B .

Ejemplo 5.5. Continuando con el ejemplo 5.3, realice la tabla de análisis de varianza y concluya.

Solución. Sean las matrices:

$$X_\alpha = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}; \quad X_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix} \quad \text{y} \quad X_{(\alpha\beta)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & 0 \\ -1 & 0 \\ -1 & 0 \\ 0 & -1 \\ 0 & -1 \\ 0 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

entonces

$$\begin{aligned} \mathbf{y}^\top \mathbf{X}_\alpha \hat{\beta}_\alpha &= 18\,860\,727, & \mathbf{y}^\top \mathbf{X}_\alpha \hat{\beta}_\alpha &= 84\,767.66, \\ \mathbf{y}^\top \mathbf{X}_\beta \hat{\beta}_\beta &= 180\,390.91, & \mathbf{y}^\top \mathbf{X}_{(\alpha\beta)} \hat{\beta}_{(\alpha\beta)} &= 90\,392.62, \\ \mathbf{y}^\top \mathbf{y} &= 18\,860\,728, & n\bar{y}^2 &= 18\,505\,175.59. \end{aligned}$$

Si se cumplen los supuestos y dado que, en la tabla 5.3, el P-valor relacionado con la interacción de los dos factores es pequeño, se rechaza la hipótesis de que la interacción es no significativa y se concluye que tanto la temperatura como la presión y su interacción tienen efecto significativo en el promedio de contenido de fenoles.

Tabla 5.3. Tabla de análisis de varianza para el ejemplo 5.3

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Presión	2	180 390.91	90 195.45	825 251	<0.0001
Temperatura	1	84 767.66	84 767.66	775 589	<0.0001
Interac. T $\times$ P	2	90 392.62	45 196.31	413 528	<0.0001
Error	12	1.3115	0.1093		
Total	17	355 552.5			

Si la interacción no hubiera sido significativa, se volvería a ajustar el modelo retirando dicha interacción. En la práctica significa que sus grados de libertad y suma de cuadrados se añaden a los grados de libertad y sumas de cuadrados del error, respectivamente.

Las instrucciones en R para el análisis de varianza presentado en la tabla 5.3 son:

```
### correr instrucciones página 214 ###
temperatura=as.factor(temperatura); presion=as.factor(presion)
contrasts(temperatura)=contr.sum(levels(temperatura))
contrasts(presion)=contr.sum(levels(presion))
ajuste1=aov(CTF~temperatura+presion+temperatura*presion,
            data=datos) # ajusta el modelo de análisis de varianza
                        # con interacción

ajuste1$coeff
summary(ajuste1) # imprime en pantalla el análisis de varianza
hnp::hnp(ajuste1, halfnormal=FALSE)
### estimando efectos de tratamientos
datos=data.frame(temperatura,presion,CTF)
tempPres=with(datos, interaction(temperatura,presion))
contrasts(tempPres)=contr.sum(levels(tempPres))
```

```
ajuste2=aov(CTF~tempPres,data=datos)
ajuste2$coeff
```



Nota 5.6. Para realizar **pruebas de comparaciones múltiples**:

- Si la interacción es significativa, se comparan los *ab* tratamientos, y para comparar niveles de factores: para cada nivel del factor *A*, se comparan los niveles del factor *B* y, viceversa, para cada nivel del factor *B*, se comparan los niveles del factor *A*.
- Si la interacción no es significativa, se pueden hacer comparaciones múltiples para cada factor del modelo en forma independiente.

Nota 5.7. Tamaños de muestra. Se debe tener en cuenta el número de unidades experimentales que hay por nivel de factor (*A* o *B* o *AB*) y la diferencia mínima que se desea detectar con el experimento [41, p. 188], es decir, se genera la tabla de potencias de acuerdo al anexo 4.1, teniendo en cuenta el número de tratamientos en cada caso, esto es:

- Para el factor *A*, “*a*” tratamientos, también se determina Δ_A y se usa:

$$r_A = \frac{2a\sigma_e^2\phi^2}{b\Delta_A^2}.$$

- Para el factor *B*, “*b*” tratamientos, también se determina Δ_B y se usa:

$$r_B = \frac{2b\sigma_e^2\phi^2}{a\Delta_B^2}.$$

- Para la interacción *AB*, “ $(a - 1)(b - 1) + 1$ ” para tratamientos, también se determina Δ_{AB} y se usa:

$$r_{AB} = \frac{2[(a - 1)(b - 1) + 1]\sigma_e^2\phi^2}{\Delta_{AB}^2},$$

y los grados de libertad del error, v_2 , son $v_2 = ab(r - 1)$ si se considera el modelo con interacciones, o $v_2 = abr - a - b + 1$ si se considera el modelo solo con efectos principales. Si el cálculo de r_A , r_B y r_{AB} no coinciden, se toma el valor más grande de los tres para el tamaño de muestra.

Nota 5.8. *Antes de obtener conclusiones del diseño con arreglos factoriales, igual que todos los casos vistos anteriormente, se deben verificar los supuestos del modelo, es decir, errores provienen de una distribución normal, verificación de residuales extremos, homogeneidad de varianza y observaciones independientes.*

2.2. Verificación de supuestos

Para la verificación de supuestos del modelo, se consideran los gráficos y test vistos en la sección 5, es decir:

- Errores independientes e idénticamente distribuidos.
- Homogeneidad de varianza.
- Errores provienen de una distribución normal.

La verificación de supuestos se realiza sobre los residuales:

$$r_{ijk} = y_{ijk} - \hat{y}_{ijk} \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad \text{y} \quad k = 1, 2, \dots, r,$$

que matricialmente se obtienen de:

$$\begin{aligned} \mathbf{R} &= \mathbf{Y} - \hat{\mathbf{Y}} \quad \text{con} \quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{P}_X \mathbf{Y} \\ \mathbf{R} &= \mathbf{Y} - \mathbf{P}_X \mathbf{Y} = (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Y}. \end{aligned}$$

Los gráficos y test más usados para hacer un chequeo de residuales son:

- Q-Q plot y test de Shapiro-Wilk para verificar que los residuales provienen de una distribución normal.
- **Envelope.** Además de indicar alejamientos del supuesto de normalidad de los errores, sirve para detectar errores asimétricos, a derecha o izquierda, si hay colas pesadas y para detectar falta de coeficientes en el modelo.
- Gráficas para verificar problemas de homogeneidad de varianza:
 - $\hat{y}_{ijk}$ versus r_{ijk} .
 - Niveles de A versus r_{ijk} .
 - Niveles de B versus r_{ijk} .

No debe haber ningún patrón ni creciente, ni decreciente, ni en forma de embudo.

- Gráfica de i versus r_i , con i la secuencia en que fueron tomados los datos. Sirve para verificar que no hayan problemas de dependencia. No debe haber ningún patrón ni creciente, ni decreciente.

3. Diseño con arreglos factoriales $A \times B$ en bloques completos

En un diseño con arreglos factoriales $A \times B$ en bloques completos, el modelo lineal es:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\beta)_{ij} + e_{ijk},$$

en que:

- y_{ijk} es la respuesta de la unidad experimental asignada al bloque k , asignada al nivel i del factor A y al nivel j del factor B .

- e_{ijk} es el componente asociado con el error aleatorio que se asume $e_{ijk} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones:

$$\begin{aligned} \sum_{i=1}^a \alpha_i &= 0, & \sum_{j=1}^b \beta_j &= 0, & \sum_{k=1}^L \delta_k &= 0, \\ \sum_{i=1}^a (\alpha\beta)_{ij} &= 0 \quad \text{para } j = 1, 2, \dots, b, \\ \sum_{j=1}^b (\alpha\beta)_{ij} &= 0 \quad \text{para } i = 1, 2, \dots, a. \end{aligned}$$

- μ es la media general o global.
- α_i es el efecto del nivel i del factor A con respecto a la media general, μ .
- β_j es el efecto del nivel j del factor B con respecto a la media general, μ .
- δ_k es el efecto del bloque k con respecto a la media general, μ .
- $(\alpha\beta)_{ij}$ es el efecto de la interacción de los niveles i de A y j de B con respecto a la media general, μ .

3.1. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay cuatro fuentes de variación en el modelo que consisten en el ajuste por el factor A , el ajuste por el factor B , el ajuste por la interacción de los factores A y B , y el ajuste por el bloque. Lo anterior implica separar la varianza debida a las cuatro fuentes de variación.

En orden, las hipótesis de interés son:

1.

$$\begin{aligned} H_0 : (\alpha\beta)_{11} &= (\alpha\beta)_{12} = \dots = (\alpha\beta)_{ab} = 0 \quad \text{versus} \\ H_a : \text{Al menos un } (\alpha\beta)_{ij} &\neq 0, \quad i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, b. \end{aligned}$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.4, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de las interacción de los factores A y B .

En caso de retirar del modelo la interacción, se procede a analizar las hipótesis 2 y 3. Si no se retira la interacción, no se deben examinar las hipótesis 2 y 3.

2.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{a-1} = \alpha_a = 0 \quad \text{versus} \\ H_a : \text{Al menos un } \alpha_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado al factor A es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.4, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor A en consideración.

3.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{b-1} = \beta_b = 0 \quad \text{versus} \\ H_a : \text{Al menos un } \beta_j \neq 0, \quad j = 1, 2, \dots, b.$$

H_0 se rechazará si el P-valor asociado al factor B es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.4, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor B en consideración.

Al igual que en el modelo para un diseño con arreglos factoriales $A \times B$, en todos los casos se puede usar el criterio AIC para comparar entre los modelos, siempre y cuando se mantenga el principio de heredabilidad de efectos.

Tabla 5.4. Análisis de varianza para un diseño con arreglos factoriales $A \times B$ en bloques completos. Diseño balanceado

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado
Bloques	$L - 1$	$\mathbf{Y}^\top \mathbf{X}_\delta \hat{\boldsymbol{\beta}}_\delta$	$\frac{SCFBloq}{L - 1}$	
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{X}_\alpha \hat{\boldsymbol{\beta}}_\alpha$	$\frac{SCFA}{a - 1}$	$\frac{CMFA}{CME}$
Factor B	$b - 1$	$\mathbf{Y}^\top \mathbf{X}_\beta \hat{\boldsymbol{\beta}}_\beta$	$\frac{SCFB}{b - 1}$	$\frac{CMFB}{CME}$
Int. AB	$(a - 1)(b - 1)$	$\mathbf{Y}^\top \mathbf{X}_{(\alpha\beta)} \hat{\boldsymbol{\beta}}_{(\alpha\beta)}$	$\frac{SCAB}{(a - 1)(b - 1)}$	$\frac{CMAB}{CME}$
Error	$(ab - 1)(L - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X} \hat{\boldsymbol{\beta}}$	$\frac{SCE}{gle}$	
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$		
$n = a \times b \times L$ y gle=grados de libertad del error. $\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^L y_{ijk}$ $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad \hat{\boldsymbol{\beta}}_\alpha = (\mathbf{X}_\alpha^\top \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^\top \mathbf{Y} \quad \hat{\boldsymbol{\beta}}_\beta = (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top \mathbf{Y}$ $\hat{\boldsymbol{\beta}}_\delta = (\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top \mathbf{Y} \quad \hat{\boldsymbol{\beta}}_{(\alpha\beta)} = (\mathbf{X}_{(\alpha\beta)}^\top \mathbf{X}_{(\alpha\beta)})^{-1} \mathbf{X}_{(\alpha\beta)}^\top \mathbf{Y}$				

Ejemplo 5.6. Considere un experimento realizado en la Universidad Autónoma de Sinaloa, México, sobre el efecto que tienen las aplicaciones de silicio en la producción de pepino bajo condiciones de invernadero [49]. Los investigadores midieron, entre otros, la altura de la planta (m) durante la etapa vegetativa y consideraron como factores de interés la dosis de silicio (0 ml, 500 ml, 1000 ml y 1500 ml por 100 litros de agua) y el nivel de riego (50 % (lámina de riego) y 100 % (riego normal)). Los tratamientos fueron asignados en tres bloques. Datos **ilustrativos** se presentan en la tabla 5.5, ya que la tesis no presenta datos individuales y un gráfico de los promedios en la figura 5.3.

Realizar el análisis de varianza y concluir.

Tabla 5.5. Datos del ejemplo sobre el efecto de silicio en la producción de pepino

Dosis de silicio	Bloque I		Bloque II		Bloque III	
	Riego		Riego		Riego	
	50 %	100 %	50 %	100 %	50 %	100 %
0	4.26	4.72	4.07	4.55	4.40	4.2
500	4.91	5.22	5.21	4.69	5.17	4.80
1000	4.35	4.62	4.95	5.12	4.61	4.78
1500	4.44	5.25	4.55	5.08	4.88	5.11

Fuente: datos simulados con base en los promedios reportados en [49].

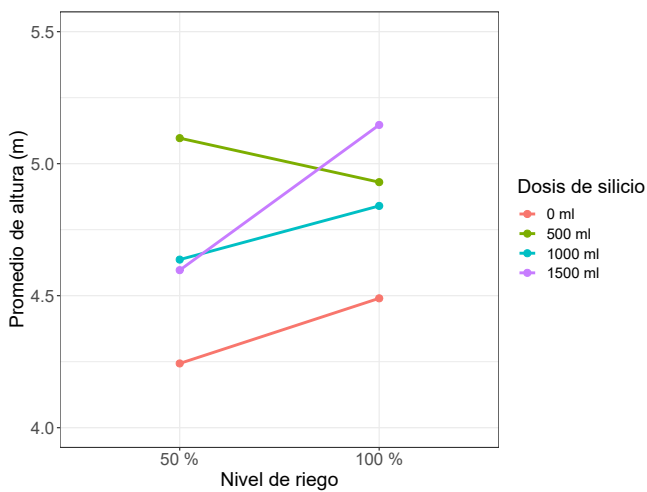


Figura 5.3. Promedios de altura por dosis de silicio y niveles de riego - ejemplo 5.6

Solución. El análisis de varianza se presenta en la tabla 5.6 en que se concluye con $p_S = 0.20$ que hay interacción entre el nivel del riego y la dosis de silicio aplicada. También se calculó el AIC para el modelo con la interacción obteniendo $AIC = -62.36$ y sin la interacción $AIC = -58.62$ concluyendo que se deja la interacción en el modelo.

Tabla 5.6. Análisis de varianza para los datos del ejemplo 5.6 sobre la producción de pepino

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Bloques	2	0.0128	0.0064	0.116	
Riego	1	0.2604	0.2604	4.699	0.048
Silicio	3	1.3872	0.4624	8.343	0.002
Int. R, y S	3	0.3883	0.1294	2.335	0.118
Error	14	0.7759	0.0554		
Total	23	2.8246			

Una vez más, cabe resaltar que estas conclusiones son válidas solo si al realizar una verificación de supuestos del modelo, no hay violación de los mismos. Un envelope se presenta en la figura 5.4.

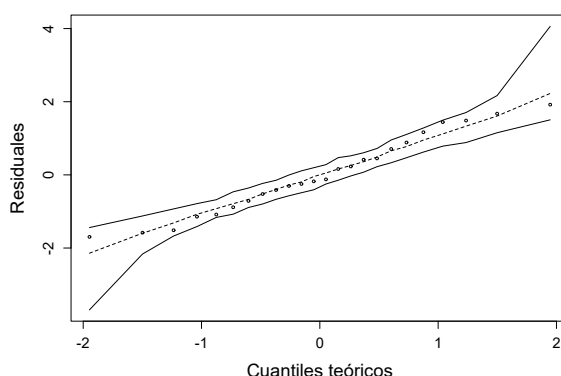


Figura 5.4. Envelope para evaluar el cumplimiento de los supuestos del modelo ajustado a los datos del ejemplo 5.6

Las instrucciones en R usadas para el ejemplo 5.6 son:

```
library(ggplot2); library(dplyr); library(readxl)
remove(list=ls())
datos = read_excel("BasesDatosDE.xlsx",sheet="pepino")
attach(datos)
datos$riego=factor(datos$riego,levels=c(0.5,1),
                  labels=c("50 %", "100 %"))
datos$silicio=factor(datos$silicio,
                    levels=c("0 ml", "500 ml","1000 ml","1500 ml"),
                    labels=c("0 ml", "500 ml","1000 ml","1500 ml"))
datos$bloque=factor(datos$bloque)
datos.agr <- datos %>% mutate(Riego = riego %>% as.factor,
                             Silicio = silicio %>% as.factor) %>%
  group_by(Riego,Silicio) %>%
  summarise(Promedio = mean(altura, na.rm = TRUE))
ggplot(datos.agr,aes(Riego, Promedio)) +
  geom_line(size = 1.2, aes(group = Silicio, color = Silicio)) +
  geom_point(size = 2.6, aes(color = Silicio))+theme_bw()+
  ylab("Promedio de altura (m)")+xlab("Nivel de riego")+
  ylim(4,5.5)+labs(colour = "Dosis de silicio")+
  theme(axis.text=element_text(size=15),
        axis.title=element_text(size=17),
        legend.text=element_text(size=13),
        legend.title=element_text(size=17))
ajuste1 = aov(altura ~ riego + silicio + riego*silicio +
             bloque,data=datos)

summary(ajuste1)
extractAIC(ajuste1)
# quitando la interacción
ajuste2 = aov(altura ~ bloque+riego + silicio , data=datos)
summary(ajuste2)
extractAIC(ajuste2)
# envelope
hnp::hnp(ajuste1,sim=99,conf=0.95,resid.type="student",
        halfnormal=FALSE,cex.axis=1.2,cex.lab=1.5)
# ajustando los efectos de tratamiento
datos$RieSil=with(datos,interaction(riego,silicio))
```

```
contrasts(datos$RieSil)=contr.sum(levels(datos$RieSil))
ajuste3=aov(altura~RieSil-1+bloque,data=datos); ajuste3$coef
ajuste4=aov(altura~RieSil+bloque,data=datos); ajuste4$coef
```



4. Diseño con arreglos factoriales $A \times B \times C$

En un diseño con arreglos factoriales con tres factores, el modelo lineal general es dado por:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\beta)_{ij} + (\alpha\delta)_{ik} + (\beta\delta)_{jk} + (\alpha\beta\delta)_{ijk} + e_{ijkl},$$

en que:

- y_{ijkl} es la respuesta de la unidad experimental l , asignada al nivel i del factor A , al nivel j del factor B y al nivel k del factor C .
- e_{ijkl} es el componente asociado con el error aleatorio que se asume $e_{ijkl} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones:

$$\begin{aligned} \sum_{i=1}^a \alpha_i &= 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{k=1}^c \delta_k = 0, \\ \sum_{i=1}^a (\alpha\beta)_{ij} &= 0 \text{ para } j = 1, \dots, b, \quad \sum_{j=1}^b (\alpha\beta)_{ij} = 0 \text{ para } i = 1, \dots, a, \\ \sum_{i=1}^a (\alpha\delta)_{ik} &= 0 \text{ para } k = 1, \dots, c, \quad \sum_{k=1}^c (\alpha\delta)_{ik} = 0 \text{ para } i = 1, \dots, a, \\ \sum_{j=1}^b (\beta\delta)_{jk} &= 0 \text{ para } k = 1, \dots, c, \quad \sum_{k=1}^c (\beta\delta)_{jk} = 0 \text{ para } j = 1, \dots, b, \\ \sum_{i=1}^a (\alpha\beta\delta)_{ijk} &= 0 \text{ para } j = 1, 2, \dots, b, \text{ y } k = 1, 2, \dots, c, \end{aligned}$$

$$\sum_{j=1}^b (\alpha\beta\delta)_{ijk} = 0 \text{ para } i = 1, 2, \dots, a, \text{ y } k = 1, 2, \dots, c,$$

$$\sum_{k=1}^c (\alpha\beta\delta)_{ijk} = 0 \text{ para } i = 1, 2, \dots, a, \text{ y } j = 1, 2, \dots, b,$$

en que:

- μ es la media general o global.
- α_i es el efecto del nivel i del factor A con respecto a la media general, μ .
- β_j es el efecto del nivel j del factor B con respecto a la media general, μ .
- δ_k es el efecto del nivel k del factor C con respecto a la media general, μ .
- $(\alpha\beta)_{ij}$ es el efecto de la interacción de los niveles i de A y j de B con respecto a la media general, μ .
- $(\alpha\delta)_{ik}$ es el efecto de la interacción de los niveles i de A y k de C con respecto a la media general, μ .
- $(\beta\delta)_{jk}$ es el efecto de la interacción de los niveles j de B y k de C con respecto a la media general, μ .
- $(\alpha\beta\delta)_{ijk}$ es el efecto de la interacción de los niveles i de A , j de B y k de C con respecto a la media general, μ .

Ejemplo 5.7. Continuando con el ejemplo 1.2 sobre el porcentaje de recuperación de compuestos volátiles bajo diferentes condiciones de viscosidad (viscosidad del agua (1 cP²), 50 cP y 100 cP), salinidad (0 g NaCl/100 g de alimentación, 2.91 g NaCl/100 g de alimentación y 5.66 g NaCl/100 g de alimentación) y tiempo de extracción (8 horas, 16 horas y 24 horas) [6], se presentan los datos para uno de los compuestos en la tabla 5.7 y gráficos en las figura 5.5, 5.6, 5.7, 5.8 y 5.9.

²cP es un centipoise en que 100 centipoise = 1 poise que corresponde 1 g/(s × cm).

Tabla 5.7. Datos ejemplo sobre porcentaje de recuperación de compuestos volátiles

Tiempo de extrac.	Viscosidad: 1 cP			Viscosidad: 50 cP			Viscosidad: 100 cP		
	Salinidad			Salinidad			Salinidad		
	0 g	2.91 g	5.66 g	0 g	2.91 g	5.66 g	0 g	2.91 g	5.66 g
8 h	0.6663	0.7091	0.5964	0.7941	0.4824	0.6260	0.3802	0.3562	0.5243
	0.6536	0.7136	0.5843	0.4131	0.5077	0.4935	0.3119	0.4024	0.5532
16 h	0.6752	0.7251	0.7421	0.4492	0.6735	0.8115	0.3167	0.4533	0.5804
	0.6814	0.7235	0.7585	0.4625	0.4008	0.6946	0.4489	0.4559	0.6561
24 h	0.6924	0.7790	0.8231	0.4358	0.6162	0.5683	0.5396	0.4635	0.5407
	0.7319	0.7429	0.8144	0.6647	0.6067	0.7857	0.5806	0.483	0.5384

Fuente: [6, p. 119].

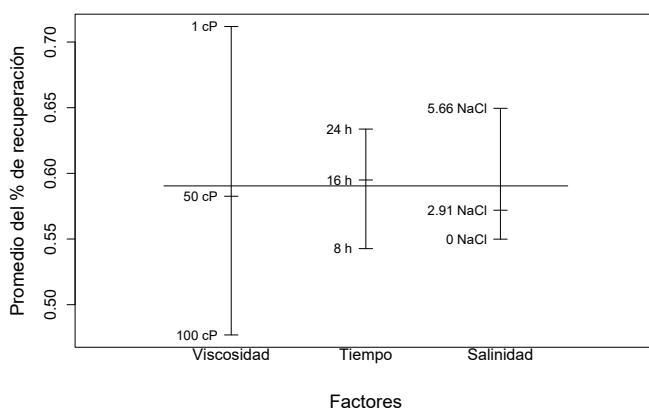


Figura 5.5. Promedios de porcentaje de recuperación de compuestos volátiles. Efectos principales

Para los niveles del factor *A* (tiempo), $i = 1$ indicará tiempo de extracción igual a 8 horas, $i = 2$ indicará tiempo de extracción igual a 16 horas e $i = 3$ indicará tiempo de extracción igual a 24 horas. Para los niveles del factor *B* (salinidad), $j = 1$ indicará una salinidad igual a 0 g NaCl/100 g de alimentación, $j = 2$ indicará una salinidad igual a 2.91 g NaCl/100 g de alimentación y $j = 3$ indicará una salinidad igual a 5.66 g NaCl/100 g de alimentación. Para los niveles del factor *C* (viscosidad), $k = 1$

indicará la viscosidad del agua (1cP), $k = 2$ indicará viscosidad igual a 50cP y $k = 3$ indicará viscosidad igual a 100cP.

Presente el modelo matricialmente.

Solución. Considerando las restricciones sobre los α , los β , los δ , las interacciones $\alpha\beta$, las interacciones $\alpha\delta$, las interacciones $\beta\delta$ y las interacciones $\alpha\beta\delta$, se tiene que:

$$Y = X\beta + E,$$

$$\begin{pmatrix} 0.6663 \\ 0.6536 \\ 0.6752 \\ 0.6814 \\ 0.6924 \\ 0.7319 \\ 0.7091 \\ 0.7136 \\ 0.7251 \\ 0.7235 \\ 0.7790 \\ \vdots \\ 0.4024 \\ 0.4533 \\ 0.4559 \\ 0.4635 \\ 0.4830 \\ 0.5243 \\ 0.5532 \\ 0.5804 \\ 0.6561 \\ 0.5407 \\ 0.5384 \end{pmatrix} = \left(\mathbf{1}_{54} || \mathbf{X}_{\alpha} || \mathbf{X}_{\beta} || \mathbf{X}_{\delta} || \mathbf{X}_{\alpha\beta} || \mathbf{X}_{\alpha\delta} || \mathbf{X}_{\beta\delta} || \mathbf{X}_{\alpha\beta\delta} \right) \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \delta_1 \\ \delta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\delta)_{11} \\ (\alpha\delta)_{12} \\ (\alpha\delta)_{21} \\ (\alpha\delta)_{22} \\ (\beta\delta)_{11} \\ (\beta\delta)_{12} \\ (\beta\delta)_{21} \\ (\beta\delta)_{22} \\ (\alpha\beta\delta)_{111} \\ (\alpha\beta\delta)_{121} \\ \vdots \\ (\alpha\beta\delta)_{222} \end{pmatrix} + \begin{pmatrix} e_{1111} \\ e_{1112} \\ e_{2111} \\ e_{2112} \\ e_{3111} \\ e_{3112} \\ e_{1211} \\ e_{1212} \\ e_{2211} \\ e_{2212} \\ e_{2211} \\ \vdots \\ e_{1232} \\ e_{2231} \\ e_{2232} \\ e_{3231} \\ e_{3232} \\ e_{1331} \\ e_{1332} \\ e_{2331} \\ e_{2332} \\ e_{3331} \\ e_{3332} \end{pmatrix},$$

Diseño y análisis de experimentos

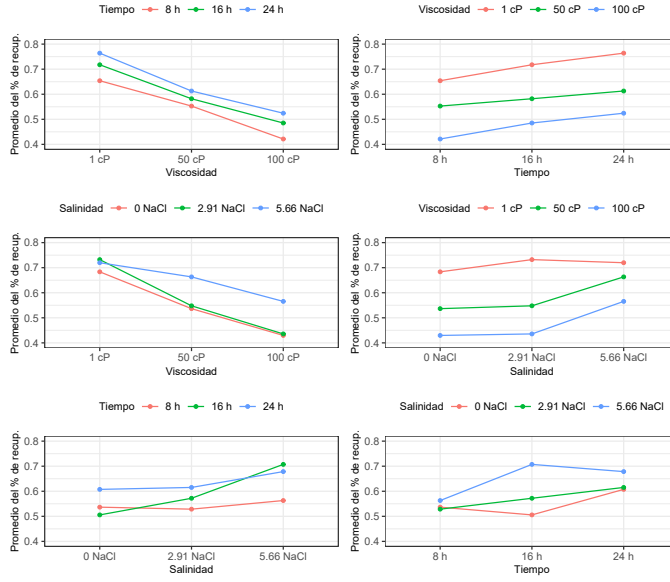


Figura 5.6. Promedios de porcentaje de recuperación de compuestos volátiles. Interacciones dobles

con

$$\mathbf{X}_\alpha = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix}, \mathbf{X}_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix}, \mathbf{X}_\delta = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{pmatrix}, \mathbf{X}_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 \\ 0 & -1 & 0 & -1 \\ -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

Diseños con arreglos factoriales

$$\mathbf{X}_{\alpha\delta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{X}_{\beta\delta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

$$\mathbf{X}_{\alpha\beta\delta} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}.$$

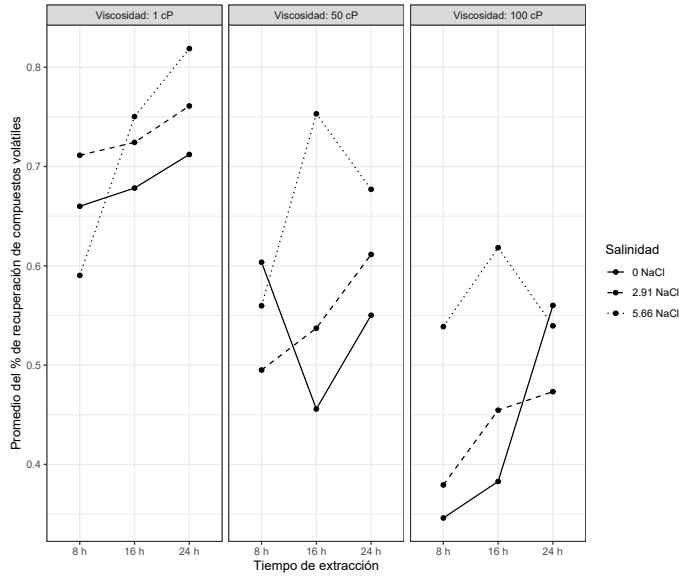


Figura 5.7. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple

Las instrucciones en R para los gráficos del ejemplo 5.7 son:

```
library(ggplot2); library(dplyr); library(readxl)
library(cowplot); library(dae)
remove(list=ls())
acuosas = read_excel("BasesDatosDE.xlsx",sheet="acuosas")
attach(acuosas)
acuosas$Viscosidad=factor(acuosas$Viscosidad,levels=c(0,1,2),
                           labels = c("1 cP","50 cP", "100 cP"))
acuosas$Salinidad=factor(acuosas$Salinidad,levels=c(0,1,2),
                           labels = c("0 NaCl","2.91 NaCl", "5.66 NaCl"))
acuosas$Tiempo=factor(acuosas$Tiempo,levels=c(1,2,3),
                       labels = c("8 h","16 h", "24 h"))
plot.design(recupera~Viscosidad*Tiempo*Salinidad,
            ylab="Promedio del % de recuperación",
```

Diseños con arreglos factoriales

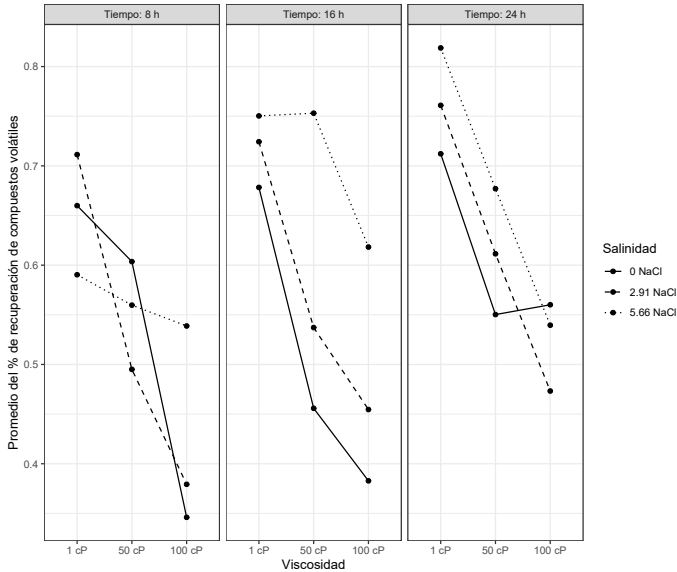


Figura 5.8. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple

```
xlab="Factores",cex.axis=1.2,cex.lab=1.4,data=acuosas)
datos.agr1 <- acuosas %>% mutate(
  Viscosidad = Viscosidad %>% as.factor,
  Tiempo=Tiempo%>%as.factor) %>% group_by(Viscosidad,Tiempo) %>%
  summarise(Promedio = mean(recupera, na.rm = TRUE))
datos.agr2 <- acuosas %>% mutate(
  Viscosidad = Viscosidad %>% as.factor,
  Salinidad = Salinidad %>% as.factor) %>%
  group_by(Viscosidad,Salinidad) %>%
  summarise(Promedio = mean(recupera, na.rm = TRUE))
datos.agr3 <- acuosas %>% mutate(Tiempo = Tiempo %>% as.factor,
  Salinidad=Salinidad%>%as.factor)%>%group_by(Salinidad,Tiempo)%>%
  summarise(Promedio = mean(recupera, na.rm = TRUE))
tema <- theme_bw()+ theme(axis.text=element_text(size=9),
  axis.title=element_text(size=9),
  legend.text=element_text(size=9),
  legend.title=element_text(size=9),legend.position="top")
```

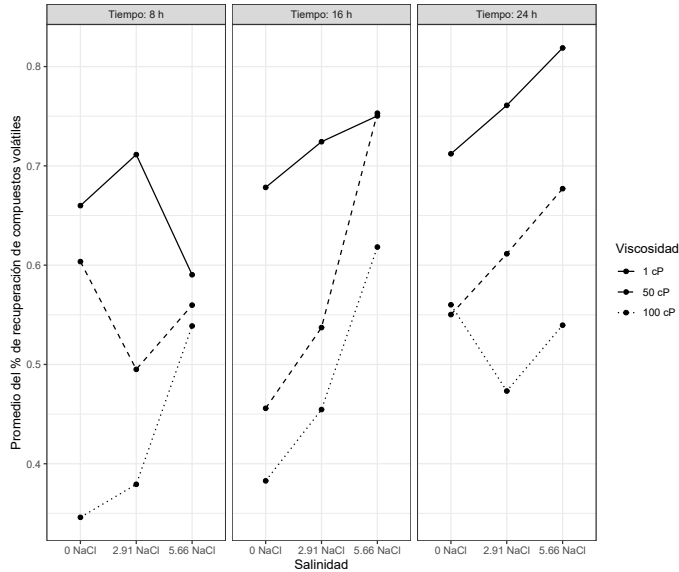


Figura 5.9. Promedios de porcentaje de recuperación de compuestos volátiles. Interacción triple

```
p1 <- ggplot(datos.agr1,aes(Viscosidad, Promedio)) +
  geom_line(aes(group = Tiempo, color = Tiempo)) +
  geom_point(size = 1.2, aes(color = Tiempo))+ylim(0.4,0.8)+
  ylab("Promedio del % de recup.")+tema
p2 <- ggplot(datos.agr1,aes(Tiempo, Promedio)) +
  geom_line(aes(group = Viscosidad, color = Viscosidad)) +
  geom_point(size = 1.2, aes(color = Viscosidad)) +
  ylim(0.4,0.8)+ylab("Promedio del % de recup.")+tema
p3 <- ggplot(datos.agr2,aes(Viscosidad, Promedio)) +
  geom_line(aes(group = Salinidad, color = Salinidad)) +
  geom_point(size = 1.2, aes(color = Salinidad))+
  ylab("Promedio del % de recup.")+ylim(0.4,0.8)+tema
p4 <- ggplot(datos.agr2,aes(Salinidad, Promedio)) +
  geom_line(aes(group = Viscosidad, color = Viscosidad)) +
  geom_point(size = 1.2, aes(color = Viscosidad))+
  ylab("Promedio del % de recup.")+ylim(0.4,0.8)+tema
p5 <- ggplot(datos.agr3,aes(Salinidad, Promedio)) +
```

```

geom_line(aes(group = Tiempo, color = Tiempo)) +
geom_point(size = 1.2, aes(color = Tiempo))+
ylab("Promedio del % de recup.")+ylim(0.4,0.8)+tema
p6 <- ggplot(datos.agr3,aes(Tiempo, Promedio)) +
geom_line(aes(group = Salinidad, color = Salinidad)) +
geom_point(size = 1.2, aes(color = Salinidad))+
ylab("Promedio del % de recup.")+ylim(0.4,0.8)+tema
plot_grid(p1,p2,p3,p4,p5,p6,nrow = 3)
a=c(1,1,1); b=c(1,2,3)
c="Promedio del % de recuperación de compuestos volátiles"
interaction.ABC.plot(recupera,Tiempo,Salinidad,Viscosidad,
data=acuosas,
ggplotFunc=list(theme_bw(),scale_color_manual(values=a),
scale_shape_manual(values=b),scale_linetype_manual(values=b),
labs(y=c, x="Tiempo de extracción")),cex.axis=1.2,cex.lab=1.5)
interaction.ABC.plot(recupera,Viscosidad,Salinidad,Tiempo,
data=acuosas,
ggplotFunc=list(theme_bw(),scale_color_manual(values=a),
scale_shape_manual(values=b),scale_linetype_manual(values=b),
labs(y=c, x="Viscosidad")),cex.axis=1.2,cex.lab=1.5)
interaction.ABC.plot(recupera,Salinidad,Viscosidad,Tiempo,
data=acuosas,
ggplotFunc=list(theme_bw(),scale_color_manual(values=a),
scale_shape_manual(values=b),scale_linetype_manual(values=b),
labs(y=c, x="Salinidad")),cex.axis=1.2,cex.lab=1.5)

```



4.1. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que hay siete fuentes de variación en el modelo que consisten en el ajuste por el factor *A*, el ajuste por el factor *B*, el ajuste por el factor *C*, el ajuste por la interacción de los factores *A* y *B*, el ajuste por la interacción de los factores *A* y *C*, el ajuste por la interacción de los factores *B* y *C*, y el ajuste por la interacción de los factores *A*, *B* y *C*. Lo anterior implica separar la varianza debida a las siete fuentes de variación.

Para identificar cuáles parámetros son estadísticamente significativos, se inicia por las interacciones de orden superior, esto es, primero se verifica la hipótesis:

$$H_0 : (\alpha\beta\delta)_{111} = (\alpha\beta\delta)_{112} = \cdots = (\alpha\beta\delta)_{abc} = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } (\alpha\beta\delta)_{ijk} \neq 0, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b$$

$$\text{y } k = 1, 2, \dots, c.$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de la interacción de los tres factores.

En caso de rechazar la hipótesis nula NO se procede a analizar el resto de las hipótesis, pero si se retira la interacción $(\alpha\beta\delta)_{ijk}$ del modelo, es válido revisar las hipótesis 1, 2 y 3.

1.

$$H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{ab} = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } (\alpha\beta)_{ij} \neq 0, \quad i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, b.$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de las interacción de los factores A y B . En caso de rechazar la hipótesis nula NO se procede a analizar las hipótesis 4 y 5.

2.

$$H_0 : (\alpha\delta)_{11} = (\alpha\delta)_{12} = \cdots = (\alpha\delta)_{ac} = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } (\alpha\delta)_{ik} \neq 0, \quad i = 1, 2, \dots, a \text{ y } k = 1, 2, \dots, c.$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de las interacción de los factores A y C . En caso de rechazar la hipótesis nula NO se procede a analizar las hipótesis 4 y 6.

3.

$$H_0 : (\beta\delta)_{11} = (\beta\delta)_{12} = \cdots = (\beta\delta)_{bc} = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } (\beta\delta)_{jk} \neq 0, \quad j = 1, 2, \dots, b \text{ y } k = 1, 2, \dots, c.$$

H_0 se rechazará si el P-valor asociado a las interacciones es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de las interacción de los factores B y C . En caso de rechazar la hipótesis nula NO se procede a analizar las hipótesis 5 y 6.

4.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{a-1} = \alpha_a = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \alpha_i \neq 0, \quad i = 1, 2, \dots, a.$$

H_0 se rechazará si el P-valor asociado al factor A es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor A en consideración.

5.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{b-1} = \beta_b = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \beta_j \neq 0, \quad j = 1, 2, \dots, b.$$

H_0 se rechazará si el P-valor asociado al factor B es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor B en consideración.

6.

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_{c-1} = \delta_c = 0 \quad \text{versus}$$

$$H_a : \text{Al menos un } \delta_k \neq 0, \quad k = 1, 2, \dots, c.$$

H_0 se rechazará si el P-valor asociado al factor C es menor a un nivel de significancia dado α en la tabla del análisis de varianza, tabla 5.8, y se concluirá que hay diferencia estadísticamente significativa en el efecto medio de los niveles del factor C en consideración.

En todos los casos se puede usar el criterio AIC para comparar entre los modelos.

Tabla 5.8. Análisis de varianza para un diseño con arreglos factoriales $A \times B \times C$. Diseño balanceado

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{SCFA}{a - 1}$	$\frac{CMFA}{CME}$
Factor B	$b - 1$	$\mathbf{Y}^\top \mathbf{P}_\beta \mathbf{Y}$	$\frac{SCFB}{b - 1}$	$\frac{CMFB}{CME}$
Factor C	$c - 1$	$\mathbf{Y}^\top \mathbf{P}_\delta \mathbf{Y}$	$\frac{SCFC}{c - 1}$	$\frac{CMFC}{CME}$
Interacción AB	$(a - 1)(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{(\alpha\beta)} \mathbf{Y}$	$\frac{SCAB}{(a - 1)(b - 1)}$	$\frac{CMAB}{CME}$
Interacción AC	$(a - 1)(c - 1)$	$\mathbf{Y}^\top \mathbf{P}_{(\alpha\delta)} \mathbf{Y}$	$\frac{SCAC}{(a - 1)(c - 1)}$	$\frac{CMAC}{CME}$
Interacción BC	$(b - 1)(c - 1)$	$\mathbf{Y}^\top \mathbf{P}_{(\beta\delta)} \mathbf{Y}$	$\frac{SCBC}{(b - 1)(c - 1)}$	$\frac{CMBC}{CME}$
Interacción ABC	d	$\mathbf{Y}^\top \mathbf{P}_{(\alpha\beta\delta)} \mathbf{Y}$	$\frac{SCABC}{d}$	$\frac{CMABC}{CME}$
Error	$abc(r - 1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}$	$\frac{SCE}{gle}$	
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$		

$n = a * b * c * r$, $d = (a - 1)(b - 1)(c - 1)$ y gle =grados de libertad del error.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^r y_{ijkl}$$

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \quad \mathbf{P}_\alpha = \mathbf{X}_\alpha(\mathbf{X}_\alpha^\top \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^\top \quad \mathbf{P}_\beta = \mathbf{X}_\beta(\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top$$

$$\mathbf{P}_\delta = \mathbf{X}_\delta(\mathbf{X}_\delta^\top \mathbf{X}_\delta)^{-1} \mathbf{X}_\delta^\top \quad \mathbf{P}_{(\alpha\beta)} = \mathbf{X}_{(\alpha\beta)}(\mathbf{X}_{(\alpha\beta)}^\top \mathbf{X}_{(\alpha\beta)})^{-1} \mathbf{X}_{(\alpha\beta)}^\top$$

$$\mathbf{P}_{(\alpha\delta)} = \mathbf{X}_{(\alpha\delta)}(\mathbf{X}_{(\alpha\delta)}^\top \mathbf{X}_{(\alpha\delta)})^{-1} \mathbf{X}_{(\alpha\delta)}^\top \quad \mathbf{P}_{(\beta\delta)} = \mathbf{X}_{(\beta\delta)}(\mathbf{X}_{(\beta\delta)}^\top \mathbf{X}_{(\beta\delta)})^{-1} \mathbf{X}_{(\beta\delta)}^\top$$

$$\mathbf{P}_{(\alpha\beta\delta)} = \mathbf{X}_{(\alpha\beta\delta)}(\mathbf{X}_{(\alpha\beta\delta)}^\top \mathbf{X}_{(\alpha\beta\delta)})^{-1} \mathbf{X}_{(\alpha\beta\delta)}^\top$$

Ejemplo 5.8. Continuando con el ejemplo sobre porcentaje de recuperación de compuestos volátiles, ejemplo 5.7, realice el análisis de varianza y concluya.

Solución. Inicialmente se realiza el análisis de varianza considerando todas las interacciones, ver tabla 5.9, en que se observa que el P-valor para la interacción triple es mayor a $p_S = 0.20$, por lo que se realiza el análisis de varianza retirando dicha interacción del modelo, ver tabla 5.10. En la práctica significa que sus grados de libertad y suma de cuadrados se añaden a los grados de libertad y sumas de cuadrados del error, respectivamente.

Tabla 5.9. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Viscosidad	2	0.4981	0.2491	34.680	<0.001
Tiempo	2	0.0751	0.0376	5.231	0.012
Salinidad	2	0.0986	0.0493	6.865	0.004
Viscosidad y tiempo	4	0.0048	0.0012	0.166	0.954
Viscosidad y salinidad	4	0.0388	0.0097	1.350	0.277
Tiempo y salinidad	4	0.0501	0.0125	1.743	0.170
Viscosidad, tiempo y salinidad	8	0.0753	0.0094	1.311	0.2800
Error	27	0.1939	0.0072		
Total	53	1.0347			

En la tabla 5.10 se observa que las interacciones viscosidad:tiempo y viscosidad:salinidad tienen sus P-valor mayores a $p_S = 0.20$ por lo que se realiza un análisis de varianza sin estas dos interacciones, ver tabla 5.11.

Tabla 5.10. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles sin interacción triple

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Viscosidad	2	0.4981	0.2491	32.379	<0.001
Tiempo	2	0.0751	0.0376	4.884	0.013
Salinidad	2	0.0986	0.0493	6.409	0.004
Viscosidad y tiempo	4	0.0048	0.0012	0.155	0.969
Viscosidad y salinidad	4	0.0388	0.0097	1.260	0.304
Tiempo y salinidad	4	0.0501	0.0125	1.627	0.189
Error	35	0.2692	0.0077		
Total	53	1.0347			

Tabla 5.11. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles considerando la interacción tiempo y salinidad

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Viscosidad	2	0.4981	0.2491	32.264	<0.001
Tiempo	2	0.0751	0.0376	5.165	0.010
Salinidad	2	0.0986	0.0493	6.778	0.003
Tiempo y salinidad	4	0.0501	0.0125	1.721	0.1629
Error	43	0.3128	0.0073		
Total	53	1.0347			

De la tabla 5.11 se observa que el P-valor asociado a la interacción tiempo y salinidad es menor que $p_S = 0.20$ pero mayor, por ejemplo, a 0.05, por lo cual para decidir si se deja en el modelo o no se usa el criterio AIC. Se encontró que con la interacción el valor del $AIC = -256.1721$ y sin la interacción es $AIC = -256.1538$, luego la diferencia entre los dos valores AIC es mínima y se opta por el modelo sin la interacción tiempo y salinidad, ver tabla 5.12.

Tabla 5.12. Análisis de varianza para los datos del ejemplo sobre porcentaje de recuperación de compuestos volátiles considerando solo efectos principales

Causa de variación	Grad. de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Viscosidad	2	0.4981	0.2491	32.264	<0.001
Tiempo	2	0.0751	0.0376	4.866	0.012
Salinidad	2	0.0986	0.0493	6.386	0.004
Error	47	0.3628	0.0077		
Total	53	1.0347			

De la tabla 5.12 se concluye que los factores viscosidad, tiempo y salinidad son significativos; sin embargo, al realizar una validación de supuestos se observa que el supuesto de normalidad se cumple, pero el supuesto de homogeneidad de varianzas, teniendo en cuenta los niveles de viscosidad y los niveles de salinidad, no se cumple. De igual forma en el envelope se observan puntos fuera de las bandas, ver figura 5.10.

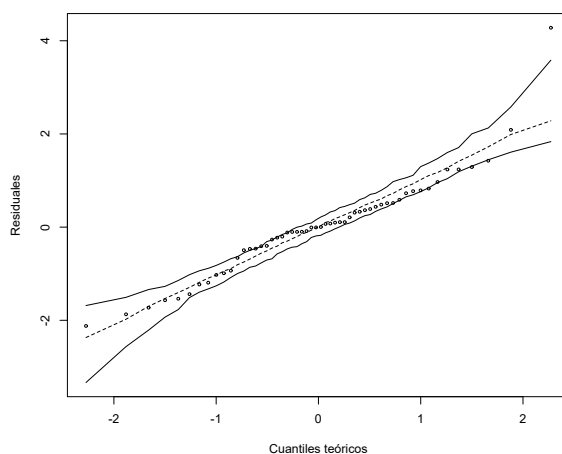


Figura 5.10. Envelope para el modelo ajustado a los datos del ejemplo 5.7

También se observa, ver figura 5.11, que hay tres residuales estudentizados, ver (3.15), mayores a 2 (en valor absoluto), en particular el residual asociado a

la observación 19, cuya porcentaje de recuperación es 79.41 % y corresponde a la viscosidad 50 cP, a la salinidad 0 g y al tiempo de extracción de 8 horas, es mayor a 3 por lo que se decide retirar esta observación del análisis y volver a realizar los análisis de varianza. Un resumen se presenta en la tabla 5.13.

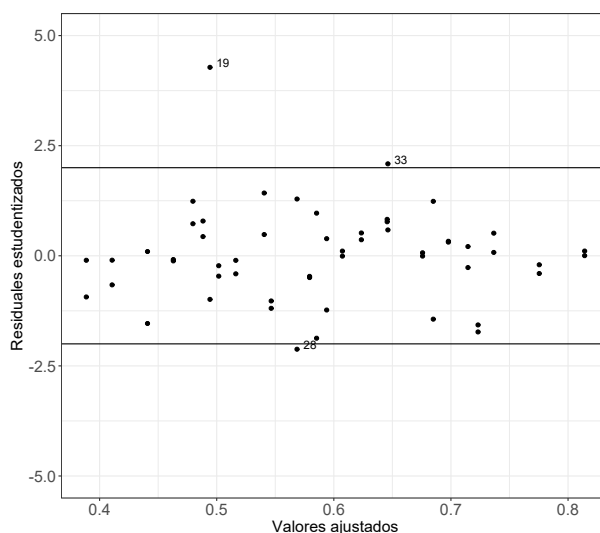


Figura 5.11. Valores ajustados versus residuales estudentizados para el modelo ajustado a los datos del ejemplo 5.7

Tabla 5.13. Modelos ajustados para los datos del ejemplo 5.7 con su respectivo *AIC*, sin la observación 19

El modelo	<i>AIC</i>
Considerando la interacción triple	-268.2186
Sin la interacción triple, pero con las interacciones dobles	-266.17
Viscosidad+Tiempo+Salinidad+Viscosidad*Salinidad+Tiempo*Salinidad	-273.6063
Viscosidad+Tiempo+Salinidad+Viscosidad*Salinidad	-271.9059

Finalmente, al realizar un envelope sobre el modelo se concluye que no hay indicios de violaciones de supuestos, ver figura 5.12.

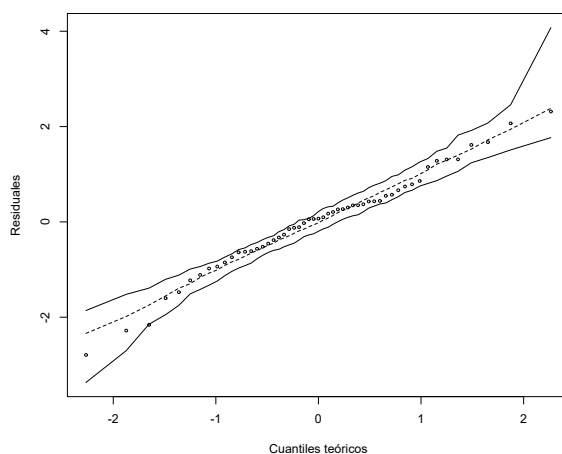


Figura 5.12. Envelope para el modelo ajustado a los datos del ejemplo 5.7 retirando la observación 19

El modelo seleccionado es:

$$\mathbb{E}(\% \text{ de recuperación}) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\beta\delta)}_{jk}. \quad (5.4)$$

Las respectivas medias esperadas se presentan en la tabla 5.14.

Tabla 5.14. Porcentajes esperados de recuperación de compuestos volátiles, modelo (5.4)

Tiempo de extracción	Viscosidad: 1 cP			Viscosidad: 50 cP			Viscosidad: 100 cP		
	Salinidad			Salinidad			Salinidad		
	0 g	2.91 g	5.66 g	0 g	2.91 g	5.66 g	0 g	2.91 g	5.66 g
8 h	0.6284	0.6888	0.6332	0.4190	0.5045	0.5767	0.3746	0.3923	0.4789
16 h	0.6601	0.7323	0.7775	0.4506	0.5480	0.7209	0.4062	0.4358	0.6232
24 h	0.7619	0.7755	0.7487	0.5525	0.5912	0.6922	0.5081	0.4790	0.5944

Las instrucciones en R para ajustar los modelos presentados en el ejemplo 5.8 son:

```
### correr instrucciones página 243 ###
library(MASS)
ajuste1 = aov(recupera~Viscosidad*Tiempo*Salinidad,
              data=acuosas)
```

```
summary(ajuste1);extractAIC(ajuste1)
ajuste2 = aov(recupera~Viscosidad+Tiempo+Salinidad+
              Viscosidad*Tiempo+Viscosidad*Salinidad+
              Tiempo*Salinidad, data=acuosas)
summary(ajuste2);extractAIC(ajuste2)
ajuste3 = aov(recupera~Viscosidad+Tiempo+Salinidad+
              Tiempo*Salinidad,data=acuosas)
summary(ajuste3);extractAIC(ajuste3)
ajuste4 = aov(recupera~Viscosidad+Tiempo+Salinidad,
              data=acuosas)
summary(ajuste4);extractAIC(ajuste4)
hnp::hnp(ajuste4,sim=99,conf=0.95,resid.type="student",
         halfnormal=FALSE)
## studres: residuales estudentizados
acuosas1 <- data.frame(acuosas,
                      Observación=as.numeric(rownames(acuosas)),
                      td=studres(ajuste4),Ajustados=ajuste4$fitted)
ggplot(acuosas1, aes(x=Ajustados,y=td))+geom_point()+
  theme_bw()+geom_hline(yintercept=c(-2,2))+
  geom_text(aes(label=ifelse(abs(td)>2,Observación,"")),
            hjust=-0.5,vjust=0)+
  ylim(-5,5)+xlab("Valores ajustados")+
  ylab("Residuales estudentizados")+
  theme(axis.title=element_text(size=15),
        axis.text=element_text(size=15))
# Retirando la observación 19 de la base de datos
acuosas2=acuosas[-19,]
# Volver a correr los modelos anteriores considerando
# data=acuosas2
ajuste7 = aov(recupera~Viscosidad+Tiempo+Salinidad+
              Viscosidad*Salinidad+Tiempo*Salinidad, data=acuosas2)
summary(ajuste7);extractAIC(ajuste7)
hnp::hnp(ajuste7,sim=99,conf=0.95,resid.type="student",
         halfnormal=FALSE)
```



5. Diseño con arreglos factoriales 2^k

Un diseño 2^k significa que hay k factores, en que cada uno tiene dos niveles, es decir, es un diseño $\underbrace{2 \times 2 \times 2 \cdots \times 2}_{k \text{ veces}}$. Se usan principalmente en experimentos de tamizaje o de selección de factores.

Los niveles, al igual que en los otros diseños, pueden ser cualitativos o cuantitativos, por ejemplo:

- Niveles cuantitativos:
 - Dos diferentes temperaturas.
 - Dos diferentes concentraciones.
- Niveles cualitativos:
 - Dos tipos de catalizador.
 - Presencia/ausencia de algún componente químico.

Entre las *ventajas* del diseño 2^k están que requiere pocas unidades experimentales e interpretación sencilla. Si no hay datos faltantes, es un diseño balanceado porque cada nivel del factor aparece con el mismo número de unidades experimentales y es un diseño ortogonal porque todas las combinaciones de niveles aparecen en el mismo número de unidades experimentales [61, p. 102].

En la tabla 5.15 se presentan algunas de las notaciones usadas para indicar los niveles y los tratamientos en diseños con arreglos factoriales con dos niveles, en que usualmente un nivel de factor es bajo y el otro nivel de factor es alto.

Nota 5.9. En la tabla 5.15 los niveles bajos de los factores A y B se denotan con el signo menos y los niveles altos se denotan con el signo más. Adicionalmente, la combinación de tratamientos (o corrida) se denota por la respectiva letra minúscula del factor que esté en nivel alto. La notación “(1)” indica los dos factores están en su nivel bajo.

Tabla 5.15. Notación de los tratamientos en un diseño con arreglos factoriales 2^2

Factor A	Factor B	
	Bajo(0)	Alto(1)
Bajo(0)	00	01
	- -	- +
	(1)	<i>b</i>
Alto(1)	10	11
	+ -	+ +
	<i>a</i>	<i>ab</i>

Ejemplo 5.9. Retomando el ejemplo 5.2 sobre producción de biodiésel en que se consideraron cuatro factores (relación molar aceite/alcohol - factor A, porcentaje en peso de catalizador - factor B, temperatura - factor C, y agente de lavado - factor D) y dos réplicas por cada uno, en un diseño completamente aleatorizado, ver [10], se midió el rendimiento del biodiésel bajo cada tratamiento. Datos **ilustrativos** se presentan en la tabla 5.16, teniendo en cuenta que la tesis no presenta los datos individuales, un gráfico de los promedios por nivel de factor en la figura 5.13 y gráficos de los promedios por nivel de factor cruzados dos a dos en la figura 5.14.

Tabla 5.16. Datos ejemplo 5.9 sobre el rendimiento de biodiésel, datos ilustrativos

Combinación de tratamientos	Réplica	
	I	II
(1)	69.5	77.5
<i>b</i>	88.0	95.3
<i>d</i>	80.5	79.9
<i>ac</i>	73.4	82.2
<i>bc</i>	79.6	86.9
<i>cd</i>	81.2	81.6
<i>abd</i>	75.2	65.3
<i>bcd</i>	75.2	76.5

Combinación de tratamientos	Réplica	
	I	II
<i>a</i>	81.9	74.1
<i>c</i>	83.2	82.7
<i>ab</i>	70.4	63.0
<i>ad</i>	80.9	78.5
<i>bd</i>	83.1	73.8
<i>abc</i>	85.1	83.5
<i>acd</i>	80.9	90.8
<i>abcd</i>	87.6	89.8

Diseños con arreglos factoriales

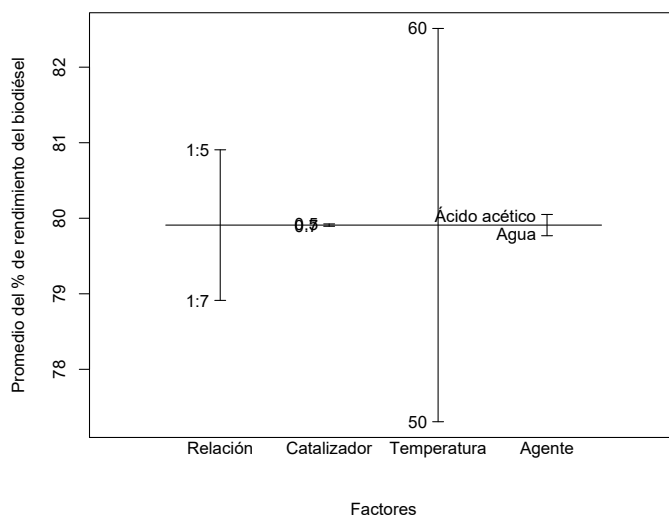


Figura 5.13. Promedios por niveles de factor - ejemplo 5.9

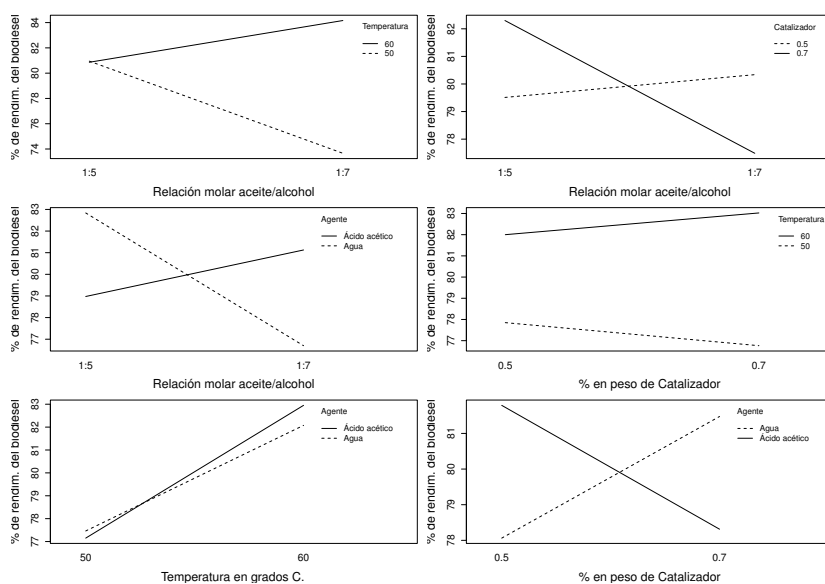


Figura 5.14. Promedios por niveles de factor cruzados dos a dos - ejemplo 5.9

Las instrucciones en R para realizar los gráficos del ejemplo 5.9 son:

```
library(readxl)
remove(list=ls())
biodiesel = read_excel("BasesDatosDE.xlsx",sheet="biodiesel")
attach(biodiesel)
Relacion=ordered(factor(Relacion,levels=c("-1","1"),
                        labels=c("1:5","1:7")))
Catalizador=ordered(factor(Catalizador,levels=c(-1,1),
                          labels=c("0.5","0.7")))
Temperatura=ordered(factor(Temperatura,levels=c(-1,1),
                          labels=c("50","60")))
Agente=ordered(factor(Agente,levels=c(-1,1),
                      labels=c("Agua","ácido acético")))
plot.design(Rendimiento~Relacion*Catalizador*Temperatura*Agente,
            ylab="Promedio del % de rendimiento del biodiésel",
            xlab="Factores")
a="% de rendim. del biodiésel"; b="Relación molar aceite/alcohol"
c="% en peso de Catalizador"
par(mar=c(4,4,1,1))
layout(matrix(c(1,2,3,4,5,6), 3, 2, byrow = TRUE))
interaction.plot(Relacion,Temperatura,Rendimiento,ylab=a,xlab=b,
                cex.axis=1.2,cex.lab=1.5,cex=2)
interaction.plot(Relacion,Catalizador,Rendimiento,ylab=a,xlab=b,
                cex.axis=1.2,cex.lab=1.5)
interaction.plot(Relacion,Agente,Rendimiento,ylab=a,xlab=b,
                cex.axis=1.2,cex.lab=1.5)
interaction.plot(Catalizador,Temperatura,Rendimiento,ylab=a,
                xlab=c,cex.axis=1.2,cex.lab=1.5)
interaction.plot(Temperatura,Agente,Rendimiento,ylab=a,
                xlab="Temperatura en grados C.",cex.axis=1.2,cex.lab=1.5)
interaction.plot(Catalizador,Agente,Rendimiento,ylab=a,xlab=c,
                cex.axis=1.2,cex.lab=1.5)
par(mfrow=c(1,1))
```



5.1. Diseño 2^2

Para asociar la notación de la tabla 5.15 a un modelo lineal, los subíndices de los parámetros asociados a los niveles de tratamiento serán 0 o 1. Así, el modelo es:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad i = 0, 1 \quad y \quad j = 0, 1,$$

con las restricciones:

$$\alpha_0 + \alpha_1 = 0 \implies \alpha_0 = -\alpha_1$$

$$\beta_0 + \beta_1 = 0 \implies \beta_0 = -\beta_1$$

$$(\alpha\beta)_{00} + (\alpha\beta)_{01} = 0 \implies (\alpha\beta)_{00} = -(\alpha\beta)_{01} \implies (\alpha\beta)_{00} = (\alpha\beta)_{11}$$

$$(\alpha\beta)_{10} + (\alpha\beta)_{11} = 0 \implies (\alpha\beta)_{10} = -(\alpha\beta)_{11}$$

$$(\alpha\beta)_{00} + (\alpha\beta)_{10} = 0 \quad y$$

$$(\alpha\beta)_{01} + (\alpha\beta)_{11} = 0 \implies (\alpha\beta)_{01} = -(\alpha\beta)_{11}.$$

Usando las restricciones adecuadamente, es decir, dejando en la matriz diseño los niveles altos $(\alpha_1, \beta_1 \text{ y } (\alpha\beta)_{11})$, el vector de parámetros estimado, $\hat{\beta}$, es dado por:

$$\hat{\beta} = \begin{pmatrix} \bar{Y}_{***} \\ \frac{(\bar{Y}_{1**} - \bar{Y}_{0**})}{2} \\ \frac{(\bar{Y}_{*1*} - \bar{Y}_{*0*})}{2} \\ \frac{(\bar{Y}_{00*} - \bar{Y}_{01*} - \bar{Y}_{10*} + \bar{Y}_{11*})}{2^2} \end{pmatrix}, \quad (5.5)$$

en que:

$$\bar{Y}_{***} = \frac{1}{2^2 r} \sum_{k=1}^r (Y_{00k} + Y_{01k} + Y_{10k} + Y_{11k})$$

$$\bar{Y}_{i**} = \frac{1}{2r} \sum_{k=1}^r (Y_{i0k} + Y_{i1k})$$

$$\bar{Y}_{*j*} = \frac{1}{2r} \sum_{k=1}^r (Y_{0jk} + Y_{1jk}) \quad y$$

$$\bar{Y}_{ijk} = \frac{1}{r} \sum_{k=1}^r Y_{ijk}.$$

La interpretación de los parámetros estimados es:

- $\hat{\mu} = \bar{Y}_{***}$ estima la media general.
- $(\widehat{\alpha\beta})_{11} = \frac{(\bar{Y}_{00*} - \bar{Y}_{01*} - \bar{Y}_{10*} + \bar{Y}_{11*})}{2^2} = \frac{(\bar{Y}_{11*} - \bar{Y}_{01*}) - (\bar{Y}_{10*} - \bar{Y}_{00*})}{2^2}$ estima el efecto de pasar del nivel alto de A al nivel bajo de A dado que el factor B está en su nivel alto menos el efecto de pasar del nivel alto al nivel bajo de A dado que el factor B está en su nivel bajo.
- $\hat{\alpha}_1 = \frac{(\bar{Y}_{1**} - \bar{Y}_{0**})}{2}$ estima el efecto medio del factor A al pasar del nivel alto al nivel bajo, por tanto, el efecto de pasar del nivel alto de A al nivel bajo de A es $2\hat{\alpha}_1$.
- $\hat{\beta}_1 = \frac{(\bar{Y}_{*1*} - \bar{Y}_{*0*})}{2}$ estima el efecto medio del factor B al pasar del nivel alto al bajo, por tanto, el efecto de pasar del nivel alto de B al nivel bajo de B es $2\hat{\beta}_1$.

El análisis de varianza se presenta en la tabla 5.17.

Tabla 5.17. Análisis de varianza para un diseño 2^2

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado
A	1	$\frac{1}{4r}(Y_{1**} - Y_{0**})^2$	SC_A	$\frac{SC_A}{CME}$
B	1	$\frac{1}{4r}(Y_{*1*} - Y_{*0*})^2$	SC_B	$\frac{SC_B}{CME}$
AB	1	$\frac{1}{4r}(Y_{00*} + Y_{11*} - Y_{10*} - Y_{01*})^2$	SC_{AB}	$\frac{SC_{AB}}{CME}$
Error	$2^2(r-1)$	Diferencia	SCE	
Total	2^2r-1	$\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^r Y_{ijk}^2 - 2^2r\bar{Y}^2$		

Ejemplo 5.10. Considere un experimento realizado en la Universidad de Cartagena sobre la efectividad de los procesos de oxidación en la disminución de plaguicidas organofosforados (clorpirifós) en que tuvieron en cuenta dos niveles de sales ferrosas (Fe^{+2} - factor A, niveles 2.5 ppm y 5 ppm) y dos niveles de peróxido de hidrógeno (H_2O_2 , factor

B, niveles 540 ppm y 1080 ppm) y midieron el porcentaje de degradación de clorpirifós [5]. Los datos se presentan en la tabla 5.18.

Realizar el análisis de varianza y concluir.

Tabla 5.18. Porcentaje de degradación de clorpirifós

A	B	Combinación de tratamientos	Réplica	
			I	II
-	-	(1)	88.2204	88.2283
+	-	a	84.57	83.14
-	+	b	76.95	77.47
+	+	ab	79.94	80.92

Fuente: [5, p. 72].

Solución. En la tabla 5.19 se presenta el análisis de varianza, del cual se concluye que la interacción entre los factores sal ferrosa y peróxido de hidrógeno es significativa para el porcentaje medio esperado de degradación de clorpirifós. Las medias estimadas para los 4 tratamientos son: 88.224 (tratamiento (1)), 83.855 (tratamiento a), 77.21 (tratamiento b) y 80.43 (tratamiento ab).



Tabla 5.19. Análisis de varianza sobre los datos de la degradación de clorpirifós

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Fe^{+2}	1	0.66	0.66	1.613	0.273
H_2O_2	1	104.25	104.25	254.591	<0.001
$Fe^{+2}:H_2O_2$	1	28.80	28.80	70.333	0.001
Error	4	1.64	0.41		
Total	7	135.35			

5.2. Diseños con arreglos factoriales 2^k con una sola réplica o no replicado

Se usan en experimentos de exploración cuando hay un número relativamente grande de factores y niveles bajo consideración. El principal problema de un diseño no replicado es que los grados de libertad del error son cero y, por tanto, no se cuenta con una estimación de la varianza del error, σ_e^2 .

En la literatura hay varias metodologías para analizar diseños con arreglos factoriales con una sola réplica, entre ellas están:

- Asumir que alguna(s) interacción(es) de orden superior no son significativas y usarlas para estimar la varianza del error. Esta solución tiene el inconveniente de retirar del modelo interacciones que no se sabe si realmente pueden ser importantes o no.
- Hacer un gráfico de probabilidad normal (Q-Q plot), ver sección 5.1, sobre los parámetros estimados, $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$, desconsiderando a μ . Los efectos que no son significativos siguen una distribución normal con media cero y varianza σ_e^2 y se localizarán sobre una recta. Los efectos que son significativos estarán lejos de la recta. Luego, las interacciones que no son significativas se pueden mandar al término de error [33, p. 441].
- Hacer un *half-normal plot*. Sea $|\hat{\theta}|_{(1)} \leq |\hat{\theta}|_{(2)} \leq \dots \leq |\hat{\theta}|_{(i)} \leq \dots \leq |\hat{\theta}|_{(k)}$, los valores ordenados de los valores absolutos de los efectos estimados sin considerar la media, μ . El método consiste en hacer un Q-Q plot, ver sección 5.1, pero sobre los valores absolutos de los efectos estimados. Los efectos importantes aparecen en la esquina superior derecha del gráfico.
- El **método de Lenth** [37], que consiste en:
 1. Calcular la mediana de los valores absolutos de los k efectos,

$$m = \text{med}|\theta_i|,$$

en que $\theta_i, i = 1, \dots, k$, es el efecto.

2. Calcular el pseudo error estándar:

$$s_0 = 1.5m.$$

3. Excluir los valores de $\hat{\theta}_i$ que exceden $2.5s_0$.
4. Recalcular m y el pseudo error estándar (PSE):

$$PSE = 1.5m.$$

5. Calcular el margen de error simultáneo:

$$SME = t_{\gamma,d} \times PSE,$$

en que $\gamma = \frac{1+0.95^{\frac{1}{k}}}{2}$ con $d = \frac{k}{3}$.

6. Todos los efectos $|\theta_i|$ mayores a SME se consideran **activos** o **significativos**.
7. Calcular el margen de error:

$$ME = t_{0.975;d} \times PSE.$$

8. Todos los efectos $|\theta_i|$ menores a ME se consideran **inertes**.
9. Los efectos restantes son “posiblemente” activos.

Ejemplo 5.11. Continuando con el ejemplo 5.2 sobre producción de biodiésel en que se consideraron cuatro factores: relación molar aceite/alcohol, porcentaje en peso de catalizador, temperatura y agente de lavado, y se midió el rendimiento del biodiésel bajo la combinación de estos factores (tratamientos), los investigadores realizaron dos réplicas, pero solo reportan en la tesis un valor por tratamiento, que se asume es el promedio de las mismas. Para efectos **ilustrativos** utilice los datos reportados en [10, p. 85] y realice el análisis como un diseño con arreglos factoriales no replicado. Los datos se presentan en la tabla 5.20. La relación molar aceite/alcohol (factor A), asóciela con el parámetro α ; el porcentaje en peso de catalizador (factor B), asócielo con el parámetro β ; la temperatura (factor C), asóciela con el parámetro γ ; y el agente de lavado (factor D), asócielo con el parámetro δ .

Tabla 5.20. Datos ejemplo 5.11

Combinación de tratamientos		Combinación de tratamientos		Combinación de tratamientos	
(1)	73	a	78.5	b	84.6
ab	71.9	c	80.8	ac	74.4
bc	76.8	abc	88.9	d	75
ad	80	bd	75.1	abd	68.5
cd	78.6	acd	80.7	bcd	74.3
abcd	88.1				

Fuente: [10, p. 85].

1. Identificar si la interacción de orden superior es significativa usando un gráfico de probabilidad normal (Q-Q plot) sobre los parámetros estimados, y si la interacción de orden superior NO es significativa, realizar el análisis de varianza enviando la interacción de orden superior al término de error.
2. Realizar un gráfico half-normal plot sobre los valores absolutos de los efectos estimados e identificar efectos importantes.
3. Usando el método de Lenth, identificar efectos activos (o significativos) y efectos inertes.

Solución

1. El vector de parámetros estimado, retirando el primer término que está asociado a μ , es:

$$\hat{\beta} = \begin{pmatrix} \hat{\alpha}_1 \rightarrow 1 \\ \hat{\beta}_1 \rightarrow 2 \\ \hat{\gamma}_1 \rightarrow 3 \\ \hat{\delta}_1 \rightarrow 4 \\ (\widehat{\alpha\beta})_{11} \rightarrow 5 \\ (\widehat{\alpha\gamma})_{11} \rightarrow 6 \\ (\widehat{\beta\gamma})_{11} \rightarrow 7 \\ (\widehat{\alpha\delta})_{11} \rightarrow 8 \\ (\widehat{\beta\delta})_{11} \rightarrow 9 \\ (\widehat{\gamma\delta})_{11} \rightarrow 10 \\ (\widehat{\alpha\beta\gamma})_{11} \rightarrow 11 \\ (\widehat{\alpha\beta\delta})_{11} \rightarrow 12 \\ (\widehat{\alpha\gamma\delta})_{11} \rightarrow 13 \\ (\widehat{\beta\gamma\delta})_{11} \rightarrow 14 \\ (\widehat{\alpha\beta\gamma\delta})_{111} \rightarrow 15 \end{pmatrix} = \begin{pmatrix} 0.8000 \\ 0.4500 \\ 2.2500 \\ -0.5375 \\ 0.0250 \\ 1.9000 \\ 1.2500 \\ 0.9875 \\ -1.4875 \\ 0.6375 \\ 3.7500 \\ -0.0125 \\ 0.2875 \\ 0.5625 \\ -0.8375 \end{pmatrix}.$$

Su gráfico de probabilidad normal se presenta en la figura 5.15. Los números dentro del gráfico están asociados a los efectos de tratamiento y los resaltados son los que se localizan fuera de la recta. En particular la interacción de orden superior, es decir, $(\alpha\beta\gamma\delta)$, número 15, está entre los efectos sobre la recta, por lo tanto se puede considerar que la interacción de orden superior no es significativa. También se observa que la interacción triple relación*catalizador* temperatura ($\alpha\beta\gamma$), número 11, es significativa.

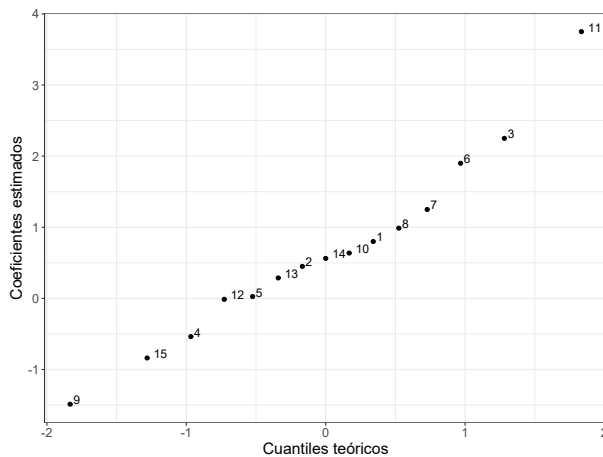


Figura 5.15. Gráfico de probabilidad normal de los efectos estimados para los datos del ejemplo 5.11

Se proceden a realizar los respectivos análisis de varianza sin la interacción $(\alpha\beta\gamma\delta)$, ver tabla 5.21. Se observa que todas las interacciones triples tienen P-valor mayor a 0.20 excepto la interacción relación*catalizador*temperatura, por lo cual se procede a retirarlas del análisis de varianza, ver tabla 5.22.

Tabla 5.21. Análisis de varianza sobre los datos del ejemplo 5.11, retirando la interacción cuádruple del modelo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Relación	1	10.24	10.24	0.912	0.515
Catalizador	1	3.24	3.24	0.289	0.686
Temperatura	1	81.00	81.00	7.218	0.227
Agente	1	4.62	4.62	0.412	0.637
Relación:catalizador	1	0.01	0.01	0.001	0.981
Temperatura:agente	1	6.50	6.50	0.579	0.586
Relación:temperatura	1	57.76	57.76	5.147	0.264
Catalizador:agente	1	35.40	35.40	3.155	0.326
Catalizador:temperatura	1	25.00	25.00	2.228	0.376
Relación:agente	1	15.60	15.60	1.390	0.448
Relación:catalizador:temperat.	1	225	225	20.049	0.140
Relación:catalizador:agente	1	0.0025	0.0025	0.0002	0.990
Relación:temperatura:agente	1	1.32	1.32	0.118	0.789
Catalizador:temperatura:agente	1	5.06	5.06	0.451	0.623
Error	1	11.22	11.22		
Total	15	482			

Tabla 5.22. Análisis de varianza sobre los datos del ejemplo 5.11 con la intracción triple relación:catalizador:temperatura

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Relación	1	10.24	10.24	2.326	0.202
Catalizador	1	3.24	3.24	0.736	0.439
Temperatura	1	81.00	81.00	18.399	0.013
Agente	1	4.62	4.62	1.050	0.363
Relación:catalizador	1	0.01	0.01	0.002	0.964
Temperatura:agente	1	6.50	6.50	1.477	0.291
Relación:temperatura	1	57.76	57.76	13.120	0.022
Catalizador:agente	1	35.40	35.40	8.041	0.047
Catalizador:temperatura	1	25.00	25.00	5.679	0.076
Relación:agente	1	15.60	15.60	3.544	0.133
Relación:catalizador:temperat.	1	225	225	51.107	0.002
Error	4	17.61	4.40		
Total	15	482			

De la tabla 5.22 se observa que la interacción triple relación*catalizador*temperatura es significativa. También se podría haber retirado la interacción doble temperatura*agente (P-valor mayor a 0.20 y el factor agente no está incluido en la interacción triple que se dejó en el modelo), sin embargo el modelo empeoraba de acuerdo al valor AIC, prefiriendo dejar esta interacción doble en el modelo. Finalmente ya no se retiran más efectos del modelo, teniendo en cuenta el principio de jerarquía. Vale resaltar que la interacción doble catalizador*agente es significativa por lo cual no se retira el factor “agente” del modelo.

2. Los valores absolutos de los 15 efectos, ordenados, son:

(1) $\widehat{(\alpha\beta\delta)}_{111}$ 0.0125	(2) $\widehat{(\alpha\beta)}_{11}$ 0.025	(3) $\widehat{(\alpha\gamma\delta)}_{111}$ 0.2875	(4) $\hat{\beta}_1$ 0.45	(5) $\hat{\delta}_1$ 0.5375	(6) $\widehat{(\beta\gamma\delta)}_{111}$ 0.5625
(7) $\widehat{(\gamma\delta)}_{11}$ 0.6375	(8) $\hat{\alpha}_1$ 0.8	(9) $\widehat{(\alpha\beta\gamma\delta)}_{1111}$ 0.8375	(10) $\widehat{(\alpha\delta)}_{11}$ 0.9875	(11) $\widehat{(\beta\gamma)}_{11}$ 1.25	(12) $\widehat{(\beta\delta)}_{11}$ 1.4875
(13) $\widehat{(\alpha\gamma)}_{11}$ 1.9	(14) $\hat{\gamma}_1$ 2.25	(15) $\widehat{(\alpha\beta\gamma)}_{111}$ 3.75			

El gráfico de z_i versus $|\hat{\theta}|_{(i)}$ se presenta en la figura 5.16. Del gráfico se concluye que el efecto importante es $\widehat{(\alpha\beta\gamma)}_{111}$.

3. Para el método de Lenth:

3.1) La mediana es $m = 0.8$ que corresponde a $\hat{\alpha}_{11}$.

3.2) El pseudo error estándar es:

$$s_0 = 1.5 \times m = 1.5 \times 0.8 = 1.2.$$

3.3) Se excluyen los valores mayores a $2.5s_0 = 2.5 \times 1.2 = 3$, es decir, se excluyen el valor asociado a $\widehat{\alpha\beta\gamma}_{111}$.

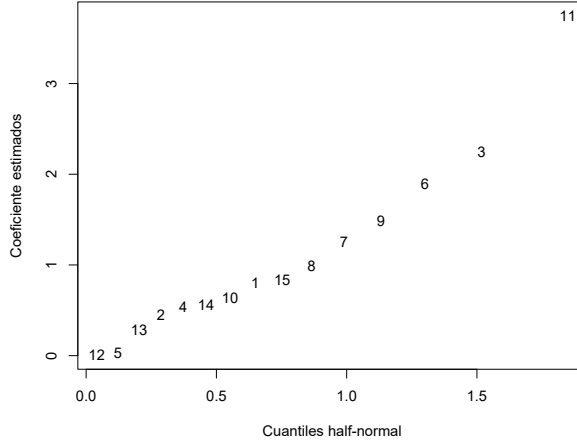


Figura 5.16. Half-normal plot de los efectos estimados para los datos del ejemplo 5.11

3.4) Se recalcula m y se calcula PSE .

$$m = \frac{(\widehat{\gamma\delta})_{11} + \hat{\alpha}_1}{2} = \frac{0.6375 + 0.8}{2} = 0.71875$$

$$PSE = 1.5 \times m = 1.5 \times 0.71875 = 1.0781.$$

3.5) Para el margen de error simultáneo, SME , se necesita

$$\gamma = \frac{1+0.95^{\frac{1}{k}}}{2} = \frac{1+0.95^{\frac{1}{15}}}{2} = 0.9983 \text{ y } d = \frac{k}{3} = \frac{15}{3} = 5:$$

$$SME = t_{\gamma,d} \times PSE = t_{0.9983,5} \times 1.0781 = 5.2235 \times 1.0781 = 5.6316.$$

3.6) No hay efectos **activos** o **significativos**, ya que ningún $|\hat{\theta}|_{(i)}$ tiene un valor mayor a 5.6316.

3.7) El margen de error es:

$$ME = t_{0.975;d} \times PSE = t_{0.975;5} \times 1.0781 = 2.5706 \times 1.0781 = 2.9654.$$

3.8) Luego todos los efectos se consideran inertes excepto $(\widehat{\alpha\beta\gamma})_{111}$.

- 3.9) Finalmente, el efecto $\widehat{(\alpha\beta\gamma)}_{111}$, que corresponde al efecto de la interacción entre la relación molar aceite/alcohol, el porcentaje en peso de catalizador y la temperatura es “posiblemente” activo, porque fue el único efecto que siendo menor de 5.6316 es mayor a 2.9654.

Las instrucciones en R para realizar el ejemplo 5.11 son:

```
library(MASS); library(ggplot2)
library(faraway); library(unrepx)
remove(list=ls())
R=c(-1,1,-1,-1,-1,1,1,1,-1,-1,-1,1,1,1,-1,1) #relación
C=c(-1,-1,1,-1,-1,1,-1,-1,1,1,-1,1,1,-1,1,1) # catalizador
T=c(-1,-1,-1,1,-1,-1,1,-1,1,-1,1,1,-1,1,1,1) # temperatura
A=c(-1,-1,-1,-1,1,-1,-1,1,-1,1,1,-1,1,1,1,1) # agente
REN=c(73,78.5,84.6,80.8,75,71.9,74.4,80,76.8,75.1,78.6,88.9,68.5,
      80.7,74.3,88.1) # rendimiento
datos=data.frame(R,C,T,A,REN)
ajuste1=aov(REN~R*C*T*A,data=datos)
coefic=ajuste1$coefficients
coef=data.frame(coeficientes=subset(coefic,
  names(coefic)!="(Intercept)"),Efecto=seq(length(coefic)-1))
p1<-ggplot(coef, aes(sample = coeficientes)) + stat_qq()
coef1<-ggplot_build(p1)$data[[1]]
coef1$Efecto<-coef$Efecto[order(coef$coeficientes)]
ggplot(coef1,aes(theoretical,sample,label=Efecto))+
  geom_text(hjust=-0.5,vjust=0,size=4)+geom_point()+
  theme_bw()+theme(text=element_text(size=15))
ajuste2=aov(REN~R+C+T+A+R*C+T*A+R*T+C*A+C*T+R*A+R*C*T+R*C*A+
  R*T*A+C*T*A)
summary(ajuste2)
extractAIC(ajuste2)
ajuste3=aov(REN~R+C+T+A+R*C+T*A+R*T+C*A+C*T+R*A+R*C*T)
summary(ajuste3)
extractAIC(ajuste3)
ajuste4=aov(REN~R+C+T+A+R*C+R*T+C*A+C*T+R*A+R*C*T)
summary(ajuste4)
extractAIC(ajuste4)
halfnorm(coef$coeficientes,nlab=15)#realiza el half-normal plot
```

```
PSE(coef$coeficientes,method ="Lenth") # pseudo error estándar
parplot(coef$coeficientes,method="Lenth") # Gráfico Pareto de
# los efectos de tratamiento estimados
# Observe que no coinciden los valores ME y SME con los
# obtenidos en el ejemplo porque la función parplot usa
# simulación para obtener los valores de t.
```



6. Diseños con arreglos factoriales 3^k

Los diseños con dos niveles por factor solo permiten estudiar efectos lineales para factores cuantitativos o comparar entre dos niveles si son factores cualitativos. En ocasiones es necesario establecer efectos de curvatura para factores cuantitativos o es necesario comparar más de dos niveles para factores cualitativos.

Un diseño 3^k significa que hay k factores, en que cada factor tiene tres niveles, es decir, es un diseño $\underbrace{3 \times 3 \times 3 \times \cdots \times 3}_{k \text{ veces}}$.

Al igual que los diseños 2^k , si no hay datos faltantes, el diseño es balanceado porque cada nivel del factor aparece con el mismo número de unidades experimentales, lo que lo hace un diseño ortogonal. La notación usual para la combinación de tratamientos está en la tabla 5.23.

Tabla 5.23. Notación de los tratamientos en un diseño con arreglos factoriales 3^2

Factor A	Factor B		
	Bajo(0)	Medio(1)	Alto(2)
Bajo(0)	00	01	02
Medio(1)	10	11	12
Alto(2)	20	21	22

Nota 5.10. Los principios fundamentales para efectos de diseños con arreglos factoriales vistos en la sección 5 también aplican para diseños 3^k , es decir, principios de orden jerárquico, principio de parsimonia de efectos y principio de heredabilidad de efectos [61, p. 111].

Ejemplo 5.12. Retomando el ejemplo 5.1 sobre la optimización de un sistema cortador de bagazo de caña de azúcar en que medían, entre otros, la energía específica de corte (J/mm^2) [27], bajo diferentes condiciones del ángulo de filo, la velocidad de corte y humedad, se va a considerar solo los dos primeros factores de forma **ilustrativa**, mientras que el porcentaje de humedad se va a dejar fijo en 30 % de base húmeda. Los investigadores tomaron 7 datos por tratamiento, los cuales no se presentan en forma individual en la tesis. Para efectos **ilustrativos**, se generaron 4 datos por tratamiento, teniendo en cuenta las medias y las desviaciones estándar reportadas en [27, p. 110].

Los datos generados se presentan en la tabla 5.24, gráficos boxplot en la figura 5.17, un gráfico de los promedios por nivel de factor en la figura 5.18 y promedios por niveles de factor cruzados en la figura 5.19.

Tabla 5.24. Energía específica de corte (J/mm^2) bajo 30 % de base húmeda - ejemplo 5.12

Velocidad de corte (m/s)	Ángulo de filo (grados)					
	20°		40°		60°	
2.3	0.048	0.067	0.030	0.041	0.045	0.042
	0.103	0.097	0.037	0.036	0.028	0.022
3.4	0.088	0.097	0.054	0.047	0.035	0.035
	0.108	0.077	0.052	0.034	0.022	0.043
4.5	0.086	0.110	0.041	0.039	0.023	0.034
	0.078	0.067	0.043	0.055	0.024	0.028

Fuente: datos simulados con base en los promedios y desviaciones estándar reportados en [27, p. 110].

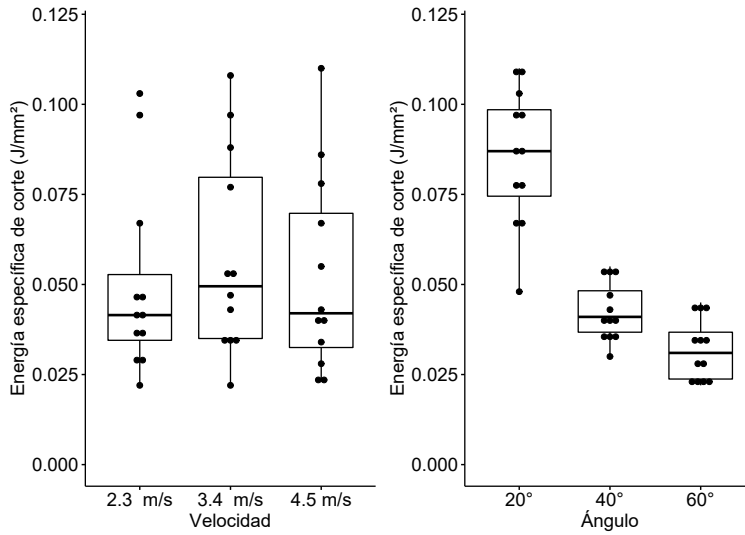


Figura 5.17. Boxplot según velocidad de corte y según ángulo de filo - ejemplo 5.12

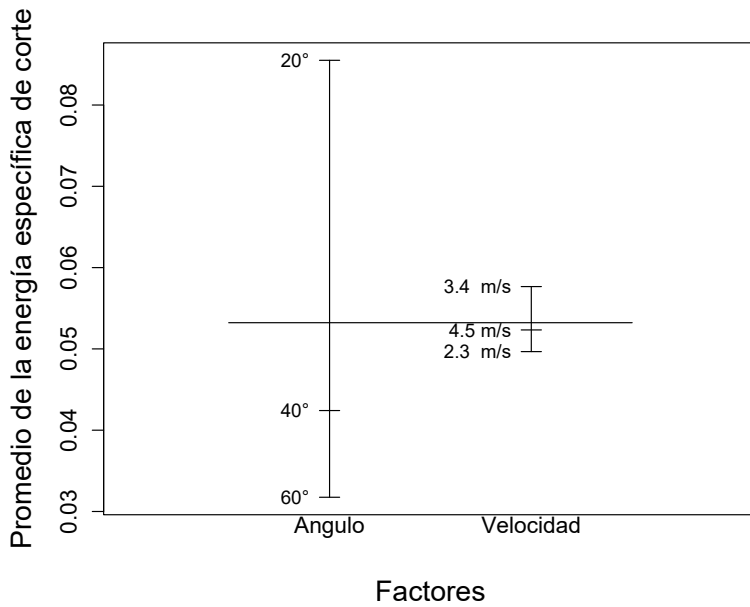


Figura 5.18. Promedios por niveles de factor - ejemplo 5.12

Diseños con arreglos factoriales

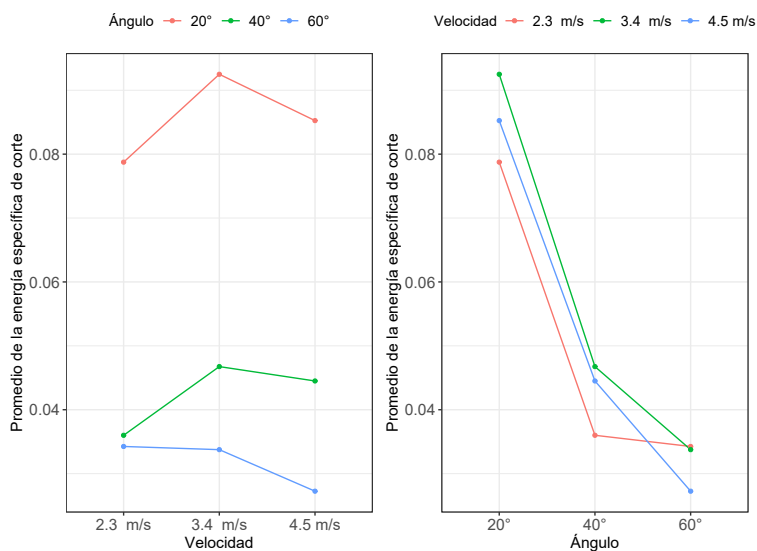


Figura 5.19. Promedios por niveles de factor cruzados - ejemplo 5.12

Las instrucciones en R para los gráficos presentados en el ejemplo 5.12 son:

```
library(ggplot2); library(ggpubr); library(dplyr)
library(cowplot); library(readxl)
remove(list=ls())
energy = read_excel("BasesDatosDE.xlsx",sheet="energia")
attach(energy)
Velocidad=factor(Velocidad,levels=c(2.3,3.4,4.5),
                 labels=c("2.3 m/s","3.4 m/s","4.5 m/s"))
Angulo=factor(angulo,levels=c(20,40,60),
              labels=c("20°","40°","60°"))
energy=data.frame(Velocidad,Angulo,energia)
tema <- theme(legend.position = "none",
              axis.text= element_text(size=15),
              axis.title = element_text(size=15))
a <- "Velocidad"; b <- "energia"
c <- "Angulo"; d <- "Energía específica de corte (J/mm²)"
p1 <- ggboxplot(energy,x=a,y=b,add="dotplot",shape=a,
               add.params=list(size=0.4))+labs(x=a,y=d)+ylim(0,0.12)+
  tema
```



```
p2 <- ggboxplot(energ,x=c,y=b,add="dotplot",shape=c,
  add.params=list(size=0.4))+labs(x="Ángulo",y=d)+
  ylim(0,0.12)+tema
plot_grid(p1,p2,nrow = 1)
plot.design(energia~Angulo*Velocidad,xlab="Factores",
  cex.axis=1.2,
  ylab="Promedio de la energía específica de corte",
  cex.lab=1.5)
datos.agr <- energ %>% mutate(
  Velocidad = Velocidad %>% as.factor,
  ángulo = Angulo %>% as.factor) %>%
  group_by(Velocidad,ángulo) %>%
  summarise(Promedio = mean(energia, na.rm = TRUE))
tema <- theme_bw()+ theme(axis.text = element_text(size=12),
  axis.title = element_text(size=12),
  legend.text=element_text(size=11),
  legend.title=element_text(size=11),legend.position="top")
p1 <- ggplot(datos.agr,aes(Velocidad, Promedio)) +
  geom_line(aes(group = ángulo, color = ángulo)) +
  geom_point(size = 1.2, aes(color = ángulo))+
  ylab("Promedio de la energía específica de corte")+tema
p2 <- ggplot(datos.agr,aes(ángulo, Promedio)) +
  geom_line(aes(group = Velocidad, color = Velocidad)) +
  geom_point(size = 1.2, aes(color = Velocidad))+
  ylab("Promedio de la energía específica de corte")+tema
plot_grid(p1,p2,nrow = 1)
```

El modelo para un diseño 3^2 es dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \quad (5.6)$$

con $i = 0, 1, 2, j = 0, 1, 2$ y $k = 1, 2, \dots, r$.

Bajo las restricciones:

$$\alpha_0 + \alpha_1 + \alpha_2 = 0 \Rightarrow \alpha_0 = -\alpha_1 - \alpha_2$$

$$\beta_0 + \beta_1 + \beta_2 = 0 \Rightarrow \beta_0 = -\beta_1 - \beta_2$$

$$\begin{cases} (\alpha\beta)_{00} + (\alpha\beta)_{01} + (\alpha\beta)_{02} = 0 \\ (\alpha\beta)_{10} + (\alpha\beta)_{11} + (\alpha\beta)_{12} = 0 \\ (\alpha\beta)_{20} + (\alpha\beta)_{21} + (\alpha\beta)_{22} = 0 \end{cases}$$

$$\begin{cases} (\alpha\beta)_{00} + (\alpha\beta)_{10} + (\alpha\beta)_{20} = 0 \\ (\alpha\beta)_{01} + (\alpha\beta)_{11} + (\alpha\beta)_{21} = 0 \\ (\alpha\beta)_{02} + (\alpha\beta)_{12} + (\alpha\beta)_{22} = 0, \end{cases}$$

se tiene que el modelo, matricialmente, es dado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E},$$

$$\begin{bmatrix} Y_{001} \\ \vdots \\ Y_{00r} \\ Y_{011} \\ \vdots \\ Y_{01r} \\ Y_{021} \\ \vdots \\ Y_{02r} \\ Y_{101} \\ \vdots \\ Y_{10r} \\ Y_{111} \\ \vdots \\ Y_{11r} \\ \vdots \\ Y_{221} \\ \vdots \\ Y_{22r} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \end{bmatrix} + \mathbf{E}.$$

Por tanto, el vector de parámetros estimado, $\hat{\beta}$, es dado por:

$$\hat{\beta} = \frac{1}{3^2 r} \begin{pmatrix} Y_{***} \\ 3Y_{1**} - Y_{***} \\ 3Y_{2**} - Y_{***} \\ 3Y_{*1*} - Y_{***} \\ 3Y_{*2*} - Y_{***} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} \quad (5.7)$$

con, por ejemplo:

$$d_1 = 3Y_{11*} - 3Y_{01*} - 3Y_{10*} - 3Y_{12*} - 3Y_{21*} + Y_{***}.$$

Ejemplo 5.13. Continuando con el ejemplo 5.12, sobre la optimización de un sistema cortador de bagazo de caña de azúcar, realice el análisis de varianza.

Solución. En la tabla 5.25 se presenta el análisis de varianza, del cual se concluye que la interacción entre velocidad de corte y ángulo de filo no es significativa. Se procede a ajustar el modelo sin la interacción, ver tabla 5.26. De esta tabla se concluye que solo es significativo el efecto del ángulo de filo. Luego se ajusta el modelo solo con este efecto, ver tabla 5.27.

Tabla 5.25. Análisis de varianza sobre los datos de energía específica de corte - ejemplo 5.12

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Velocidad de corte	2	0.00040	0.0002	1.156	0.330
Ángulo de filo	2	0.01944	0.0097	56.421	<0.0001
Velocidad:ángulo	4	0.00036	0.0001	0.522	0.721
Error	27	0.00465	0.0002		
Total	35	0.02485			

Vale la pena recordar que las conclusiones son válidas una vez se verifique el cumplimiento de los supuestos del modelo.

Tabla 5.26. Análisis de varianza sobre los datos de energía específica de corte
- ejemplo 5.12 sin considerar la interacción

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Velocidad de corte	2	0.00040	0.0002	1.232	0.306
Ángulo de filo	2	0.01944	0.0097	60.132	<0.0001
Error	31	0.00501	0.0002		
Total	35	0.02485			

Tabla 5.27. Análisis de varianza sobre los datos de energía específica de corte
- ejemplo 5.12, considerando en el modelo solo el ángulo de filo

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Ángulo de filo	2	0.01944	0.0097	59.3	<0.0001
Error	33	0.00541	0.0002		
Total	35	0.02485			

Las instrucciones en R para los análisis de varianza presentados en el ejemplo 5.13 son:

```
### correr instrucciones página 274 ###
ajuste1=aov(energia~Velocidad*Angulo,data=energ)
summary(ajuste1)
extractAIC(ajuste1)
ajuste2=aov(energia~Velocidad+Angulo,data=energ)
summary(ajuste2)
extractAIC(ajuste2)
ajuste3=aov(energia~Angulo,data=energ)
summary(ajuste3)
extractAIC(ajuste3)
hnp::hnp(ajuste3,halfnorma=FALSE)
```



En el caso de tener niveles cuantitativos, se puede considerar trabajar con polinomios ortogonales. Si los niveles son igualmente espaciados, la matriz diseño quedará construida de acuerdo a la tabla 5.28.

Tabla 5.28. Entradas matriz diseño con arreglos factoriales 3^2

Factor A	Factor B	γ_0	γ_L^A	γ_Q^A	γ_L^B	γ_Q^B	γ_{LL}^{AB}	γ_{LQ}^{AB}	γ_{QL}^{AB}	γ_{QQ}^{AB}
Bajo	Bajo	1	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{2}$	$\frac{-1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	$\frac{1}{6}$
Bajo	Medio	1	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	0	$\frac{-2}{\sqrt{6}}$	0	$\frac{2}{2\sqrt{3}}$	0	$\frac{-2}{6}$
Bajo	Alto	1	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{-1}{2}$	$\frac{-1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$\frac{1}{6}$
Medio	Bajo	1	0	$\frac{-2}{\sqrt{6}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	0	0	$\frac{2}{2\sqrt{3}}$	$\frac{-2}{6}$
Medio	Medio	1	0	$\frac{-2}{\sqrt{6}}$	0	$\frac{-2}{\sqrt{6}}$	0	0	0	$\frac{4}{6}$
Medio	Alto	1	0	$\frac{-2}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	0	0	$\frac{-2}{2\sqrt{3}}$	$\frac{-2}{6}$
Alto	Bajo	1	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{-1}{2}$	$\frac{1}{2\sqrt{3}}$	$\frac{-1}{2\sqrt{3}}$	$\frac{1}{6}$
Alto	Medio	1	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	0	$\frac{-2}{\sqrt{6}}$	0	$\frac{-2}{2\sqrt{3}}$	0	$\frac{-2}{6}$
Alto	Alto	1	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{2}$	$\frac{1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$\frac{1}{6}$

La matriz $\mathbf{X}^\top \mathbf{X}$ sería dada por:

$$\mathbf{X}^\top \mathbf{X} = r \begin{pmatrix} 3^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{X}^\top \mathbf{Y} = \begin{pmatrix} 3^2 r \bar{Y}_{***} \\ \frac{-Y_{0**} + Y_{2**}}{\sqrt{2}} \\ \frac{Y_{0**} - 2Y_{1**} + Y_{2**}}{\sqrt{6}} \\ \frac{Y_{*2*} - Y_{*0*}}{\sqrt{2}} \\ \frac{Y_{*0*} - 2Y_{*1*} + Y_{*2*}}{\sqrt{6}} \\ \frac{Y_{00*} + Y_{22*} - Y_{02*} - Y_{20*}}{2} \\ \vdots \end{pmatrix} = \begin{pmatrix} 3^2 r \bar{Y}_{***} \\ \frac{-3r \bar{Y}_{0**} + 3r \bar{Y}_{2**}}{\sqrt{2}} \\ \frac{3r \bar{Y}_{0**} - 6r \bar{Y}_{1**} + 3r \bar{Y}_{2**}}{\sqrt{6}} \\ \frac{3r \bar{Y}_{*2*} - 3r \bar{Y}_{*0*}}{\sqrt{2}} \\ \frac{3r \bar{Y}_{*0*} - 6r \bar{Y}_{*1*} + 3r \bar{Y}_{*2*}}{\sqrt{6}} \\ \frac{r \bar{Y}_{00*} + r \bar{Y}_{22*} - r \bar{Y}_{02*} - r \bar{Y}_{20*}}{2} \\ \vdots \end{pmatrix}, \quad (5.8)$$

por tanto

$$\begin{aligned}
 \hat{\gamma}_0 &= \bar{Y}_{***} \\
 \hat{\gamma}_L^A &= \frac{\bar{Y}_{2**} - \bar{Y}_{0**}}{\sqrt{2}} \\
 \hat{\gamma}_Q^A &= \frac{\bar{Y}_{0**} - 2\bar{Y}_{1**} + \bar{Y}_{2**}}{\sqrt{6}} \\
 \hat{\gamma}_L^B &= \frac{\bar{Y}_{*2*} - \bar{Y}_{*0*}}{\sqrt{2}} \\
 \hat{\gamma}_Q^B &= \frac{\bar{Y}_{*0*} - 2\bar{Y}_{*1*} + \bar{Y}_{*2*}}{\sqrt{6}}, \\
 \hat{\gamma}_{LL}^{AB} &= \frac{(\bar{Y}_{22*} - \bar{Y}_{20*}) - (\bar{Y}_{02*} - \bar{Y}_{00*})}{2}, \\
 &\text{etc.}
 \end{aligned} \tag{5.9}$$

En la tabla 5.29 se presenta el correspondiente análisis de varianza con causa de variación, grados de libertad, cuadrados medios y las respectivas esperanzas de los cuadrados medios.

Tabla 5.29. Análisis de varianza para un diseño con arreglos factoriales 3^2 con niveles cuantitativos e igualmente espaciados

Causa de variación	Grados de libertad	Cuadrado medio	$\mathbb{E}(CM)$
γ_L^A	1	$SC(\gamma_L^A)$	$\sigma_e^2 + 3r(\gamma_L^A)^2$
γ_Q^A	1	$SC(\gamma_Q^A)$	$\sigma_e^2 + 3r(\gamma_Q^A)^2$
γ_L^B	1	$SC(\gamma_L^B)$	$\sigma_e^2 + 3r(\gamma_L^B)^2$
γ_Q^B	1	$SC(\gamma_Q^B)$	$\sigma_e^2 + 3r(\gamma_Q^B)^2$
γ_{LL}^{AB}	1	$SC(\gamma_{LL}^{AB})$	$\sigma_e^2 + r(\gamma_{LL}^{AB})^2$
γ_{LQ}^{AB}	1	$SC(\gamma_{LQ}^{AB})$	$\sigma_e^2 + r(\gamma_{LQ}^{AB})^2$
γ_{QL}^{AB}	1	$SC(\gamma_{QL}^{AB})$	$\sigma_e^2 + r(\gamma_{QL}^{AB})^2$
γ_{QQ}^{AB}	1	$SC(\gamma_{QQ}^{AB})$	$\sigma_e^2 + r(\gamma_{QQ}^{AB})^2$
Error	$3^2(r-1)$	$\frac{SCE}{3^2(r-1)}$	σ_e^2
Total	$3^2r - 1$		

Ejemplo 5.14. Continuando con el ejemplo 5.12, presente el análisis de varianza considerando polinomios ortogonales y concluya.

Solución. En la tabla 5.30 se presenta el análisis de varianza considerando polinomios ortogonales. Se puede observar que solo los ajustes lineal y cuadrático para el efecto del ángulo de filo son significativos. Sin embargo, teniendo en cuenta que el P-valor para el término cuadrático de velocidad era menor de 0.20, pero mayor a 0.05, se compararon los respectivos AIC, encontrando que tiene menor valor AIC el modelo que solo considera ángulo de filo. La tabla 5.31 presenta el análisis de varianza solo con el efecto lineal y cuadrático del ángulo de filo. Un gráfico de dispersión y medias ajustadas se presenta en la figura 5.20.

Tabla 5.30. Análisis de varianza para los datos del ejemplo 5.12 considerando polinomios ortogonales

Causa de variación	Grados de libertad	Cuadrado medio	F-calculado	P-valor
$\gamma_{L}^{veloc.}$	1	0.0000427	0.248	>0.20
$\gamma_{Q}^{veloc.}$	1	0.0003556	2.064	0.1623
$\gamma_{L}^{ángulo}$	1	0.0173344	100.640	<0.001
$\gamma_{Q}^{ángulo}$	1	0.0021017	12.202	<0.01
γ_{LL}^{va}	1	0.0001823	1.058	>0.20
γ_{LQ}^{va}	1	0.0001021	0.593	>0.20
γ_{QL}^{va}	1	0.0000750	0.435	>0.20
γ_{QQ}^{va}	1	0.0000001	0.001	>0.20
Error	27	0.000172		
Total	35			

Tabla 5.31. Análisis de varianza para los datos del ejemplo 5.12 considerando polinomios ortogonales, solo con el efecto del ángulo de filo

Causa de variación	Grados de libertad	Cuadrado medio	F-calculado	P-valor
$\gamma_{L}^{ángulo}$	1	0.0173344	105.77	<0.001
$\gamma_{Q}^{ángulo}$	1	0.0021017	12.82	0.001
Error	33	0.000164		
Total	35			

Diseños con arreglos factoriales

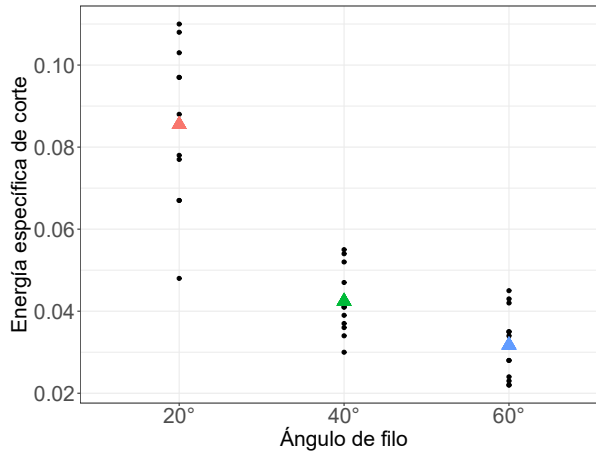


Figura 5.20. Gráfico de dispersión (puntos) y valores ajustados (triángulos) de acuerdo al modelo presentado en la tabla 5.31 - ejemplo 5.12

Las instrucciones en R para ajustar el modelo y realizar la figura 5.20 son:

```
### correr instrucciones página 274 ###
library(MASS); options(scipen=999)
contrasts(energ$Angulo)=contr.poly(levels(energ$Angulo))
contrasts(energ$Velocidad)=contr.poly(levels(energ$Velocidad))
ajuste4 = lm(energia~Angulo+Velocidad+Angulo*Velocidad,
             data=energ)
summary(ajuste4); contrasts(energ$Angulo)
contrasts(energ$Velocidad); ajuste4$coefficients
ajuste5 = lm(energia~Angulo+Velocidad,data=energ)
summary(ajuste5)
ajuste6 = lm(energia~Angulo,data=energ)
summary(ajuste6)
extractAIC(ajuste5);extractAIC(ajuste6)
X=model.matrix(ajuste6); ajuste6= lm(energia~X-1)
datos=data.frame(energia,Angulo,medias.ajus=ajuste6$fitted)
ggplot(datos, aes(x=Angulo, y=energia)) + geom_point()+
  geom_point(aes(y=medias.ajus,size=1.1,col=Angulo),shape=17)+
  theme_bw()+ labs(y="Energía específica de corte",
                  x="ángulo de filo")+
  
```



```
theme(legend.position = "none",axis.text=element_text(size=20),
      axis.title=element_text(size=20))
```



6.1. Contrastes ortogonales en diseños con arreglos factoriales 3^k

En un diseño con arreglos factoriales con tres niveles por factor, cada efecto principal tiene 2 grados de libertad asociados, cada interacción de dos factores tiene 4 grados de libertad asociados, cada interacción de tres niveles tiene 8 grados de libertad asociados, etc. Esto implica que se pueden encontrar dos contrastes ortogonales para cada efecto principal, cuatro contrastes ortogonales para cada interacción de dos factores, ocho contrastes ortogonales para cada interacción de tres factores, etc. Si los factores tienen niveles cuantitativos se puede usar polinomios ortogonales, y si son factores con niveles cualitativos se pueden encontrar contrastes ortogonales. Por ejemplo para un diseño 3^2 , con niveles cuantitativos igualmente espaciados, en la tabla 5.28 se presentan los contrastes. Para un diseño 3^2 con niveles cualitativos, un ejemplo de contrastes ortogonales se muestra en la tabla 5.32.

Tabla 5.32. Ejemplo de entradas matriz diseño en un diseño con arreglos factoriales 3^2 con niveles cualitativos

Factor A	Factor B	μ	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
a_0	b_0	1	1	1	1	1	1	1	1	1
a_0	b_1	1	1	1	-1	1	0	-2	-1	1
a_0	b_2	1	1	1	0	-2	-1	1	0	-2
a_1	b_0	1	-1	1	1	1	0	-2	0	-2
a_1	b_1	1	-1	1	-1	1	-1	1	1	1
a_1	b_2	1	-1	1	0	-2	1	1	-1	1
a_2	b_0	1	0	-2	1	1	-1	1	-1	1
a_2	b_1	1	0	-2	-1	1	1	1	0	-2
a_2	b_2	1	0	-2	0	-2	0	-2	1	1

Cabe notar que en la tabla 5.32, los contrastes C_1 y C_2 corresponden a los dos grados de libertad del factor A ; los contrastes C_3 y C_4 , a los dos grados de libertad del factor B , y los contrastes C_5 , C_6 , C_7 y C_8 , a los cuatro grados de libertad de la interacción $A \times B$.

Para la construcción de C_1 y C_2 ver (5.11) y para la construcción de C_3 y C_4 ver (5.12). Para la construcción de C_5 , C_6 , C_7 y C_8 , sean x_1 y x_2 denotando los subíndices de los niveles del factor A y B , respectivamente, es decir, $x_1 = 0, 1, 2$ y $x_2 = 0, 1, 2$, y sean:

$$\begin{aligned} AB &= x_1 + x_2 \quad y \\ AB^2 &= x_1 + 2x_2, \end{aligned} \tag{5.10}$$

juntos en módulo 3. Los contrastes se pueden ver en (5.13) y (5.14).

$$\text{Factor } A \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & a_0 & a_1 & a_2 \\ \hline C_1 & 1 & -1 & 0 \\ \hline C_2 & 1 & 1 & -2 \\ \hline \end{array} \right. \tag{5.11}$$

$$\text{Factor } B \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & b_0 & b_1 & b_2 \\ \hline C_3 & 1 & -1 & 0 \\ \hline C_4 & 1 & 1 & -2 \\ \hline \end{array} \right. \tag{5.12}$$

$$\begin{array}{l} \text{factor } AB \\ \text{usando (5.10)} \end{array} \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & x_1 + x_2 = 0 & x_1 + x_2 = 1 & x_1 + x_2 = 2 \\ \hline C_5 & 1 & 0 & -1 \\ \hline C_6 & 1 & -2 & 1 \\ \hline \end{array} \right. \tag{5.13}$$

$$\begin{array}{l} \text{factor } AB^2 \\ \text{usando (5.10)} \end{array} \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & x_1 + 2x_2 = 0 & x_1 + 2x_2 = 1 & x_1 + 2x_2 = 2 \\ \hline C_7 & 1 & 0 & -1 \\ \hline C_8 & 1 & -2 & 1 \\ \hline \end{array} \right. \tag{5.14}$$

entonces:

Factor A	Factor B	x_1	x_2	$AB : x_1 + x_2$	$AB^2 : x_1 + 2x_2$
a_0	b_0	0	0	$0 = 0 \text{ (mód 3)}$	$0 = 0 \text{ (mód 3)}$
a_0	b_1	0	1	$1 = 1 \text{ (mód 3)}$	$2 = 2 \text{ (mód 3)}$
a_0	b_2	0	2	$2 = 2 \text{ (mód 3)}$	$1 = 4 \text{ (mód 3)}$
a_1	b_0	1	0	$1 = 1 \text{ (mód 3)}$	$1 = 1 \text{ (mód 3)}$
a_1	b_1	1	1	$2 = 2 \text{ (mód 3)}$	$0 = 3 \text{ (mód 3)}$
a_1	b_2	1	2	$0 = 3 \text{ (mód 3)}$	$2 = 5 \text{ (mód 3)}$
a_2	b_0	2	0	$2 = 2 \text{ (mód 3)}$	$2 = 2 \text{ (mód 3)}$
a_2	b_1	2	1	$0 = 3 \text{ (mód 3)}$	$1 = 4 \text{ (mód 3)}$
a_2	b_2	2	2	$1 = 4 \text{ (mód 3)}$	$0 = 6 \text{ (mód 3)}$

Nota 5.11. Hay más alternativas para la construcción de $C_1, C_2, C_3, C_4, C_5, C_6, C_7$ y C_8 , por ejemplo (5.15).

$$\begin{aligned}
 \text{Factor A} & \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & a_0 & a_1 & a_2 \\ \hline C_1 & 0 & 1 & -1 \\ \hline C_2 & -2 & 1 & 1 \\ \hline \end{array} \right. \\
 \text{Factor B} & \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & b_0 & b_1 & b_2 \\ \hline C_3 & 1 & 0 & -1 \\ \hline C_4 & 1 & -2 & 1 \\ \hline \end{array} \right. \\
 \text{factor } AB \\ \text{usando (5.10)} & \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & x_1 + x_2 = 0 & x_1 + x_2 = 1 & x_1 + x_2 = 2 \\ \hline C_5 & 1 & -1 & 0 \\ \hline C_6 & 1 & 1 & -2 \\ \hline \end{array} \right. \\
 \text{factor } AB^2 \\ \text{usando (5.10)} & \left\{ \begin{array}{|c|c|c|c|} \hline \text{Contraste} & x_1 + 2x_2 = 0 & x_1 + 2x_2 = 1 & x_1 + 2x_2 = 2 \\ \hline C_7 & 1 & -1 & 0 \\ \hline C_8 & 1 & 1 & -2 \\ \hline \end{array} \right.
 \end{aligned}
 \tag{5.15}$$

Nota 5.12. Para tener en cuenta:

- Observe que los contrastes son ortogonales.
- La notación para $x_1 + x_2$ es AB , la notación para $x_1 + 2x_2$ es AB^2 y la notación para la interacción entre los factores A y B es $A \times B$.

- C_7 y C_8 en (5.14) se pueden obtener usando $2x_1 + x_2$, ver (5.16).

$$\text{factor } A^2B \left\{ \begin{array}{c|c|c|c} \text{Contr.} & 2x_1 + x_2 = 0 & 2x_1 + x_2 = 1 & 2x_1 + x_2 = 2 \\ \hline C_7 & 1 & -1 & 0 \\ \hline C_8 & 1 & 1 & -2 \end{array} \right. \quad (5.16)$$

porque

$$(AB^2)^2 = A^2B^4 = A^2B. \quad (5.17)$$

Dada (5.17), la recomendación es usar siempre la primera letra con exponente 1, por ejemplo:

$$\begin{aligned} CDE^2 &\Rightarrow x_3 + x_4 + 2x_5, & AB^2C &\Rightarrow x_1 + 2x_2 + x_3, \\ AB^2C^2 &\Rightarrow x_1 + 2x_2 + 2x_3, & BD^2 &\Rightarrow x_2 + 2x_4, \\ AC^2E &\Rightarrow x_1 + 2x_3 + x_5, & AB^2E^2 &\Rightarrow x_1 + 2x_2 + 2x_5. \end{aligned}$$

- Los componentes AB y AB^2 no tienen interpretabilidad en sí, pero son útiles para ganar grados de libertad en el error cuando no hay réplicas del diseño con arreglos factoriales y para diseñar sistemas de confusión capítulo 6.
- La descomposición de un diseño 3^3 en sus 13 componentes, cada una con dos grados de libertad, es dada por:

$$\begin{array}{llll} A & \Rightarrow x_1, & AC & \Rightarrow x_1 + x_3, & ABC & \Rightarrow x_1 + x_2 + x_3 \\ B & \Rightarrow x_2, & AC^2 & \Rightarrow x_1 + 2x_3, & ABC^2 & \Rightarrow x_1 + x_2 + 2x_3 \\ C & \Rightarrow x_3, & BC & \Rightarrow x_2 + x_3, & AB^2C & \Rightarrow x_1 + 2x_2 + x_3 \\ AB & \Rightarrow x_1 + x_2, & BC^2 & \Rightarrow x_2 + 2x_3, & AB^2C^2 & \Rightarrow x_1 + 2x_2 + 2x_3 \\ AB^2 & \Rightarrow x_1 + 2x_2. & & & & \end{array}$$

Ejemplo 5.15. Continuando con el ejemplo 5.12, realizar el análisis de varianza considerando el conjunto de contrastes ortogonales presentado en la tabla 5.32.

Solución. En la tabla 5.33 se presenta el análisis de varianza, en que el factor A es la velocidad de corte y el factor B es al ángulo de filo. Se puede observar que solo C_3 y C_4 son contrastes significativos, indicando que el ángulo de filo de

20 grados genera un efecto medio en la energía específica de corte diferente al ángulo de filo de 40 grados, y el efecto medio de estos dos ángulos genera un efecto medio diferente en la energía específica de corte con respecto al ángulo de filo 60 grados.

Tabla 5.33. Análisis de varianza para los datos del ejemplo 5.12 considerando contrastes ortogonales

Causa de variación	Grados de libertad	Cuadrado medio	F-calculado	P-valor
C_1	1	0.000384	2.229	>0.10
C_2	1	0.000014	0.083	> 0.20
C_3	1	0.011137	64.660	<0.0001
C_4	1	0.008299	48.183	<0.0001
C_5	1	0.000057	0.331	> 0.20
C_6	1	0.000031	0.178	> 0.20
C_7	1	0.000003	0.020	> 0.20
C_8	1	0.000268	1.558	> 0.20
Error	27	0.000172		
Total	35			

Las instrucciones en R usadas en el ejemplo 5.15 son:

```
### correr instrucciones página 274 ###
### hasta el attach(energ) ####
Velocidad=ifelse(Velocidad==2.3,0,ifelse(Velocidad==3.4,1,2))
Angulo=ifelse(angulo==20,0,ifelse(angulo==40,1,2))
av=(Velocidad+Angulo)%%3 # módulo 3
av2=(Velocidad+2*Angulo)%%3 # módulo 3
Velocidad=factor(Velocidad); Angulo=factor(Angulo);
av=factor(av); av2=factor(av2)
contraste1= cbind(c(1,-1,0),c(1,1,-2))
contraste2= cbind(c(1,0,-1),c(1,-2,1))
contrasts(Velocidad) <- contrasts(Angulo) <- contraste1
contrasts(av) <- contrasts(av2) <-contraste2
ajuste=summary.lm(aov(energia~Velocidad+Angulo+av+av2))
ajuste$coef[,3]^2 # F-calculado
```



Nota 5.13. Para el caso de diseños con arreglos factoriales 3^k con una sola réplica, o no replicados, es importante tener en cuenta que, al tener más de un grado de libertad por efecto, se puede dividir las interacciones en sus respectivos componentes de un grado de libertad y luego, sí proceder a establecer cuál componente mandar al error.

Ejemplo 5.16. Como ejemplo **ilustrativo**, suponga que se tiene un diseño con arreglos factoriales 3^2 con una sola réplica. Los dos factores a tener en cuenta son el factor A (C_1 y C_2) y el factor B (C_3 y C_4). El conjunto de datos se presenta en la tabla 5.34. Considere los contrastes ortogonales presentados en la tabla 5.32 y realice un análisis de varianza.

Tabla 5.34. Datos ejemplo ilustrativo 5.16

	Factor B		
Factor A	b_0	b_1	b_2
a_0	236	298	302
a_1	225	285	250
a_2	236	270	260

Solución. El modelo matricialmente escrito es:

$$Y = X\beta + E,$$

$$\begin{bmatrix} 236 \\ 298 \\ 302 \\ 225 \\ 285 \\ 250 \\ 236 \\ 270 \\ 260 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 0 & -2 & -1 & 1 \\ 1 & 1 & 1 & 0 & -2 & -1 & 1 & 0 & -2 \\ 1 & -1 & 1 & 1 & 1 & 0 & -2 & 0 & -2 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 & -2 & 1 & 1 & -1 & 1 \\ 1 & 0 & -2 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 0 & -2 & -1 & 1 & 1 & 1 & 0 & -2 \\ 1 & 0 & -2 & 0 & -2 & 0 & -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{bmatrix} + E.$$

El vector de parámetros estimado, $\hat{\beta}$, es:

$$\hat{\beta} = \begin{pmatrix} \hat{\mu} \\ \hat{C}_1 \rightarrow 1 \\ \hat{C}_2 \rightarrow 2 \\ \hat{C}_3 \rightarrow 3 \\ \hat{C}_4 \rightarrow 4 \\ \hat{C}_5 \rightarrow 5 \\ \hat{C}_6 \rightarrow 6 \\ \hat{C}_7 \rightarrow 7 \\ \hat{C}_8 \rightarrow 8 \end{pmatrix} = \begin{pmatrix} 262.4444 \\ 12.6667 \\ 3.5556 \\ -26 \\ -4.1111 \\ -11.1667 \\ 0.7222 \\ -0.5 \\ -1.6111 \end{pmatrix}. \quad (5.18)$$

Un gráfico half-normal sobre los efectos estimados se presenta en la figura 5.21. Del gráfico se observa que, en particular, los contrastes C_7 y C_8 , es decir, el componente AB^2 de la interacción $A \times B$, se pueden asumir no significativos, por lo cual se mandan al término de error.

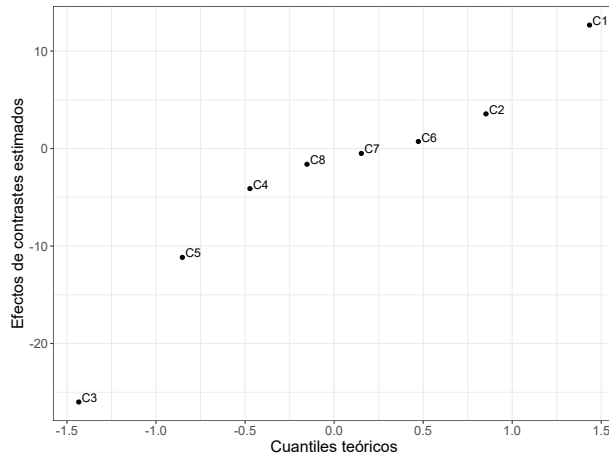


Figura 5.21. Half-normal plot de los efectos estimados para los datos del ejemplo 5.16

El análisis de varianza usando los contrastes se presenta en la tabla 5.35 y el análisis de varianza presentando los factores A y B en la tabla 5.36.

Tabla 5.35. Análisis de varianza usando contrastes para los datos del ejemplo 5.16 sin considerar los contrastes C_7 y C_8

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
C_1	1	963	963	39.926	0.02414
C_2	1	228	228	9.438	0.09163
C_3	1	4056	4056	168.221	0.00589
C_4	1	304	304	12.618	0.07093
C_5	1	748	748	31.030	0.03075
C_6	1	9	9	0.389	0.59630
Error(C_7 y C_8)	2	48	24		
Total	8	6356			

Tabla 5.36. Análisis de varianza para los datos del ejemplo 5.16 sin considerar los contrastes C_6 , C_7 y C_8

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Factor A	2	1190	595.1	30.99	0.00992
Factor B	2	4360	2180.1	113.53	0.00149
C_5	1	748	748.2	38.96	0.00829
Error(C_6, C_7 y C_8)	3	58	19.2		
Total	8	6356			

El problema, como señalan [61, p. 212], es si la significancia estadística del componente de la interacción AB , del cual hace parte el contraste C_5 puede dar alguna interpretación intuitiva y la respuesta es que no es obvia la interpretación. Si los niveles fueran cuantitativos, esta interpretación sería más fácil porque los contrastes estarían asociados a tendencias lineales o cuadráticas.

Las instrucciones en R para el ejemplo 5.16 son:

```
library(ggplot2);library(unrepx)
remove(list=ls())
A=c(0,0,0,1,1,1,2,2,2)
B=rep(c(0,1,2),3)
rta=c(236,298,302,225,285,250,236,270,260)
```



```

AB=(A+B)%%3 # módulo 3
AB2=(A+2*B)%%3 # módulo 3
A=factor(A);B=factor(B);AB=factor(AB);AB2=factor(AB2)
contraste1= cbind(c(1,-1,0),c(1,1,-2))
contraste2= cbind(c(1,0,-1),c(1,-2,1))
contrasts(A) <- contrasts(B) <- contraste1
contrasts(AB) <- contrasts(AB2) <-contraste2
ajuste1=summary.lm(aov(rta~A+B+AB+AB2))
coefic=ajuste1$coef[,1]
coef=data.frame(coeficientes=subset(coefic,
    names(coefic)!="(Intercept)"),
    Efecto=paste("C",seq(length(coefic)-1),sep=""))
p1<-ggplot(coef, aes(sample = coeficientes)) + stat_qq()
coef1<-ggplot_build(p1)$data[[1]]
coef1$Efecto<-coef$Efecto[order(coef$coeficientes)]
ggplot(coef1,aes(theoretical,sample,label=Efecto))+
    geom_text(hjust=-0.2,vjust=0,size=4)+geom_point()+
    theme_bw()+theme(text=element_text(size=15))
parplot(coef$coeficientes,method="Lenth") # Pareto de
                                         # los coeficientes
ajuste2=summary.lm(aov(rta~A+B+AB))
ajuste2$coef[,3]^2 # F-calculado
X=model.matrix(ajuste2)
ajuste3=aov(rta~A+B+X[,6]); anova(ajuste3)

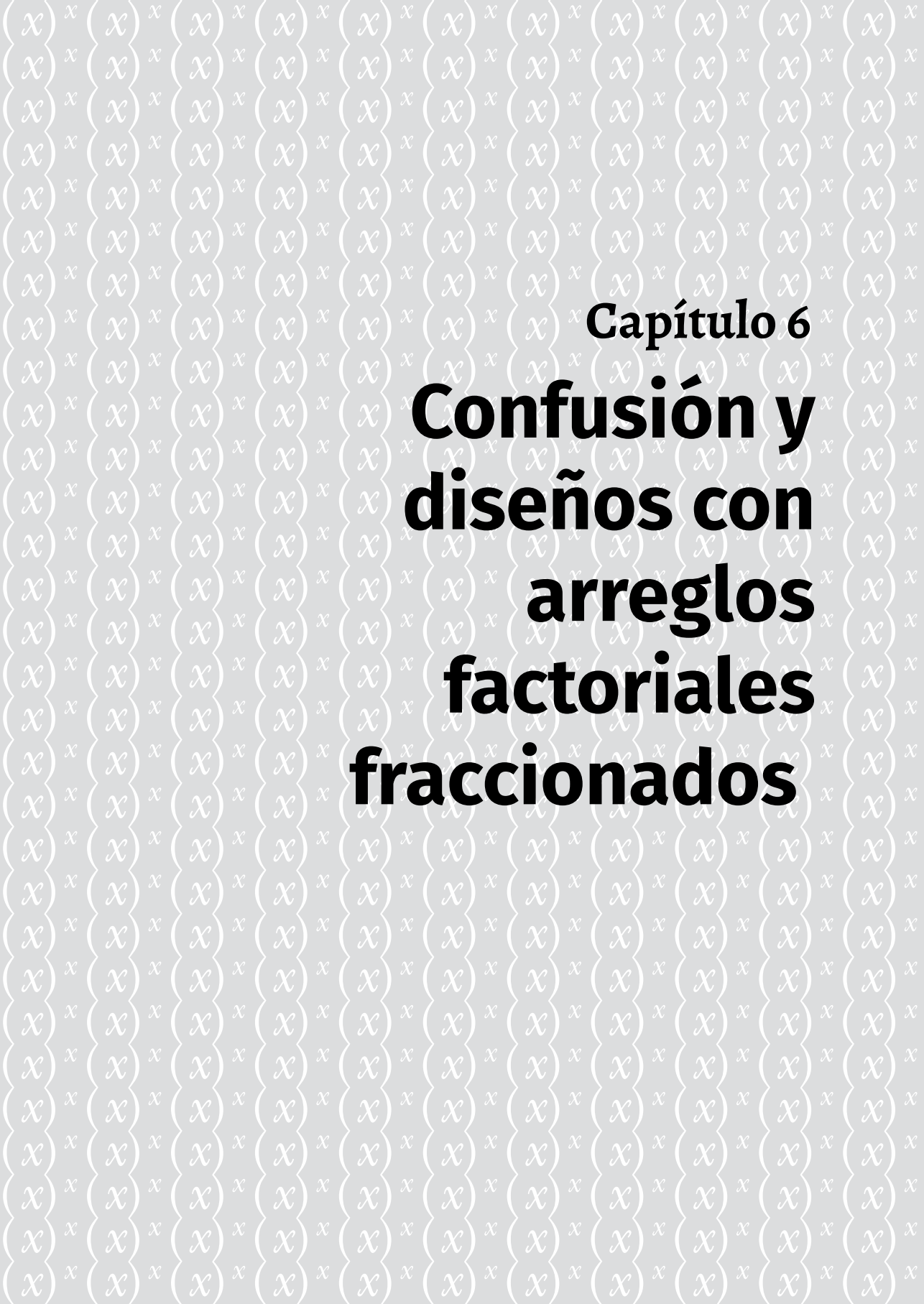
```



7. Ejercicios

- 7.1 Usando los datos del ejemplo 5.3 como prueba piloto, calcular un tamaño de muestra mínimo considerando una potencia de 0.90 y un nivel de significancia de 0.01.
- 7.2 Analizar los datos del ejemplo 5.10 asumiendo que los tratamientos son aleatoriamente asignados a las unidades experimentales dentro de cada réplica.

- 7.3 Realizar las figuras 5.13 y 5.14 para los datos del ejemplo 5.11.
- 7.4 Encontrar d_2 , d_3 y d_4 para el vector de parámetros estimado dado en (5.7).
- 7.5 Mostrar que bajo el modelo (5.6) el vector de parámetros estimado es dado por (5.7).
- 7.6 Completar $\mathbf{X}^\top \mathbf{Y}$ en (5.8).
- 7.7 Completar el vector de parámetros estimados en (5.9).
- 7.8 Para un diseño 3^4 , los 80 grados de libertad asociados a los efectos principales e interacciones se pueden dividir en 40 componentes. Escribir los 40 componentes.



Capítulo 6

Confusión y diseños con arreglos factoriales fraccionados

En este capítulo se presentan alternativas en diseños con arreglos factoriales cuando es imposible efectuar todas las corridas de un experimento bajo condiciones homogéneas, o cuando no se pueden usar todos los tratamientos en el experimento. En las secciones 1 y 2 se presentan sistemas de confusión en diseños con arreglos factoriales 2^k y factoriales fraccionados 2^{k-1} , respectivamente. En las secciones 3 y 4 se presentan sistemas de confusión en diseños con arreglos factoriales 3^k y factoriales fraccionados 3^{k-1} , respectivamente. En la sección 5 se hace una introducción a pseudofactores. Finalmente, como caso especial, en la sección 6 se presenta el diseño central compuesto, que es una variación del diseño con arreglos factoriales 2^k .

1. Confusión en diseños con arreglos factoriales 2^k

Según [41, p. 288], **un sistema de confusión** se usa cuando es imposible efectuar todas las corridas de un experimento 2^k bajo condiciones homogéneas. **Confusión o mezclado** es una técnica de diseño en que un experimento con arreglos factoriales completo se distribuye en **bloques**, donde el tamaño del bloque es menor que el número de tratamientos de una réplica. La técnica hace que la información de ciertos efectos de tratamiento, por ejemplo, las interacciones de orden superior, sean **indistinguibles** de los bloques o estén **confundidas** con los bloques.

Este sacrificio de la interacción de orden superior está basado en el principio de orden jerárquico [61, p. 112].

Se puede construir diseños con arreglos factoriales 2^k en 2^p bloques incompletos en que $p < k$.

1.1. Confusión de un diseño con arreglos factoriales 2^k en dos bloques

Considere inicialmente un diseño con arreglos factoriales 2^2 en 2 bloques. Para construir los bloques se tendrá en cuenta la interacción de orden superior que es AB , ver tabla 6.1.

Tabla 6.1. Asignación de 2^2 tratamientos a dos bloques

Combinación de tratamientos	A	B	AB	Asignación bloque
(1)	-	-	+	$\Rightarrow$ Bloque I
b	-	+	-	$\Rightarrow$ Bloque II
a	+	-	-	$\Rightarrow$ Bloque II
ab	+	+	+	$\Rightarrow$ Bloque I

Por tanto, los bloques quedan conformados como se presenta en la figura 6.1.

Bloque I	Bloque II
A_0B_0	A_0B_1
A_1B_1	A_1B_0
Bloque I	Bloque II
(1)	b
ab	a

Figura 6.1. Esquema de asignación de un diseño con 2^2 tratamientos a dos bloques

Si en el modelo se consideraran como fuente de variación los bloques (asociados a δ) y la interacción AB , se tiene que:

$$\begin{aligned}
 \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \\
 \begin{bmatrix} Y_{00} \\ Y_{01} \\ Y_{10} \\ Y_{11} \end{bmatrix} &= \begin{bmatrix} 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ (\alpha\beta)_{11} \\ \delta_1 \end{bmatrix} + \mathbf{E}. \tag{6.1}
 \end{aligned}$$

Se observa en la expresión (6.1) que las últimas dos columnas son idénticas, y corresponden a las columnas asociadas al efecto de la interacción AB y al efecto de los bloques, es decir, $(\alpha\beta)_{11}$ y δ_1 , respectivamente. Esto implica que NO se puede distinguir el efecto del bloqueo del efecto de la interacción y por esto se dice que estos dos efectos están **confundidos** y se decide retirar una de las columnas repetidas de la matriz diseño, $\mathbf{X}$, ver la expresión (6.2).

$$\begin{bmatrix} Y_{00} \\ Y_{01} \\ Y_{10} \\ Y_{11} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ (\alpha\beta)_{11} + \delta_1 \end{bmatrix} + \mathbf{E}. \quad (6.2)$$

Nota 6.1. Observe que sigue siendo un diseño con arreglos factoriales sin réplicas y que se tendrán 0 grados de libertad para el error.

Ahora, considere un diseño con arreglos factoriales 2^3 en 2 bloques. Para construir los bloques se tendrá en cuenta la interacción de orden superior que es ABC , ver tabla 6.2. Por tanto los bloques quedarán conformados según la figura 6.2.

Tabla 6.2. Asignación de 2^3 tratamientos a dos bloques

Combinación de tratamientos	A	B	C	ABC	Asignación bloque
(I)	-	-	-	-	$\Rightarrow$ Bloque I
b	-	+	-	+	$\Rightarrow$ Bloque II
a	+	-	-	+	$\Rightarrow$ Bloque II
ab	+	+	-	-	$\Rightarrow$ Bloque I
c	-	-	+	+	$\Rightarrow$ Bloque II
bc	-	+	+	-	$\Rightarrow$ Bloque I
ac	+	-	+	-	$\Rightarrow$ Bloque I
abc	+	+	+	+	$\Rightarrow$ Bloque II

Bloque I	Bloque II
$A_0B_0C_0$	$A_0B_1C_0$
$A_1B_1C_0$	$A_1B_0C_0$
$A_0B_1C_1$	$A_0B_0C_1$
$A_1B_0C_1$	$A_1B_1C_1$
Bloque I	Bloque II
(1)	b
ab	a
bc	c
ac	abc

Figura 6.2. Esquema de asignación de un diseño con 2^3 tratamientos a dos bloques

Nota 6.2. Observe que en el primer bloque se asigna la combinación de tratamiento “(1)”. Este bloque se llama el **bloque principal**.

Si en el modelo se consideraran como fuentes de variación los bloques y la interacción ABC , se tiene que:

$$\begin{bmatrix} Y_{000} \\ Y_{001} \\ Y_{010} \\ Y_{011} \\ Y_{100} \\ Y_{101} \\ Y_{110} \\ Y_{111} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \gamma_1 \\ (\alpha\beta)_1 \\ (\alpha\gamma)_{11} \\ (\beta\gamma)_{11} \\ (\alpha\beta\gamma)_{111} \\ \delta_2 \end{bmatrix} + \mathbf{E}. \quad (6.3)$$

Se observa en (6.3) que las últimas dos columnas son idénticas, que corresponden a las columnas asociadas a los efectos de interacción ABC y de bloques, respectivamente. Esto implica que NO se puede distinguir el efecto del bloqueo del efecto de la interacción ABC , es decir, los dos efectos están *confundidos*, ver la expresión (6.4).

$$\begin{bmatrix} Y_{000} \\ Y_{001} \\ Y_{010} \\ Y_{011} \\ Y_{100} \\ Y_{101} \\ Y_{110} \\ Y_{111} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \gamma_1 \\ (\alpha\beta)_{11} \\ (\alpha\gamma)_{11} \\ (\beta\gamma)_{11} \\ (\alpha\beta\gamma)_{111} + \delta_2 \end{bmatrix} + \mathbf{E}. \quad (6.4)$$

De forma igual al caso anterior, los grados de libertad para el error son cero.

Ejemplo 6.1. Basados en el escenario presentado en [21, p. 465] se **simularon** datos para un diseño con arreglos factoriales 2^4 en que se estudia el desempeño de un nuevo rifle bajo cuatro factores de interés: el peso de carga (A), asociado al parámetro α ; el peso del proyectil (B), asociado al parámetro β ; el propulsor (C), asociado al parámetro γ , y el arma (D), asociada al parámetro δ , que consistía en dos rifles. Se corrió la mitad de los tratamientos un día y la otra mitad otro día, es decir, se bloqueó por día, asociada al parámetro θ . Se usó la interacción de mayor orden, es decir, $ABCD$, para el sistema de confusión. Los datos **simulados** se presentan en la tabla 6.3.

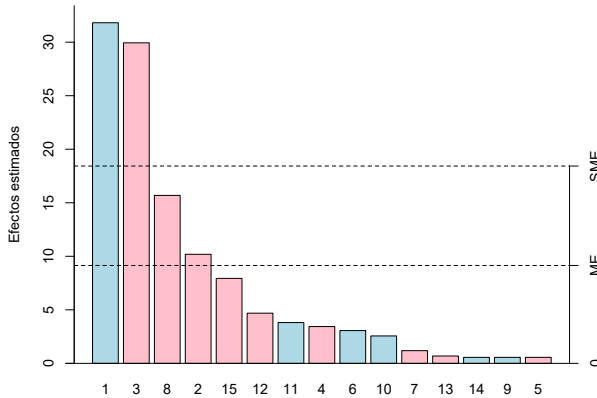
Estime el vector de parámetros, β , identifique si las interacciones triples se pueden considerar significativas o no, y realice el análisis de varianza.

Tabla 6.3. Datos ejemplo 6.1

Día 1		Día 2	
Tratamiento	Efecto	Tratamiento	Efecto
(1)	100	d	62
cd	50	c	41
bd	81	b	72
bc	3	bcd	-9
ad	115	a	148
ac	103	acd	117
ab	156	abd	160
abcd	51	abc	59

Solución. El vector de parámetros estimados se presenta en (6.5) y un gráfico de Pareto de los efectos estimados usando el método de Lenth en la figura 6.3.

$$\hat{\beta} = \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \rightarrow 1 \\ \hat{\beta}_1 \rightarrow 2 \\ \hat{\gamma}_1 \rightarrow 3 \\ \hat{\delta}_1 \rightarrow 4 \\ \hat{\theta}_1 + (\widehat{\alpha\beta\gamma\delta})_{1111} \rightarrow 5 \\ (\widehat{\alpha\beta})_{11} \rightarrow 6 \\ (\widehat{\alpha\gamma})_{11} \rightarrow 7 \\ (\widehat{\beta\gamma})_{11} \rightarrow 8 \\ (\widehat{\alpha\delta})_{11} \rightarrow 9 \\ (\widehat{\beta\delta})_{11} \rightarrow 10 \\ (\widehat{\gamma\delta})_{11} \rightarrow 11 \\ (\widehat{\alpha\beta\gamma})_{111} \rightarrow 12 \\ (\widehat{\alpha\beta\delta})_{111} \rightarrow 13 \\ (\widehat{\alpha\gamma\delta})_{111} \rightarrow 14 \\ (\widehat{\beta\gamma\delta})_{111} \rightarrow 15 \end{pmatrix} = \begin{pmatrix} 81.8125 \\ 31.8125 \\ -10.1875 \\ -29.9375 \\ -3.4375 \\ -0.5625 \\ 3.0625 \\ -1.1875 \\ -15.6875 \\ 0.5625 \\ 2.5625 \\ 3.8125 \\ -4.6875 \\ -0.6875 \\ 0.5625 \\ -7.9375 \end{pmatrix}. \quad (6.5)$$



Valores críticos de Lenth: ME = 9.14, SME = 18.4

Figura 6.3. Gráfico de Pareto de los efectos estimados para los datos del ejemplo 6.1. Margen de error (ME) y margen de error simultáneo (SME) obtenidos por el método de Lenth

En particular, se observa de la figura 6.3 que las interacciones triples son no significativas y se procede a realizar el análisis de varianza sin estas interacciones, ver tabla 6.4.

Tabla 6.4. Análisis de varianza datos ejemplo 6.1 considerando las interacciones dobles

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Carga	1	16 193	16 193	47.200	0.002
Proyectil	1	1661	1661	4.840	0.093
Propulsor	1	14 340	14 340	41.800	0.003
Arma	1	189	189	0.551	0.499
Día+ $(\alpha\beta\gamma\delta)$	1	5	5	0.015	
Carga:proyectil	1	150	150	0.437	0.545
Carga:propulsor	1	23	23	0.066	0.810
Proyectil:prop.	1	3938	3938	11.478	0.028
Carga:arma	1	5	5	0.015	0.909
Proyectil:arma	1	105	105	0.306	0.609
Propulsor:arma	1	233	233	0.678	0.457
Error	4	1372	343		

De la tabla 6.4 se observa que de las interacciones dobles, solo la interacción entre el factor peso del proyectil y propulsor es significativa, por lo cual se corre el análisis de varianza sin las otras interacciones dobles, ver tabla 6.5.

Tabla 6.5. Análisis de varianza datos ejemplo 6.1 considerando solo la interacción proyectil:propulsor

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Carga	1	16 193	16 193	77.207	<0.001
Proyectil	1	1661	1661	7.918	0.0202
Propulsor	1	14 340	14 340	68.374	<0.001
Arma	1	189	189	0.901	0.367
Día+ $(\alpha\beta\gamma\delta)$	1	5	5	0.024	
Proyectil:prop.	1	3938	3938	18.775	0.002
Error	9	1888	210		

Finalmente, también se retira del modelo el factor arma (δ), quedando el modelo:

$$\mathbb{E}(\text{desempeño})_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \theta_l, \quad i, j, k, l = 0, 1,$$

en que el desempeño promedio es $\hat{\mu} = 81.81$ y $\hat{\alpha}_1 = 31.81$, esto es, el efecto medio del peso de carga en su nivel alto sobre el desempeño del rifle es 31.81 con respecto a la media general. Así, el nivel alto del peso de carga aumenta el desempeño del rifle en 31.81. Por otro lado, $\hat{\alpha}_0 = -31.81$, esto es, el efecto medio del peso de carga en su nivel bajo sobre el desempeño del rifle es -31.81 con respecto a la media general. Así, el nivel bajo del peso de carga disminuye el desempeño del rifle en 31.81. En cuanto al efecto de medio del peso del proyectil (β_j) y del propulsor (γ_k), al interactuar se deben interpretar teniendo en cuenta precisamente la interacción,

$$\hat{\beta}_j + \hat{\gamma}_k + (\hat{\beta}\hat{\gamma})_{jk} = \begin{cases} 10.1875 + 29.9375 - 15.6875 & \text{Peso del proyectil y propulsor} \\ & \text{en niveles bajos} \\ 10.1875 - 29.9375 + 15.6875 & \text{Peso del proyectil en nivel bajo} \\ & \text{y propulsor en nivel alto} \\ -10.1875 + 29.9375 + 15.6875 & \text{Peso del proyectil en nivel alto} \\ & \text{y propulsor en nivel bajo} \\ -10.1875 - 29.9375 - 15.6875 & \text{Peso del proyectil y propulsor} \\ & \text{en niveles altos.} \end{cases}$$

$$= \begin{cases} 24.4375 & \text{Peso del proyectil y propulsor en niveles bajos.} \\ -4.0625 & \text{Peso del proyectil en nivel bajo y propulsor en nivel alto.} \\ 35.4375 & \text{Peso del proyectil en nivel alto y propulsor en nivel bajo.} \\ -55.8125 & \text{Peso del proyectil y propulsor en niveles altos.} \end{cases}$$

Observando que, por ejemplo, la combinación de niveles altos de peso del proyectil y propulsor disminuyen el desempeño del rifle en 55.81. En la figura 6.4 se presentan las estimaciones promedio por peso de carga y de acuerdo a la interacción del peso del proyectil con el propulsor.

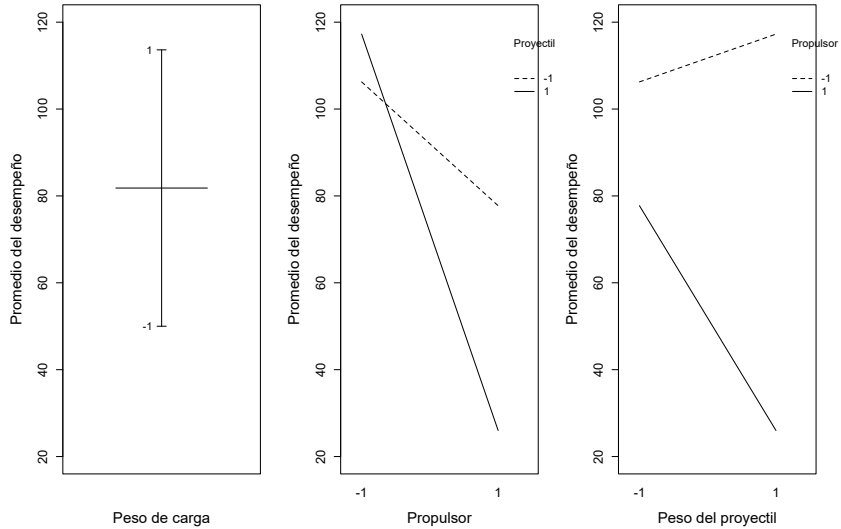


Figura 6.4. Valores medios estimados por peso de carga y para la interacción peso del proyectil con el propulsor

Para la construcción del diseño con arreglos factoriales realizado en el ejemplo 6.1 se puede usar la función *conf.design* de R:

```
library(conf.design)
conf.design(c(A=1,B=1,C=1,D=1),p=2) # factor de confusión ABCD
# p: número de niveles de los factores (p número primo)
```

Las instrucciones en R para el análisis del ejemplo 6.1 son:

```
library(ggplot2);library(unrep);library(cowplot)
remove(list=ls())
a=c(-1,-1,-1,-1, 1, 1, 1, 1,-1,-1,-1,-1, 1, 1, 1, 1) #Carga
b=c(-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1,-1,-1, 1, 1) #Proyectil
c=c(-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1,-1, 1) #Propulsor
d=c(-1, 1, 1,-1, 1,-1,-1, 1, 1,-1,-1, 1,-1, 1, 1,-1) #Arma
dia=c(-1,-1,-1,-1,-1,-1,-1,-1, 1, 1, 1, 1, 1, 1, 1, 1)
rta=c(100,50,81,3,115,103,156,51,62,41,72,-9,148,117,160,59)
ajuste1=summary.lm(aov(rta ~ a + b + a*b + c + a*c + b*c +
  a*b*c + d + a*d + b*d + a*b*d + c*d + a*c*d + b*c*d +dia))
ajuste1 # se obtiene una estimación de beta, entre otros.
```

Confusión y diseños con arreglos factoriales fraccionados

```
coefic=ajuste1$coef[,1] # se extrae beta de ajuste1
coef=data.frame(coeficientes=
  subset(coefic,names(coefic)!="(Intercept)"),
  Efecto=seq(length(coefic)-1))
p1<-ggplot(coef, aes(sample = coeficientes)) + stat_qq()
coef1<-ggplot_build(p1)$data[[1]]
coef1$Efecto<-coef$Efecto[order(coef$coeficientes)]
ggplot(coef1,aes(theoretical,sample,label=Efecto))+
  geom_text(hjust=-0.4,vjust=0,size=4)+geom_point()+
  theme_bw()+theme(text=element_text(size=15))
parplot(coef$coeficientes,method="Lenth") # Gráfico Pareto
                                     # de los coeficientes
ajuste2=aov(rta ~ a + b + a*b + c + a*c + b*c + d + a*d +
  b*d + c*d +dia); summary(ajuste2)
ajuste3=aov(rta ~ a + b + c + b*c + d +dia)
summary(ajuste3)
ajuste4=aov(rta ~ a + b + c + b*c +dia)
summary(ajuste4)
ajustados=ajuste4$fitted
a=as.factor(a)
op=par(mar=c(4,4,1,1))
layout(matrix(c(1,2,3), 1, 3, byrow = TRUE))
plot.design(ajustados~a,ylab="Promedio del desempeño",
  xlab="Peso de Carga",cex.axis=1.2,cex.lab=1.4,
  xaxt = "n",ylim=c(20,120))
interaction.plot(c,b,ajustados,
  ylab="Promedio del desempeño",xlab="Propulsor",
  trace.label="Proyectil",cex.axis=1.2,
  cex.lab=1.4,ylim=c(20,120))
interaction.plot(b,c,ajustados,ylab="Promedio del desempeño",
  xlab="Peso del proyectil",trace.label="Propulsor",
  cex.axis=1.2,cex.lab=1.4,ylim=c(20,120))
par(op); par(mfrow=c(1,1))
```



1.2. Confusión de un diseño con arreglos factoriales 2^k en cuatro bloques

El procedimiento consiste en elegir dos efectos para generar los bloques, confundiendo automáticamente un tercer efecto que es la interacción generalizada de los dos primeros.

Por ejemplo, considere un diseño con arreglos factoriales 2^4 en cuatro bloques. En este caso la interacción de orden superior es $ABCD$, si por ejemplo se elige usar también la interacción ACD para construir los bloques, se tiene que:

$$ABCD \times ACD = A^2B^1C^2D^2. \quad (6.6)$$

Como el diseño es un 2^k , los exponentes se deben expresar en módulo 2, entonces (6.6) quedaría

$$A^2B^1C^2D^2 = A^0B^1C^0D^0 = B,$$

lo que implica que el efecto del factor principal B también quedaría confundido con el efecto de los bloques.

Lo ideal es confundir interacciones de orden superior y no efectos principales. En este caso la interacción $ABCD$ debería ser una de las interacciones que quedan confundidas con los bloques:

$$\text{Confundiendo los efectos } AB \text{ y } CD \implies AB \times CD = ABCD \quad (6.7)$$

$$\text{Confundiendo los efectos } AC \text{ y } BD \implies AC \times BD = ABCD \quad (6.8)$$

$$\text{Confundiendo los efectos } BC \text{ y } AD \implies BC \times AD = ABCD. \quad (6.9)$$

En (6.7) queda confundido el bloqueo con los efectos $ABCD$, AB y CD ; en (6.8) queda confundido el bloqueo con los efectos $ABCD$, AC y BD , y en (6.9) queda confundido el bloqueo con los efectos $ABCD$, BC y AD . En cualquiera de los tres casos, las interacciones dobles podrían ser interesantes.

En la tabla 6.6 está la asignación de tratamientos a los bloques confundiendo AB y CD , y por tanto $ABCD$, ver, también, figura 6.5.

Tabla 6.6. Asignación de los 2^4 tratamientos a cuatro bloques confundiendo AB , CD y $ABCD$

Combinación de tratamientos	A	B	C	D	AB	CD	Asignación bloque
(1)	-	-	-	-	+	+	$\Rightarrow$ Bloque I
b	-	+	-	-	-	+	$\Rightarrow$ Bloque II
a	+	-	-	-	-	+	$\Rightarrow$ Bloque II
ab	+	+	-	-	+	+	$\Rightarrow$ Bloque I
c	-	-	+	-	+	-	$\Rightarrow$ Bloque III
bc	-	+	+	-	-	-	$\Rightarrow$ Bloque IV
ac	+	-	+	-	-	-	$\Rightarrow$ Bloque IV
abc	+	+	+	-	+	-	$\Rightarrow$ Bloque III
d	-	-	-	+	+	-	$\Rightarrow$ Bloque III
bd	-	+	-	+	-	-	$\Rightarrow$ Bloque IV
ad	+	-	-	+	-	-	$\Rightarrow$ Bloque IV
abd	+	+	-	+	+	-	$\Rightarrow$ Bloque III
cd	-	-	+	+	+	+	$\Rightarrow$ Bloque I
bcd	-	+	+	+	-	+	$\Rightarrow$ Bloque II
acd	+	-	+	+	-	+	$\Rightarrow$ Bloque II
abcd	+	+	+	+	+	+	$\Rightarrow$ Bloque I

Bloque I	Bloque II	Bloque III	Bloque IV
(1)	<i>a</i>	<i>c</i>	<i>bc</i>
<i>ab</i>	<i>b</i>	<i>abc</i>	<i>ac</i>
<i>cd</i>	<i>bcd</i>	<i>d</i>	<i>bd</i>
<i>abcd</i>	<i>acd</i>	<i>abd</i>	<i>ad</i>

Figura 6.5. Esquema de asignación de un diseño con 2^4 tratamientos a cuatro bloques

Las instrucciones en R para construir el diseño presentado en la figura 6.5 son:

```
library(conf.design)
G=rbind(c(A=1,B=1,C=0,D=0),c(A=0,B=0,C=1,D=1))
conf.design(G,p=2) # factor de confusión AB y CD
# p: número de niveles de los factores (p número primo)
conf.set(G,p=2) # indica los factores confundidos con los bloques
```


Nota 6.3 (Diseños 2^5 confundidos en 2^3 bloques). Siguiendo un ejemplo presentado en [61, p. 130], para realizar el sistema de confusión en este caso se necesitan tres efectos, por ejemplo, ACE , BCE y $ABCD$, entonces, también quedarán confundidos los efectos:

$$\begin{aligned} ACE \times BCE &= AB \\ ACE \times ABCD &= BDE \\ BCE \times ABCD &= ADE \\ ACE \times BCE \times ABCD &= CD. \end{aligned} \tag{6.10}$$

En total quedan confundidas 7 interacciones (ACE , BCE , $ABCD$, AB , BDE , ADE y CD) con los bloques.

Las instrucciones en R para construir el diseño con el sistema de confusión presentado en (6.10) son:

```
library(conf.design)
G=rbind(c(A=1,B=0,C=1,D=0,E=1),c(A=0,B=1,C=1,D=0,E=1),
        c(A=1,B=1,C=1,D=1,E=0))
conf.design(G,p=2)
conf.set(G,p=2) #indica los factores confundidos con los bloques
```

Otra posibilidad, dada por los mismos autores, es usar para el sistema de confusión las interacciones: AB , AC y DE :

$$\begin{aligned} AB \times AC &= BC \\ AB \times DE &= ABDE \\ AC \times DE &= ACDE \\ AB \times AC \times DE &= BCDE. \end{aligned} \tag{6.11}$$

Luego en total quedan confundidas 7 interacciones (AB , AC , DE , BC , $ABDE$, $ACDE$ y $BCDE$) con los bloques.

Nota 6.4 (Comparación entre dos diseños de confusión). De la nota 6.3 queda la pregunta: ¿cuál sistema de confusión es mejor?

La respuesta es el diseño que tenga menor **aberración**, que quiere decir el diseño que confunda menos efectos de menor orden. Para calcular esto, sea $g_i(b)$ el número de interacciones de i factores confundidos en el diseño b . El caso (6.10) se va a denotar por b_1 y tiene:

$$g_1(b_1) = 0, \quad g_2(b_1) = 2, \quad g_3(b_1) = 4 \quad y \quad g_4(b_1) = 1.$$

El caso (6.11) se va a denotar por b_2 y tiene:

$$g_1(b_2) = 0, \quad g_2(b_2) = 4, \quad g_3(b_2) = 0 \quad y \quad g_4(b_2) = 3.$$

Se trata de que el diseño seleccionado tenga el menor número de efectos de menor orden confundidos, luego se busca el menor i tal que $g_i(b_1) \neq g_i(b_2)$ y se procede a compararlos. En este caso, se inicia con $i = 1$, como $g_1(b_1) = g_1(b_2) = 0$, se pasa a $i = 2$, como $g_2(b_1) \neq g_2(b_2)$ y $g_2(b_1) < g_2(b_2)$, se concluye que el primer diseño propuesto, es decir, b_1 , es el que tiene menor aberración y será el diseño más adecuado entre los dos diseños considerados.

Nota 6.5. También se puede usar la librería FrF2 de R para diseñar sistemas de confusión de diseños 2^k :

```
library(FrF2); remove(list=ls())
FrF2(4,2,blocks="AB") # genera advertencia por alias con
                        # los bloques
FrF2(4,2,blocks="AB",alias.block.2fis=T) ## ver figura 6.1
FrF2(8,3,blocks="ABC",random=FALSE) # sin aleatorizar las
                        # corridas, figura 6.2
FrF2(8,3,blocks="ABC")
FrF2(16,4,blocks="ABCD") # ver tabla 6.3
FrF2(16,4,blocks=c("AB","CD")) # genera error por alias con los
                        # los bloques
FrF2(16,4,blocks=c("AB","CD"),alias.block.2fis=T)
                        # ver figura 6.5
dis=FrF2(16,4,blocks=c("AB","CD"),alias.block.2fis=T)
View(dis)
```

1.3. Réplicas de diseños 2^k confundidos en 2^p bloques

Para tener réplicas de este tipo de diseños, se pueden considerar dos opciones:

- Confundir las mismas interacciones en todas las réplicas. Este diseño sería un **diseño completamente confundido** o mezclado.
- Confundir en cada réplica interacciones diferentes. Este diseño se llama un **diseño parcialmente confundido** o mezclado.

Debe tenerse en cuenta que los bloques están anidados dentro de las réplicas y por tanto deben entrar entre las causas de variación y en la construcción de la matriz diseño.

Ejemplo 6.2. Considere un diseño 2^4 en 2^2 bloques. Presente un diseño completamente confundido y un diseño parcialmente confundido, cada uno con dos réplicas.

Solución

- Confundiendo en las dos réplicas tanto la interacción AB como la interacción CD , por tanto la interacción $ABCD$. El diseño se presenta en la figura 6.6 y el análisis de varianza con causa de variación y grados de libertad, en la tabla 6.7.
- Confundiendo en una réplica las interacciones AB y CD , por tanto $ABCD$, y en la otra réplica las interacciones AC y BD , por tanto $ABCD$. La asignación de los tratamientos a los bloques bajo (6.8) se presenta en la tabla 6.8. El diseño se presenta en la figura 6.7 y el análisis de varianza con causa de variación y grados de libertad, en la tabla 6.9.

Réplica I			
Bloque I	Bloque II	Bloque III	Bloque IV
(1)	<i>a</i>	<i>c</i>	<i>bc</i>
<i>ab</i>	<i>b</i>	<i>abc</i>	<i>ac</i>
<i>cd</i>	<i>bcd</i>	<i>d</i>	<i>bd</i>
<i>abcd</i>	<i>acd</i>	<i>abd</i>	<i>ad</i>
Confundidos <i>AB</i> , <i>CD</i> y <i>ABCD</i> con el bloqueo			
Réplica II			
Bloque I	Bloque II	Bloque III	Bloque IV
(1)	<i>a</i>	<i>c</i>	<i>bc</i>
<i>ab</i>	<i>b</i>	<i>abc</i>	<i>ac</i>
<i>cd</i>	<i>bcd</i>	<i>d</i>	<i>bd</i>
<i>abcd</i>	<i>acd</i>	<i>abd</i>	<i>ad</i>
Confundidos <i>AB</i> , <i>CD</i> y <i>ABCD</i> con el bloqueo			

Figura 6.6. Asignación de 2^4 tratamientos a cuatro bloques considerando un diseño completamente confundido

Réplica I			
Bloque I	Bloque II	Bloque III	Bloque IV
(1)	<i>a</i>	<i>c</i>	<i>bc</i>
<i>ab</i>	<i>b</i>	<i>abc</i>	<i>ac</i>
<i>cd</i>	<i>bcd</i>	<i>d</i>	<i>bd</i>
<i>abcd</i>	<i>acd</i>	<i>abd</i>	<i>ad</i>
Confundidos <i>AB</i> , <i>CD</i> y <i>ABCD</i> con el bloqueo			
Réplica II			
Bloque I	Bloque II	Bloque III	Bloque IV
(1)	<i>b</i>	<i>a</i>	<i>ab</i>
<i>ac</i>	<i>abc</i>	<i>c</i>	<i>bc</i>
<i>bd</i>	<i>d</i>	<i>abd</i>	<i>ad</i>
<i>abcd</i>	<i>acd</i>	<i>bcd</i>	<i>cd</i>
Confundidos <i>AC</i> , <i>BD</i> y <i>ABCD</i> con el bloqueo			

Figura 6.7. Asignación de 2^4 tratamientos a cuatro bloques considerando un diseño parcialmente confundido

Tabla 6.7. Análisis de varianza con causa de variación y grados de libertad de un diseño 2^4 en cuatro bloques considerando un diseño completamente confundido

Causa de variación	Grados de libertad
Réplicas	1
Bloq+ $AB+CD+ABCD$ (réplicas)	6
A	1
B	1
C	1
D	1
AC	1
AD	1
BC	1
BD	1
Error	16
Total	31

Tabla 6.8. Asignación de 2^4 tratamientos a cuatro bloques confundiendo AC , BD y $ABCD$

Combinación de tratam.	A	B	C	D	AC	BD	Asignación bloque
(I)	-	-	-	-	+	+	$\Rightarrow$ Bloque I
b	-	+	-	-	+	-	$\Rightarrow$ Bloque II
a	+	-	-	-	-	+	$\Rightarrow$ Bloque III
ab	+	+	-	-	-	-	$\Rightarrow$ Bloque IV
c	-	-	+	-	-	+	$\Rightarrow$ Bloque III
bc	-	+	+	-	-	-	$\Rightarrow$ Bloque IV
ac	+	-	+	-	+	+	$\Rightarrow$ Bloque I
abc	+	+	+	-	+	-	$\Rightarrow$ Bloque II
d	-	-	-	+	+	-	$\Rightarrow$ Bloque II
bd	-	+	-	+	+	+	$\Rightarrow$ Bloque I
ad	+	-	-	+	-	-	$\Rightarrow$ Bloque IV
abd	+	+	-	+	-	+	$\Rightarrow$ Bloque III
cd	-	-	+	+	-	-	$\Rightarrow$ Bloque IV
bcd	-	+	+	+	-	+	$\Rightarrow$ Bloque III
acd	+	-	+	+	+	-	$\Rightarrow$ Bloque II
abcd	+	+	+	+	+	+	$\Rightarrow$ Bloque I

En el diseño presentado en la figura 6.7 se pueden sacar conclusiones de las interacciones AB y CD de la réplica II , y conclusiones de las inter-

acciones AC y BD de la réplica I . No se puede concluir sobre el efecto de la interacción $ABCD$ porque está confundido en las dos réplicas con el bloqueo. Así, la suma de cuadrados de las interacciones AB y CD se toman solo de la réplica II , y la suma de cuadrados de las interacciones AC y BD se toman solo de la réplica I .

Tabla 6.9. Análisis de varianza con causa de variación y grados de libertad de un diseño 2^4 en cuatro bloques considerando un diseño parcialmente confundido

Causa de variación	Grados de libertad
Réplicas	1
Bloq+ $ABCD$ (réplicas)	6
A	1
B	1
C	1
D	1
AB (De la rép. II)	1
AC (De la rép. I)	1
AD	1
BC	1
BD (De la rép. I)	1
BC (De la rép. II)	1
Error	14
Total	31

2. Diseño con arreglos factoriales fraccionado 2^{k-1}

Según [61, p. 153], dado que realizar un diseño con arreglos factoriales 2^k completo puede no ser viable, en particular si k es grande, por ejemplo $k > 6$, un diseño con arreglos factoriales completo es poco usado en la práctica. Por razones económicas, es común usar el **diseño con arreglos factoriales fraccionado**, que consiste en escoger un subconjunto de tratamientos y aplicarlos, en vez

de aplicar el diseño con arreglos factoriales completo. El subconjunto de tratamientos debe ser *adecuadamente seleccionado* para que sea óptimo según el **criterio de resolución**¹. Entre los inconvenientes de este diseño están los efectos de **alias**, que indican los efectos que tendrían columnas idénticas en la matriz diseño.

Ejemplo 6.3. Diseño 2^{k-1} . Sea $k = 4$. Una interacción que se puede utilizar para seleccionar la fracción del diseño a aplicar es la interacción de orden superior, esto es $ABCD$, ver tabla 6.10, en que se señala si la combinación de factores o tratamiento es seleccionada o no. Vale resaltar que al utilizar la interacción $ABCD$ para seleccionar la fracción, se tiene que $I = ABCD$ es la **relación de definición** del diseño. En la figura 6.8 se presentan los 8 tratamientos seleccionados ($2^{4-1} = 2^3 = 8$).

Tabla 6.10. Asignación de los 2^{4-1} tratamientos usando la interacción $ABCD$

Combinación de tratamientos	A	B	C	D	$ABCD$	Tratamiento seleccionado
(I)	-	-	-	-	+	SI
b	-	+	-	-	-	NO
a	+	-	-	-	-	NO
ab	+	+	-	-	+	SI
c	-	-	+	-	-	NO
bc	-	+	+	-	+	SI
ac	+	-	+	-	+	SI
abc	+	+	+	-	-	NO
d	-	-	-	+	-	NO
bd	-	+	-	+	+	SI
ad	+	-	-	+	+	SI
abd	+	+	-	+	-	NO
cd	-	-	+	+	+	SI
bcd	-	+	+	+	-	NO
acd	+	-	+	+	-	NO
abcd	+	+	+	+	+	SI

¹Ver página 322 sobre resolución del diseño.

Combinación de tratamientos seleccionados			
$abcd$	ab	ac	ad
bc	bd	cd	(1)

Figura 6.8. Tratamientos efectivamente aplicados en el ejemplo 6.3, según tabla 6.10

El modelo matricialmente quedaría:

$$Y = X\beta + E$$

$$\begin{pmatrix} Y_{0000} \\ Y_{1100} \\ Y_{0110} \\ Y_{1010} \\ Y_{0101} \\ Y_{1001} \\ Y_{0011} \\ Y_{1111} \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu + (\alpha\beta\gamma\delta)_{1111} \\ \alpha_1 + (\beta\gamma\delta)_{111} \\ \beta_1 + (\alpha\gamma\delta)_{111} \\ \gamma_1 + (\alpha\beta\delta)_{111} \\ \delta_1 + (\alpha\beta\gamma)_{111} \\ (\alpha\beta)_{11} + (\gamma\delta)_{11} \\ (\alpha\gamma)_{11} + (\beta\delta)_{11} \\ (\alpha\delta)_{11} + (\beta\gamma)_{11} \end{pmatrix} + E.$$

Es importante resaltar que no se consideraron las columnas:

- $(\beta\gamma)_{11}$, porque es igual a la columna asociada con $(\alpha\delta)_{11}$.
- $(\beta\delta)_{11}$, porque es igual a la columna asociada con $(\alpha\gamma)_{11}$.
- $(\gamma\delta)_{11}$, porque es igual a la columna asociada con $(\alpha\beta)_{11}$.
- $(\alpha\beta\gamma)_{111}$, porque es igual a la columna asociada con δ_1 .
- $(\alpha\beta\delta)_{111}$, porque es igual a la columna asociada con γ_1 .
- $(\alpha\gamma\delta)_{111}$, porque es igual a la columna asociada con β_1 .
- $(\beta\gamma\delta)_{111}$, porque es igual a la columna asociada con α_1 .
- $(\alpha\beta\gamma\delta)_{1111}$, porque es igual a la columna asociada con μ .

A los efectos cuyas columnas en la matriz diseño son iguales se les llama **alias**, entonces:

- BC y AD son alias, y el estimador asociado a AD estima en realidad $BC + AD$.
- BD y AC son alias, y el estimador asociado a AC estima en realidad $BD + AC$.
- CD y AB son alias, y el estimador asociado a AB estima en realidad $CD + AB$.
- ABC y D son alias, y el estimador asociado a D estima en realidad $ABC + D$.
- ABD y C son alias, y el estimador asociado a C estima en realidad $ABD + C$.
- ACD y B son alias, y el estimador asociado a B estima en realidad $ACD + B$.
- BCD y A son alias, y el estimador asociado a A estima en realidad $BCD + A$.

Al efecto $ABCD = I$ se le llama la **relación de definición del diseño**.

Para obtener el diseño de tratamientos seleccionados en la figura 6.8 y sus alias, se puede usar en R:

```
library(FrF2)
dis=FrF2(nruns=8,nfactors=4,generator="ABC"); dis
summary(dis); generators(dis)
y=rnorm(8) # una respuesta generada para poder usar la
           # función aliases
aliases(lm(y~(.)^3, data = dis)) # presenta los alias
                                # del diseño
```

Note que en el generador de la salida de R aparece $D = ABC$ ya que:

$$\begin{aligned}
 D &= ABC \\
 D \times D &= ABC \times D \\
 I &= ABCD.
 \end{aligned}$$

Nota 6.6. La **relación de definición** de un diseño con arreglos factoriales fraccionado es el conjunto de todas las columnas que son iguales a la columna asociada a μ .

Nota 6.7. Una forma de obtener los **alias** es multiplicar la relación de definición del diseño por cada efecto. Por ejemplo, si $I = ABCD$, entonces:

$$I \times A = ABCD \times A = A^2BCD = BCD,$$

es decir, A es alias de BCD :

$$I \times BC = ABCD \times BC = AB^2C^2D = AD,$$

es decir, BC es alias de AD , etc.

Nota 6.8. Si se elige la otra fracción, es decir, la fracción en que $ABCD$ tiene signo “-”, se tiene que:

$$I = -ABCD,$$

y, por ejemplo,

$$I \times A = -ABCD \times A = -A^2BCD = -BCD,$$

por tanto A es alias de BCD y el estimador asociado a A estima en realidad $A - BCD$:

$$I \times BC = -ABCD \times BC = -AB^2C^2D = -AD,$$

por tanto BC es alias de $-AD$ y el estimador asociado a BC estima en realidad $BC - AD$, etc.

Ejemplo 6.4. Considere un experimento presentado en el IOCA2017² sobre las condiciones óptimas de operación para corte con plasma en que se midió, entre otros, la exactitud del corte en el largo en la pieza [11]. Los investigadores tuvieron en cuenta cinco factores: presión del gas-plasma (factor A, niveles 39 y 45 psi), distancia entre boquilla y pieza (factor B, niveles 2 y 4 mm), velocidad del corte (factor C, niveles 66 y 133 mm/s), intensidad del arco de corriente (factor D, niveles 40 y 50 amp) y método de cortado (factor E, niveles método 1 y método 2). En la tabla 6.11 se presentan datos **ilustrativos** generados a partir del modelo presentado en el documento de [11, p. 28], teniendo en cuenta que el documento no presenta los datos individuales. Realice el análisis de varianza.

²Avances en investigación de operaciones y ciencias administrativas. Proceedings book.

Tabla 6.11. Datos sobre el estudio de la exactitud del corte con plasma en el largo de la pieza - ejemplo 6.4

Combinación de tratamientos	Exactitud del largo	Combinación de tratamientos	Exactitud del largo
e	3.53	a	4.65
b	2.41	abe	1.51
c	3.47	ace	4.17
bce	3.03	abc	2.29
d	3.39	ade	3.47
bde	3.91	abd	2.26
cde	3.17	acd	4.49
bcd	2.86	abcde	1.48

Fuente: datos ilustrativos generados a partir del modelo presentado en [11, p. 28].

Solución. En (6.12) se presentan los parámetros estimados con:

α asociado con el factor A (presión del gas-plasma)

β asociado con el factor B (distancia entre boquilla y pieza)

γ asociado con el factor C (velocidad del corte medida)

δ asociado con el factor D (intensidad del arco de corriente)

η asociado con el factor E (método de cortado)

$$\begin{aligned}
 \hat{\mu}_1 &= 3.131 \\
 \hat{\alpha}_1 &= -0.091 \rightarrow 1 & \hat{\beta}_1 &= -0.662 \rightarrow 2 \\
 \hat{\gamma}_1 &= -0.011 \rightarrow 3 & \hat{\delta}_1 &= -0.002 \rightarrow 4 \\
 \hat{\eta}_1 &= -0.097 \rightarrow 5 & \widehat{(\alpha\beta)}_{11} &= -0.493 \rightarrow 6 \\
 \widehat{(\alpha\gamma)}_{11} &= 0.078 \rightarrow 7 & \widehat{(\alpha\delta)}_{11} &= -0.113 \rightarrow 8 \\
 \widehat{(\alpha\eta)}_{11} &= -0.286 \rightarrow 9 & \widehat{(\beta\gamma)}_{11} &= -0.043 \rightarrow 10 \\
 \widehat{(\beta\delta)}_{11} &= 0.161 \rightarrow 11 & \widehat{(\beta\eta)}_{11} &= 0.111 \rightarrow 12 \\
 \widehat{(\gamma\delta)}_{11} &= -0.118 \rightarrow 13 & \widehat{(\gamma\eta)}_{11} &= -0.061 \rightarrow 14 \\
 \widehat{(\delta\eta)}_{11} &= -0.024 \rightarrow 15.
 \end{aligned} \tag{6.12}$$

En la figura 6.9 se presentan gráficos half-normal y Pareto de los parámetros estimados. De la misma figura se puede detectar que las interacciones entre los factores presión, *A*, y distancia, *B*, asociada al número 6, y presión, *A*, y método, *E*, asociada al número 9, podrían ser significativas. El resto no son significativas. Escogiendo cualquiera de las otras interacciones para mandar al error, se mandó la interacción entre los factores velocidad, *C*, y método, *E*, asociada al número 14, ver tabla 6.12.

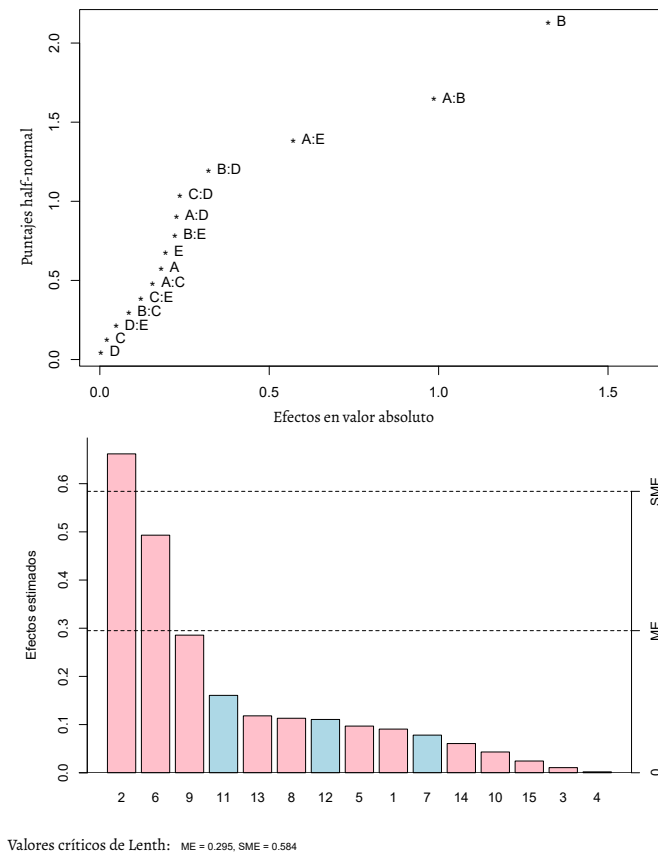


Figura 6.9. Gráfica half-normal (arriba) y Pareto (abajo) de los parámetros estimados - ejemplo 6.4

Tabla 6.12. Análisis de varianza sin considerar la interacción
velocidad* método en el ejemplo 6.4

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calc.	P-valor
Presión	1	0.131	0.131	2.235	0.3753
Distancia	1	7.009	7.009	119.192	0.0581
Velocidad	1	0.002	0.002	0.031	0.8895
Intensidad	1	0.0001	0.0001	0.001	0.9803
Método	1	0.150	0.150	2.553	0.3560
Presión:Dist.	1	3.891	3.891	66.162	0.0779
Presión:Veloc.	1	0.098	0.098	1.661	0.4201
Presión:Intens.	1	0.205	0.205	3.482	0.3132
Presión:Método	1	1.305	1.305	22.197	0.1331
Distancia:Veloc.	1	0.030	0.030	0.506	0.6064
Dist.:Intens.	1	0.413	0.413	7.020	0.2298
Dist.:Método	1	0.196	0.196	3.330	0.3192
Veloc.:Intens.	1	0.223	0.223	3.796	0.3019
Intens.:Método	1	0.010	0.010	0.162	0.7566
Error	1	0.059	0.059		
Total	15	13.7210			

De la tabla 6.12 se observa que hay interacciones dobles con P-valor mayor a 0.20, por lo cual se retiran del modelo y se vuelve a ajustar, ver tabla 6.13.

De la tabla 6.13 se observa que las interacciones presión (*A*) y distancia (*B*); y presión (*A*) y método (*E*) son significativas. Como los efectos dobles que quedaron en el modelo no incluyen ni velocidad (*C*) ni intensidad (*D*), se proceden a quitar del modelo, teniendo en cuenta que no son significativos, ver tabla 6.14.

Tabla 6.13. Análisis de varianza sin considerar algunas interacciones dobles para el ejemplo 6.4

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calc.	P-valor
Presión	1	0.131	0.131	0.853	0.3827
Distancia	1	7.009	7.009	45.502	0.0001
Velocidad	1	0.002	0.002	0.012	0.9164
Intensidad	1	0.0001	0.0001	0.000	0.9852
Método	1	0.150	0.150	0.975	0.3524
Presión:Dist.	1	3.891	3.891	25.257	0.0010
Presión:Método	1	1.305	1.305	8.474	0.0196
Error	7	1.232	0.154		
Total	15	13.7210			

Tabla 6.14. Análisis de varianza con las interacciones dobles presión y distancia, y presión y método del ejemplo 6.4

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calc.	P-valor
Presión	1	0.131	0.131	1.065	0.3265
Distancia	1	7.009	7.009	56.791	<0.0001
Método	1	0.150	0.150	1.217	0.2959
Presión:Dist.	1	3.891	3.891	31.524	0.0002
Presión:Método	1	1.305	1.305	10.576	0.0087
Error	10	1.234	0.123		
Total	15	13.7210			

Un gráfico de los valores ajustados de acuerdo a las interacciones significativas se presenta en la figura 6.10.

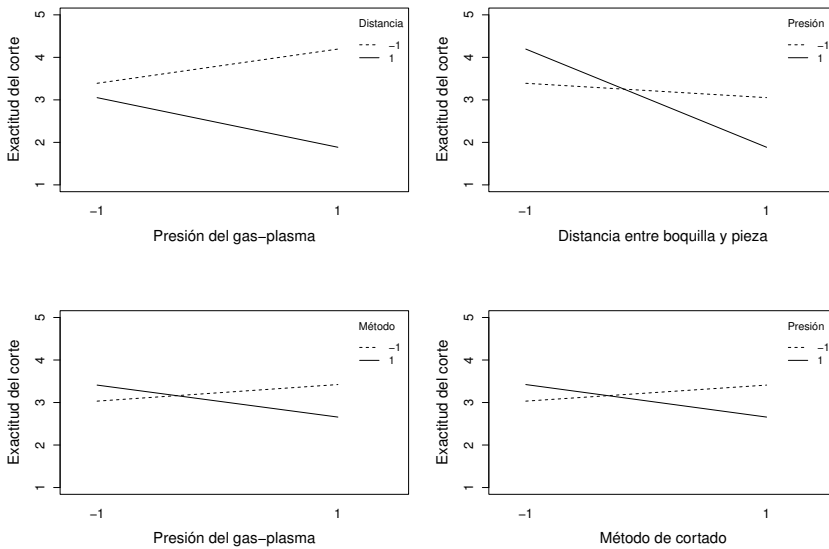


Figura 6.10. Valores promedio estimados de acuerdo a la presión del gas-plasma, la distancia entre boquilla y pieza, y el método de cortado - ejemplo 6.4

Las instrucciones en R usadas en el ajuste del modelo para los datos del ejemplo 6.4 son:

```
library(FrF2); library(ggplot2); library(unrepX)
remove(list=ls())
dis=FrF2(nruns=16,nfactors=5,generator="ABCD",random=FALSE); dis
summary(dis)
Elargo=c(3.53,4.65,2.41,1.51,3.47,4.17,3.03,2.29,3.39,3.47,3.91,
         2.26,3.17,4.49,2.86,1.48) # Exactitud del largo
alias(es=lm(Elargo~(.)^3, data = dis))
datos=add.response(dis,response=Elargo)
DanielPlot(datos, half=TRUE,alpha=1)
# A:"Presión", B:"Distancia boquilla", C:"Velocidad",
# D: "Intensidad del arco" y E:"Método de cortado"
ajuste1=aov(Elargo~(A+B+C+D+E)^2,data=datos)
```

Confusión y diseños con arreglos factoriales fraccionados

```
coef=data.frame(coeficientes=subset(ajuste1$coef,
                                   names(ajuste1$coef)!="(Intercept)"))
parplot(coef$coeficientes,method="Lenth") # Pareto-efectos estim.
ajuste2=aov(Elargo~A+B+C+D+E+A*B+A*C+A*D+A*E+B*C+B*D+B*E+C*D+D*E,
           data=datos);summary(ajuste2)
ajuste3=aov(Elargo~A+B+C+D+E+A*B+A*E,data=datos);summary(ajuste3)
ajuste4=aov(Elargo~A+B+E+A*B+A*E,data=datos);summary(ajuste4)
hnp:hnp(ajuste4,halfnormal=FALSE)
ajustados=ajuste4$fitted
t="Exactitud del corte"; a="Presión del gas-plasma"
b="Distancia entre boquilla y pieza"; e="Método de cortado"
op <- par(no.readonly = TRUE)
par(mar=c(0.5,0.5,0.5,0.5))
layout(matrix(c(1,2,3,4), 2, 2, byrow = TRUE))
interaction.plot(datos$A,datos$B,ajustados,ylab=t,xlab=a,
               trace.label="Distancia",cex.axis=1.2,cex.lab=1.4,ylim=c(1,5))
interaction.plot(datos$B,datos$A,ajustados,ylab=t,xlab=b,
               trace.label="Presión",cex.axis=1.2,cex.lab=1.4,ylim=c(1,5))
interaction.plot(datos$A,datos$E,ajustados,ylab=t,xlab=a,
               trace.label="Método",cex.axis=1.2,cex.lab=1.4,ylim=c(1,5))
interaction.plot(datos$E,datos$A,ajustados,ylab=t,xlab=e,
               trace.label="Presión",cex.axis=1.2,cex.lab=1.4,ylim=c(1,5))
par(op); par(mfrow=c(1,1))
```



Para identificar la **resolución del diseño** se pueden observar las siguientes características:

1. Diseños de resolución III.

- Ninguno de los efectos principales es alias de otro efecto principal.
- Efectos principales son alias de las interacciones de dos factores.
- Interacciones de dos factores *pueden* ser alias entre sí.

2. Diseños de resolución IV.

- Ninguno de los efectos principales es alias de otro efecto principal.

- Efectos principales NO son alias de interacciones de dos factores.
- Interacciones de dos factores son alias entre sí.

3. Diseños de resolución V.

- Ninguno de los efectos principales es alias de otro efecto principal.
- Efectos principales NO son alias de interacciones de dos factores.
- Interacciones de dos factores NO son alias de otras interacciones de dos factores.
- Interacciones de dos factores son alias de interacciones de tres factores.

Vale resaltar que entre más alta sea la resolución, mejor es el diseño para identificar efectos significativos.

Ejemplo 6.5. En el ejemplo 6.4 sobre condiciones óptimas para corte con plasma, ¿cuál es la resolución del diseño?

Solución. El diseño aplicado es 2^{5-1}_V , es decir, es un diseño con arreglos factoriales fraccionado de resolución V, ya que el conjunto de alias es:

$$\begin{array}{ll}
 AB \times ABCDE = CDE & AC \times ABCDE = BDE \\
 AD \times ABCDE = BCE & AE \times ABCDE = BCD \\
 BC \times ABCDE = ADE & BD \times ABCDE = ACE \\
 BE \times ABCDE = ACD & CD \times ABCDE = ABE \\
 CE \times ABCDE = ABD & DE \times ABCDE = ABC,
 \end{array}$$

observando que las interacciones de dos factores son alias de interacciones de tres factores. ✓

3. Confusión en diseños con arreglos factoriales 3^k

Se pueden construir diseños con arreglos factoriales 3^k en 3^p bloques incompletos en que $p < k$. Para hacer la distribución de los 3^k tratamientos en 3

bloques, por ejemplo, es necesario usar 2 grados de libertad (ya sean de interacciones o de efectos principales) para los bloques. Considere un diseño con arreglos factoriales 3^2 confundido en 3 bloques, se tienen diferentes alternativas para construir los bloques, unas más adecuadas que las otras.

- Usar un factor principal, por ejemplo el factor A . Los dos grados de libertad asociados al factor A se van a escribir como $A(x_1)$ y $A^2(2x_1)$, en que x_1 denota el nivel del factor A , sin que implique alguna ordenación. La notación 0, 1 o 2 para los niveles sirven para configurar la asignación de los tratamientos a los bloques.

En la tabla 6.15 se presentan dos esquemas de asignación de los tratamientos a los bloques, usando en cada esquema uno de los grados de libertad. Se observa que los dos esquemas, 1 y 2, generan la misma asignación de los tratamientos a los bloques, luego se usará el esquema de confusión que tenga exponente 1, esto es A , y no A^2 .

Usar grados de libertad de alguno de los factores principales **no es un buen sistema de confusión** porque los niveles del factor principal, por ejemplo el factor A , quedan confundidos con los bloques, luego no se podrá concluir si el factor A es significativo o no.

Tabla 6.15. Asignación de 3^2 tratamientos a tres bloques usando un grado de libertad del factor A

Trat.	x_1 Fact. A	x_2 Fact. B	A x_1	Mód.	Asignac. bloque	A^2 $2x_1$	Mód.	Asignac. bloque
a_0b_0	0	0	0	0	→ Bloq. I	0	0	→ Bloque I
a_0b_1	0	1	0	0	→ Bloq. I	0	0	→ Bloque I
a_0b_2	0	2	0	0	→ Bloq. I	0	0	→ Bloque I
a_1b_0	1	0	1	1	→ Bloq. II	2	2	→ Bloque III
a_1b_1	1	1	1	1	→ Bloq. II	2	2	→ Bloque III
a_1b_2	1	2	1	1	→ Bloq. II	2	2	→ Bloque III
a_2b_0	2	0	2	2	→ Bloq. III	4	1	→ Bloque II
a_2b_1	2	1	2	2	→ Bloq. III	4	1	→ Bloque II
a_2b_2	2	2	2	2	→ Bloq. III	4	1	→ Bloque II

esquema 1
esquema 2

- Usar alguno de los 4 grados de libertad de la interacción $A \times B$, que se van a denotar por $AB (x_1 + x_2)$, $AB^2 (x_1 + 2x_2)$, $A^2B (2x_1 + x_2)$ y $A^2B^2 (2x_1 + 2x_2)$, en que x_1 denota el nivel del factor A y x_2 denota el nivel del factor B . Ver tabla 6.16.

Tabla 6.16. Asignación de 3^2 tratamientos a tres bloques usando un grado de libertad de la interacción $A \times B$

Trat.	x_1 A	x_2 B	AB $x_1 + x_2$	Mód.	Asignac. bloque	AB^2 $x_1 + 2x_2$	Mód.	Asignac. bloque
a_0b_0	0	0	0	0	→ Bloq. I	0	0	→ Bloque I
a_0b_1	0	1	1	1	→ Bloq. II	2	2	→ Bloque III
a_0b_2	0	2	2	2	→ Bloq. III	4	1	→ Bloque II
a_1b_0	1	0	1	1	→ Bloq. II	1	1	→ Bloque II
a_1b_1	1	1	2	2	→ Bloq. III	3	0	→ Bloque I
a_1b_2	1	2	3	0	→ Bloq. I	5	2	→ Bloque III
a_2b_0	2	0	2	2	→ Bloq. III	2	2	→ Bloque III
a_2b_1	2	1	3	0	→ Bloq. I	4	1	→ Bloque II
a_2b_2	2	2	4	1	→ Bloq. II	6	0	→ Bloque I

esquema 3

esquema 4

Trat.	x_1 A	x_2 B	A^2B $2x_1 + x_2$	Mód.	Asignac. bloque	A^2B^2 $2x_1 + 2x_2$	Mód.	Asignac. bloque
a_0b_0	0	0	0	0	→ Bloq. I	0	0	→ Bloque I
a_0b_1	0	1	1	1	→ Bloq. II	2	2	→ Bloque III
a_0b_2	0	2	2	2	→ Bloq. III	4	1	→ Bloque II
a_1b_0	1	0	2	2	→ Bloq. III	2	2	→ Bloque III
a_1b_1	1	1	3	0	→ Bloq. I	4	1	→ Bloque II
a_1b_2	1	2	4	1	→ Bloq. II	6	0	→ Bloque I
a_2b_0	2	0	4	1	→ Bloq. II	4	1	→ Bloque II
a_2b_1	2	1	5	2	→ Bloq. III	6	0	→ Bloque I
a_2b_2	2	2	6	0	→ Bloq. I	8	2	→ Bloque III

esquema 5

esquema 6

En la tabla 6.16 se presentan los cuatro esquemas de asignación de los tratamientos a los bloques usando en cada esquema uno de los 4 grados de libertad. Se observa que los esquemas 3 y 6 generan la misma asignación de los tratamientos a los bloques, al igual que los esquemas 4 y 5,

luego se usará uno de los esquemas de confusión que tengan la primera letra con exponente 1, esto es AB o AB^2 .

Aunque no quedan comprometidos los efectos principales con los bloques, sí queda comprometida la conclusión sobre el efecto de la interacción entre los dos factores. Una alternativa para poder estudiar la interacción es realizar dos réplicas, en que se confunda AB en la primera réplica y AB^2 en la segunda réplica [32, p. 376].

En la figura 6.11 se presentan los tratamientos asignados a los bloques según el esquema 4, el cual se puede obtener en R usando:

```
library(conf.design)
G=c(A=1,B=2)
conf.design(G,p=3) # factor de confusión  $AB^2$ 
# p: número de niveles de los factores (p número primo)
```

o usando:

```
library(DoE.base)
fac.design(nlevels=c(3,3), blocks=3,block.gen=c(1,2))
# factor de confusión  $AB^2$ , orden aleatorio.
```

Bloque I	Bloque II	Bloque III
a_0b_0	a_0b_2	a_0b_1
a_1b_1	a_1b_0	a_1b_2
a_2b_2	a_2b_1	a_2b_0

Figura 6.11. Esquema de asignación de un diseño con 3^2 tratamientos a tres bloques usando AB^2

En (6.13) se presenta la matriz diseño asociada a la figura 6.11, en que δ_1 y δ_2 corresponden a los efectos de los bloques. Se observa que los contrastes c_7 y c_8 quedan confundidos con los dos bloques, δ_1 y δ_2 , esto es, se usan de los cuatro grados de libertad que tiene la interacción entre los factores A y B , dos grados de libertad para el bloqueo que corresponden a los dos grados de libertad de AB^2 y A^2B , ya que generan el mismo esquema de confusión.

$$\mathbf{X} = \begin{pmatrix} \mu & c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & \delta_1 & \delta_2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & -1 & 1 & 0 & -2 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -2 & -1 & 1 & 0 & -2 & 0 & 1 \\ 1 & -1 & 1 & 1 & 1 & 0 & -2 & 0 & -2 & 0 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & -2 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 0 & -2 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 0 & -2 & -1 & 1 & 1 & 1 & 0 & -2 & 0 & 1 \\ 1 & 0 & -2 & 0 & -2 & 0 & -2 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{matrix} a_0b_0 \\ a_0b_1 \\ a_0b_2 \\ a_1b_0 \\ a_1b_1 \\ a_1b_2 \\ a_2b_0 \\ a_2b_1 \\ a_2b_2 \end{matrix} \quad (6.13)$$

Ejemplo 6.6. Considere un experimento realizado en la Universidad Nacional de Colombia sobre consumo de combustible en vehículos tipo camión bajo diferentes condiciones de peso (11.72 toneladas, 17.18 toneladas y 27.99 toneladas, notación: P), pendiente (pendiente longitudinal del tramo, 2 %, 5.2 % y 7 %, notación: L) y velocidad (25 km/h, 35 km/h y 45 km/h, notación: V). En el estudio midieron, entre otros, el consumo de combustible en ml/km de diferentes tipos de camiones [48, p. 48], entre los cuales estaba la configuración de camión C3, que es el que se va a tomar para este ejemplo. Los investigadores realizaron tres réplicas por tratamiento (submuestreo), sin embargo para efectos **ilustrativos** se va a tomar solo un dato de los obtenidos por tratamiento y se va a asignar a bloques de tamaño 9 usando como componente de bloqueo PL^2V . En la tabla 6.17 se muestra el esquema de bloqueo para este diseño y los datos. Realizar el análisis de varianza.

Solución. En la tabla 6.18 se presenta el análisis de varianza. Cabe notar que al usar PL^2V para el bloqueo, la componente PL^2V queda confundida con los bloques y por tanto el resto de componentes de la interacción triple, es decir, PLV , PLV^2 y PL^2V^2 serán asignadas al término del error con un total 6 grados de libertad. Se observa que las interacciones dobles de los factores peso con velocidad y pendiente longitudinal del tramo con velocidad presentan un P-valor mayor 0.20, por lo que se decide retirarlas del modelo.

Tabla 6.17. Esquema de asignación del diseño 3^3 y datos del ejemplo 6.6 en tres bloques

Bloque I $x_1 + 2x_2 + x_3 = 0$	Bloque II $x_1 + 2x_2 + x_3 = 1$	Bloque III $x_1 + 2x_2 + x_3 = 2$
$a_0b_0c_0 = 286.82$	$a_1b_0c_0 = 355.11$	$a_2b_0c_0 = 592.62$
$a_1b_1c_0 = 917.19$	$a_2b_1c_0 = 1363.59$	$a_0b_1c_0 = 603.36$
$a_2b_2c_0 = 1368.70$	$a_0b_2c_0 = 661.58$	$a_1b_2c_0 = 950.84$
$a_2b_0c_1 = 558.64$	$a_0b_0c_1 = 267.13$	$a_1b_0c_1 = 334.06$
$a_0b_1c_1 = 550.04$	$a_1b_1c_1 = 869.14$	$a_2b_1c_1 = 1299.09$
$a_1b_2c_1 = 871.99$	$a_2b_2c_1 = 1316.64$	$a_0b_2c_1 = 561.80$
$a_1b_0c_2 = 347.26$	$a_2b_0c_2 = 524.43$	$a_0b_0c_2 = 216.32$
$a_2b_1c_2 = 1252.65$	$a_0b_1c_2 = 619.66$	$a_1b_1c_2 = 829.10$
$a_0b_2c_2 = 597.91$	$a_1b_2c_2 = 820.8$	$a_2b_2c_2 = 1273.05$

La asignación a los bloques es ilustrativa.

Fuente: datos tomados de [48, p. 66].

Tabla 6.18. Análisis de varianza para los datos del ejemplo 6.6

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calc.	P-valor
Peso	2	1525 863	762 932	795.149	<0.001
Pendiente	2	1765 672	882 836	920.117	<0.001
Velocidad	2	23 204	11 602	12.092	<0.001
Bloque	2	1092	546	0.569	0.5940
Peso:Pendiente	4	169 688	42 422	44.213	0.0001
Peso:Velocidad	4	3397	849	0.885	0.5257
Pendiente:Veloc.	4	2772	693	0.722	0.607
Error	6	5757	959		
Total	26	3 497 446			

Adicionalmente, teniendo en cuenta que los factores tienen niveles cuantitativos, se construyeron polinomios ortogonales para ajustar el modelo final, observando que la interacción peso-cuadrático por pendiente-cuadrático no es significativa, ver tabla 6.19. La figura 6.12 presenta los valores ajustados del modelo seleccionado.

Tabla 6.19. Análisis de varianza para los datos del ejemplo 6.6 considerando polinomios ortogonales

Causa de variación	Grados de lib.	Suma de cuadrados	F-calc.	P-valor
Velocidad.L	1	21 261.2821	24.958	0.0002
Velocidad.Q	1	1943.1601	2.281	0.1532
PesoP.L	1	1 523 256.8658	1788.118	<0.0001
PesoP.Q	1	2606.1589	3.059	0.10215
PendienteL.L	1	1 566 934.1684	1839.389	<0.0001
PendienteL.Q	1	198 738.2594	233.294	<0.0001
PesoP.L:Pend.L.L	1	145 235.6437	170.489	<0.0001
PesoP.Q:Pend.L.L	1	2661.9227	3.125	0.09889
PesoP.L:Pend.L.Q	1	21 113.2971	24.784	0.00020
PesoP.Q:Pend.L.Q	1	677.3554	0.795	0.38762
Bloques	2	1092	0.641	
Error	14	11 926.28		

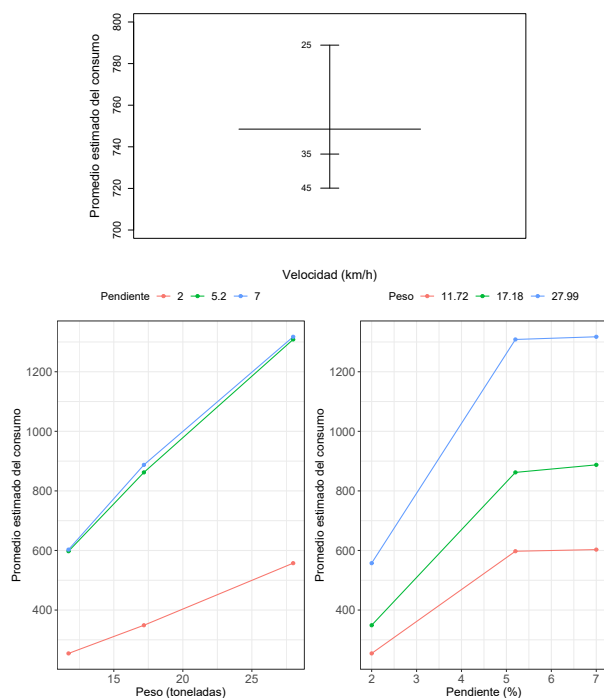


Figura 6.12. Valores ajustados promedio - ejemplo 6.6

Confusión y diseños con arreglos factoriales fraccionados

Las instrucciones en R usadas en el ajuste del modelo son:

```
library(DoE.base);library(varhandle);library(MASS)
library(ggplot2);library(dplyr);library(cowplot)
remove(list=ls())
dis=fac.design(factor.names=list(Peso=c(11.72,17.18,27.99),
  Pendiente=c(2,5.2,7),Velocidad=c(25,35,45)),nlevels=c(3,3,3),
  blocks=3,block.gen=c(1,2,1),random=F)#factor de confusión PL^2V
Consumo=c(286.82,917.19,1368.70,558.64,550.04,871.99,347.26,
  1252.65,597.91,355.11,1363.59,661.58,267.13,869.14,
  1316.64,524.43,619.66,820.8,592.62,603.36,950.84,
  334.06,1299.09,561.80,216.32, 829.10,1273.05)
datos=add.response(dis,Consumo)
summary(aov(Consumo~(Peso+Pendiente+Velocidad)^2+Blocks,
  data=datos))
## polinomios ortogonales - método Robson 1959 ##
nivelesP=unique(unfactor(datos$Peso))
nivelesL=unique(unfactor(datos$Pendiente))
nP=length(nivelesP);nL=length(nivelesL)
nivP2=nivelesP^2; nivL2=nivelesL^2
P.L=nivelesP-mean(nivelesP)
Pend.L=nivelesL-mean(nivelesL)
P.Q=nivP2-(1/nP)*sum(nivP2)-P.L*(sum(nivP2*P.L)/(sum(P.L^2)))
Pend.Q=nivL2-(1/nL)*sum(nivL2)-
  Pend.L*(sum(nivL2*Pend.L)/(sum(Pend.L^2)))
P.L=P.L/sqrt(sum(P.L^2)); P.Q=P.Q/sqrt(sum(P.Q^2))
L.L=Pend.L/sqrt(sum(Pend.L^2)); L.Q=Pend.Q/sqrt(sum(Pend.Q^2))
matrizP=cbind(P.L,P.Q); matrizP
matrizL=cbind(L.L,L.Q); matrizL
contrasts(datos$Peso)=matrizP
contrasts(datos$Pendiente)=matrizL
contrasts(datos$Velocidad)=contr.poly(levels(datos$Velocidad))
ajuste=lm(Consumo~Velocidad+Peso+Pendiente+
  Peso*Pendiente+Blocks,datos)
summary(ajuste)
X=model.matrix(ajuste)
ajuste1=lm(Consumo~X[-13]-1,datos); summary(ajuste1)
datos=data.frame(datos,ajustados=ajuste1$fitted)
```



```
ty="Promedio estimado del Consumo"
plot.design(datos$ajustados~datos$Velocidad,ylab=ty,
  xlab="Velocidad (Km/h)",cex.axis=1.2,cex.lab=1.4,xaxt="n",
  ylim=c(700,800))
datos.agr <- datos %>% group_by(Peso,Pendiente) %>%
  summarise(Promedio = mean(ajustados, na.rm = TRUE))
tema <- theme_bw()+ theme(axis.text = element_text(size=12),
  axis.title = element_text(size=12),
  legend.text=element_text(size=11),
  legend.title=element_text(size=11),legend.position="top")
p1 <- ggplot(datos.agr,aes(unfactor(Peso), Promedio)) +
  geom_line(aes(group = Pendiente, color = Pendiente))+
  geom_point(size = 1.2, aes(color = Pendiente))+
  ylab(ty)+xlab("Peso (toneladas)")+tema
p2 <- ggplot(datos.agr,aes(unfactor(Pendiente), Promedio)) +
  geom_line(aes(group = Peso, color = Peso))+
  geom_point(size = 1.2, aes(color = Peso))+
  ylab(ty)+xlab("Pendiente (%)")+tema
plot_grid(p1,p2,nrow = 1)
```



4. Diseños con arreglos factoriales fraccionados 3^{k-1}

[61, p. 214] sugieren que aún siendo poco costoso realizar todos los tratamientos, y, conociendo que una gran parte de los grados de libertad en un diseño 3^k se van para las interacciones de orden superior y que algunas de las interacciones de orden superior no son significativas, en la práctica es más eficiente realizar un tercio del experimento o un noveno del mismo. Al igual que en diseños con arreglos factoriales fraccionados 2^{k-1} , el subconjunto de tratamientos debe ser adecuadamente seleccionado para que sea óptimo según el criterio de resolución.

Ejemplo 6.7. Diseño 3^{3-1} . Para la selección de la fracción del diseño se va a usar la componente ABC^2 de la interacción $A \times B \times C$, ver tabla 6.20. Los tratamientos seleccionados ($3^{3-1} = 9$) que efectivamente se aplicarán están en la tabla 6.21.

Tabla 6.20. Asignación de los 3^{3-1} tratamientos usando la componente ABC^2

A	B	C	$ABC^2 =$ $x_1 + x_2 +$ $2x_3$	Tratam. selecc.	A	B	C	$ABC^2 =$ $x_1 + x_2 +$ $2x_3$	Tratam. selecc.
0	0	0	0	SI	1	1	2	0	SI
0	0	1	2	NO	1	2	0	0	SI
0	0	2	1	NO	1	2	1	2	NO
0	1	0	1	NO	1	2	2	1	NO
0	1	1	0	SI	2	0	0	2	NO
0	1	2	2	NO	2	0	1	1	NO
0	2	0	2	NO	2	0	2	0	SI
0	2	1	1	NO	2	1	0	0	SI
0	2	2	0	SI	2	1	1	2	NO
1	0	0	1	NO	2	1	2	1	NO
1	0	1	0	SI	2	2	0	1	NO
1	0	2	2	NO	2	2	1	0	SI
1	1	0	2	NO	2	2	2	2	NO
1	1	1	1	NO					

Tabla 6.21. Tratamientos seleccionados

A	B	C	A	B	C	A	B	C
0	0	0	0	1	1	0	2	2
1	0	1	1	1	2	1	2	0
2	0	2	2	1	0	2	2	1

Como se utilizó el componente ABC^2 para seleccionar la fracción, entonces la relación de definición del diseño es $I = ABC^2$. En la tabla 6.22 se puede observar que las entradas de las dos columnas de los contrastes asociados a ABC^2 son todas iguales a 0 y -2, respectivamente.

Tabla 6.22. Tabla de contrastes ortogonales para el ejemplo 6.7

$A B C$	A_1	A_2	B_1	B_2	C_1	C_2	AB_1	AB_2	AB_1^2	AB_2^2
0 0 0	1	1	1	1	1	1	1	1	1	1
0 1 1	1	1	-1	1	-1	1	-1	1	0	-2
0 2 2	1	1	0	-2	0	-2	0	-2	-1	1
1 0 1	-1	1	1	1	-1	1	-1	1	-1	1
1 1 2	-1	1	-1	1	0	-2	0	-2	1	1
1 2 0	-1	1	0	-2	1	1	1	1	0	-2
2 0 2	0	-2	1	1	0	-2	0	-2	0	-2
2 1 0	0	-2	-1	1	1	1	1	1	-1	1
2 2 1	0	-2	0	-2	-1	1	-1	1	1	1
$A B C$	AC_1	AC_2	AC_1^2	AC_2^2	BC_1	BC_2	BC_1^2	BC_2^2		
0 0 0	1	1	1	1	1	1	1	1		
0 1 1	0	-2	-1	1	0	-2	1	1		
0 2 2	-1	1	0	-2	-1	1	1	1		
1 0 1	-1	1	1	1	-1	1	-1	1		
1 1 2	1	1	-1	1	1	1	-1	1		
1 2 0	0	-2	0	-2	0	-2	-1	1		
2 0 2	0	-2	1	1	0	-2	0	-2		
2 1 0	-1	1	-1	1	-1	1	0	-2		
2 2 1	1	1	0	-2	1	1	0	-2		
$A B C$	ABC_1	ABC_2	ABC_1^2	ABC_2^2	$AB^2C_1^2$	$AB^2C_2^2$	AB^2C_1	AB^2C_2		
0 0 0	1	1	0	-2	1	1	1	1		
0 1 1	-1	1	0	-2	-1	1	1	1		
0 2 2	0	-2	0	-2	0	-2	1	1		
1 0 1	-1	1	0	-2	1	1	-1	1		
1 1 2	0	-2	0	-2	-1	1	-1	1		
1 2 0	1	1	0	-2	0	-2	-1	1		
2 0 2	0	-2	0	-2	1	1	0	-2		
2 1 0	1	1	0	-2	-1	1	0	-2		
2 2 1	-1	1	0	-2	0	-2	0	-2		

Cabe notar que las columnas:

- A_1 y A_2 son iguales a las columnas BC_1^2 y BC_2^2 , respectivamente, y son iguales a las columnas AB^2C_1 y AB^2C_2 , respectivamente.
- B_1 y B_2 son iguales a las columnas AC_1^2 y AC_2^2 , respectivamente, y son iguales a las columnas $AB^2C_1^2$ y $AB^2C_2^2$, respectivamente.
- C_1 y C_2 son iguales a las columnas AB_1 y AB_2 , respectivamente, y son iguales a las columnas ABC_1 y ABC_2 , respectivamente.
- AB_1^2 y AB_2^2 son iguales a las columnas AC_1 y AC_2 , respectivamente, y son iguales a las columnas BC_1 y BC_2 , respectivamente.

Así, el conjunto de alias generado es:

$$\begin{aligned}A &= BC^2 = AB^2C \\B &= AC^2 = AB^2C^2 \\C &= AB = ABC \\AC &= BC = AB^2.\end{aligned}$$

Entonces:

- A , BC^2 y AB^2C son alias, y el estimador asociado a A estima en realidad $A + BC^2 + AB^2C$.
- B , AC^2 y AB^2C^2 son alias, y el estimador asociado a B estima en realidad $B + AC^2 + AB^2C^2$.
- C , AB y ABC son alias, y el estimador asociado a C estima en realidad $C + AB + ABC$.
- AC , BC y AB^2 son alias, y el estimador asociado a AC estima en realidad $AC + BC + AB^2$.



Nota 6.9. De forma similar al diseño 2^k , una forma de obtener los alias es multiplicar la relación de definición del diseño por cada efecto, pero en este caso también se multiplica por la relación de definición al cuadrado (segundos alias) [33, p. 491]. En el ejemplo 6.7 se tiene que:

$I =$	ABC^2	$=$	$(ABC^2)^2$
$I \times A =$	A^2BC^2	$=$	$A^3B^2C^4$
$A =$	$(A^2BC^2)^2$	$=$	B^2C
$A =$	$A^4B^2C^4$	$=$	$(B^2C)^2$
$A =$	AB^2C	$=$	BC^2
$I \times B =$	AB^2C^2	$=$	$A^2B^3C^4$
$B =$	AB^2C^2	$=$	$(A^2C)^2$
$B =$	AB^2C^2	$=$	AC^2
$I \times C =$	ABC^3	$=$	$A^2B^2C^5$
$C =$	AB	$=$	$(A^2B^2C^2)^2$
$C =$	AB	$=$	ABC
$I \times AC =$	A^2BC^3	$=$	$A^3B^2C^5$
$AC =$	$(A^2B)^2$	$=$	$(B^2C^2)^2$
$AC =$	AB^2	$=$	BC

Ejemplo 6.8. Retomando el ejemplo 6.6, suponga que solo se podía realizar una fracción del experimento, la fracción asociada al bloque I, ver tabla 6.17. Realice el análisis de varianza.

Solución. Al tener solo 9 datos, antes de realizar el análisis de varianza, es necesario identificar los alias para incluir en el modelo solo aquellos efectos que pueden ser estimables si sus alias no son significativos. Vale recordar que

P indica peso en toneladas, L indica pendiente longitudinal del tramo en % y V indica velocidad en km/h .

$I =$	PL^2V	$=$	$(PL^2V)^2$
$I \times P =$	P^2L^2V	$=$	$P^3L^4V^2$
$P =$	$(P^2L^2V)^2$	$=$	LV^2
$P =$	PLV^2	$=$	LV^2
$I \times L =$	PL^3V	$=$	$P^2L^5V^2$
$L =$	PV	$=$	$(P^2L^2V^2)^2$
$L =$	PV	$=$	PLV
$I \times V =$	PL^2V^2	$=$	$P^2L^4V^3$
$V =$	PL^2V^2	$=$	$(P^2L)^2$
$V =$	PL^2V^2	$=$	PL^2
$I \times PL =$	P^2L^3V	$=$	$P^3L^5V^2$
$PL =$	$(P^2V)^2$	$=$	$(L^2V^2)^2$
$PL =$	PV^2	$=$	LV

Luego, solo se van a incluir en el modelo los efectos principales, ya que las interacciones dobles son alias entre sí y no se podría distinguir si son significativas. En la tabla 6.23 se presenta el análisis de varianza en que se observa que la velocidad no es significativa y se ajusta el modelo solo considerando el peso y la pendiente longitudinal del tramo, ver tabla 6.24. En la figura 6.13 se presentan los valores ajustados para el modelo seleccionado.

Tabla 6.23. Análisis de varianza para los datos del ejemplo 6.8

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Peso	2	514 126	257 063	26.874	0.0359
Pendiente	2	561 693	280 846	29.361	0.0329
Velocidad	2	59 801	29 900	3.126	0.2424
Error	2	19 131	9565		

Tabla 6.24. Análisis de varianza para los datos del ejemplo 6.8 considerando solo peso y pendiente longitudinal del tramo

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Peso	2	514 126	257 063	13.03	0.0177
Pendiente	2	561 693	280 846	14.23	0.0152
Error	4	78 932	19 733		

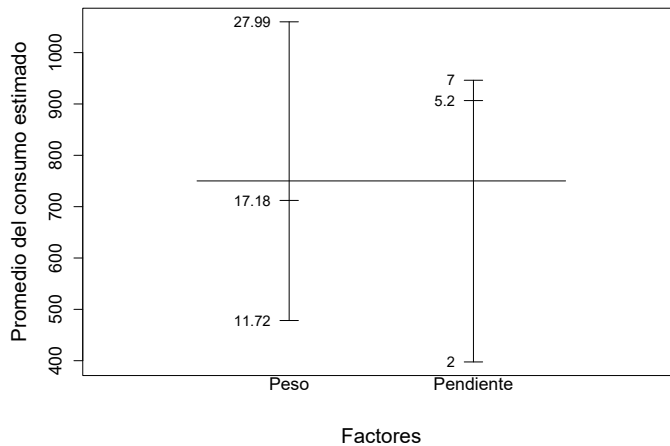


Figura 6.13. Valores ajustados promedio - ejemplo 6.8

Las instrucciones en R para el ajuste del modelo en el ejemplo 6.8 son:

```
library(DoE.base); library(dplyr)
remove(list=ls())
dis=fac.design(factor.names=list(Peso=c(11.72,17.18,27.99),
                                Pendiente=c(2,5.2,7), Velocidad=c(25,35,45)),
```

Confusión y diseños con arreglos factoriales fraccionados

```
nlevels=c(3,3,3), blocks=3,block.gen=c(1,2,1),
random=F) # factor de confusión PL^2V
dis=data.frame(dis%>%filter(Blocks==1))
Consumo=c(286.82,917.19,1368.70,558.64,550.04,871.99,347.26,
          1252.65,597.91)
datos=data.frame(dis,Consumo)
summary(aov(Consumo~Peso+Pendiente+Velocidad,data=datos))
ajuste=aov(Consumo~Peso+Pendiente,data=datos)
summary(ajuste)
Peso=datos$Peso
Pendiente=datos$Pendiente
Ajustados=ajuste$fitted
plot.design(Ajustados~Peso+Pendiente,
            ylab="Promedio del consumo estimado",
            xlab="Factores",cex.axis=1.2,cex.lab=1.4)
hnp::hnp(ajuste,halfnormal=F)
```



Nota 6.10. En cuanto a la resolución del diseño, aplica el mismo concepto visto en las notas de clase de diseño con arreglos factoriales 2^k sobre diseño con arreglos factoriales fraccionado 2^{k-1} . Desde este punto de vista el diseño usado en el ejemplo 6.8 sería de resolución III.

Nota 6.11 (Efectos clear o transparentes). Según [61, p. 215], un efecto principal es llamado **clear o transparente** si no es alias de otro efecto principal, de igual forma componentes de interacciones de dos factores son llamados clear o transparentes si no son alias de otras interacciones de dos factores. Por ejemplo para la interacción $A \times B$ es llamada clear o transparente si sus componentes AB y AB^2 no son alias de ningún otro componente de dos factores.

5. Pseudofactores

Según [21, p. 482], para diseños con más de tres niveles por factor, por ejemplo, para diseño 4^k se puede usar el concepto de pseudofactores de tal manera que el diseño sería 2^{2k} . Un diseño 4^3 se puede escribir como $2^{2 \times 3}$ implicando que tiene los siguientes pseudofactores:

$$A = A_1 A_2$$

$$B = B_1 B_2$$

$$C = C_1 C_2.$$

A	A_1	A_2	B	B_1	B_2	C	C_1	C_2
0	0	0	0	0	0	0	0	0
1	0	1	1	0	1	1	0	1
2	1	0	2	1	0	2	1	0
3	1	1	3	1	1	3	1	1

Siendo A_1 y A_2 los dos pseudofactores del factor A , B_1 y B_2 los dos pseudofactores del factor B y C_1 y C_2 los dos pseudofactores del factor C , cada pseudofactor con dos niveles. Así el diseño se puede escribir como $A_1 \times A_2 \times B_1 \times B_2 \times C_1 \times C_2$.

Otro ejemplo puede ser un diseño 6^{3-1} . En este caso se puede escribir como $2^{3-1} 3^{3-1}$ con sus respectivos pseudofactores:

$$A = A_1 A_2$$

$$B = B_1 B_2$$

$$C = C_1 C_2.$$

A	x_1 A_1	x_2 A_2	B	x_3 B_1	x_4 B_2	C	x_5 C_1	x_6 C_2
0	0	0	0	0	0	0	0	0
1	0	1	1	0	1	1	0	1
2	0	2	2	0	2	2	0	2
3	1	0	3	1	0	3	1	0
4	1	1	4	1	1	4	1	1
5	1	2	5	1	2	5	1	2

Siendo A_1 y A_2 los dos pseudofactores del factor A , B_1 y B_2 los dos pseudofactores del factor B y C_1 y C_2 los dos pseudofactores del factor C , en que el subíndice 1 indica que el pseudofactor tiene, por ejemplo, dos niveles, y el subíndice 2 indica que el pseudofactor tiene tres niveles.

Si, por ejemplo, se usan como generadores del diseño $A_1B_1C_1 (x_1 + x_3 + x_5)$ y $A_2B_2C_2^2 (x_2 + x_4 + 2x_6)$ para seleccionar los tratamientos que se van a aplicar, suponiendo que la asignación a los bloques es:

$$\begin{aligned}
 A_1B_1C_1 = 0 \quad \text{y} \quad A_2B_2C_2^2 = 0 & \quad \text{entonces} \quad \text{Bloque I} \\
 A_1B_1C_1 = 0 \quad \text{y} \quad A_2B_2C_2^2 = 1 & \quad \text{entonces} \quad \text{Bloque II} \\
 A_1B_1C_1 = 0 \quad \text{y} \quad A_2B_2C_2^2 = 2 & \quad \text{entonces} \quad \text{Bloque III} \\
 A_1B_1C_1 = 1 \quad \text{y} \quad A_2B_2C_2^2 = 0 & \quad \text{entonces} \quad \text{Bloque IV} \\
 A_1B_1C_1 = 1 \quad \text{y} \quad A_2B_2C_2^2 = 1 & \quad \text{entonces} \quad \text{Bloque V} \\
 A_1B_1C_1 = 1 \quad \text{y} \quad A_2B_2C_2^2 = 2 & \quad \text{entonces} \quad \text{Bloque VI,}
 \end{aligned}$$

en la tabla 6.25 se muestra una sección de los $6^3 = 216$ combinaciones de tratamientos, su asignación a los bloques y si son seleccionados o no para el experimento. Finalmente, la fracción seleccionada para el experimento corresponde a los 36 tratamientos asignados al bloque I.

La única diferencia es que una interacción de pseudofactores de la forma $A_1A_2B_1$ no representa una interacción de tres factores. Esta representa solo uno de los contrastes ortogonales de la interacción de los dos factores A y B .

Tabla 6.25. Asignación de los 6^{3-1} tratamientos usando las componentes $A_1B_1C_1$ y $A_2B_2C_2^2$

A_1	A_2	B_1	B_2	C_1	C_2	$A_1B_1C_1 = x_1 + x_3 + x_5 \pmod{2}$	$A_2B_2C_2^2 = x_2 + x_3 + 2x_6 \pmod{3}$	Bloque	Tratam. seleccionado
0	0	0	0	0	0	0	0	Bloq. I	SI
0	0	0	0	0	1	0	2	Bloq. III	NO
0	0	0	0	0	2	0	1	Bloq. II	NO
0	0	0	1	0	0	0	1	Bloq. II	NO
0	0	0	1	0	1	0	0	Bloq. I	SI
0	0	0	1	0	2	0	2	Bloq. III	NO
...	...	...	...	...	...	...	...	...	...
0	2	1	2	1	0	0	1	Bloq. II	NO
0	2	1	2	1	1	0	0	Bloq. I	SI
0	2	1	2	1	2	0	2	Bloq. III	NO
1	0	0	0	0	0	1	0	Bloq. IV	NO
1	0	0	0	0	1	1	2	Bloq. VI	NO
1	0	0	0	0	2	1	1	Bloq. V	NO
...	...	...	...	...	...	...	...	...	...
1	2	1	1	1	0	1	0	Bloq. IV	NO
1	2	1	1	1	1	1	2	Bloq. VI	NO
1	2	1	1	1	2	1	1	Bloq. V	NO
1	2	1	2	1	0	1	1	Bloq. V	NO
1	2	1	2	1	1	1	0	Bloq. IV	NO
1	2	1	2	1	2	1	2	Bloq. VI	NO

6. Diseño central compuesto

Cuando un investigador desea optimizar procesos es usual usar metodologías de superficies de respuesta, que precisamente buscan desarrollar, mejorar y optimizar procesos.

Considere un diseño con arreglos factoriales 2^k (en principio con k factores cuantitativos). Un problema (potencial) si se desea encontrar un óptimo, es el supuesto de linealidad en los efectos de los factores, ya que solo se tienen dos niveles por factor.

Una forma de afrontar este problema (potencial) es, en vez de usar un diseño con arreglos factoriales 2^k o un diseño con arreglos factoriales fraccionado 2^{k-p} , usar un **diseño central compuesto** que consiste en adicionar ya sea al diseño con arreglos factoriales 2^k o al diseño con arreglos factoriales 2^k fraccionado (de resolución V) n_c puntos centrales y $2k$ puntos axiales [43, p. 331].

Los **puntos centrales**, como el nombre lo indica, implica agregar puntos en el centro del campo de experimentación, por ejemplo, considere un diseño 2^2 con 3 puntos centrales en que el factor A tiene los niveles 2.5 y 5, y el factor B tiene los niveles 540 y 1080, como en el ejemplo 5.10 sobre porcentaje de degradación de clorpirifós, ver tabla 6.26.

Tabla 6.26. Diseño con arreglos factoriales 2^2 con tres puntos centrales

Niveles codificados		Niveles no codificados	
Factor A	Factor B	Factor A	Factor B
-1	-1	2.5	540
1	-1	5	540
-1	1	2.5	1080
1	1	5	1080
0	0	3.75	810
0	0	3.75	810
0	0	3.75	810

} puntos
centrales

En cuanto a los **puntos axiales**, que también se les llama **puntos estrella**, son puntos que quedan sobre los ejes del campo de experimentación. En el ejemplo de la tabla 6.26, los cuatro puntos axiales son $(-\alpha, 0)$, $(\alpha, 0)$, $(0, -\alpha)$ y $(0, \alpha)$, por ejemplo $(0, \alpha)$ indica que el factor A está en su punto central y que el factor B está en el nivel α , mientras que el punto $(-\alpha, 0)$ indicaría que el factor A está en el nivel $-\alpha$ y que el factor B está en su punto central, ver tabla 6.27.

Tabla 6.27. Diseño con arreglos factoriales 2^2 con tres puntos centrales y puntos axiales

Niveles codificados		Niveles no codificados	
Factor A	Factor B	Factor A	Factor B
-1	-1	2.5	540
1	-1	5	540
-1	1	2.5	1080
1	1	5	1080
0	0	3.75	810
0	0	3.75	810
0	0	3.75	810
$-\alpha$	0	$-1.25\alpha + 3.75$	810
α	0	$1.25\alpha + 3.75$	810
0	$-\alpha$	3.75	$-270\alpha + 810$
0	α	3.75	$270\alpha + 810$

} puntos
axiales

Para la selección de la distancia axial, α , y para el número de puntos centrales se puede usar³ [21, p. 587]:

$$\begin{aligned}\alpha &= (n_f)^{1/4} \\ n_c &= 4\sqrt{n_f} + 4 - 2k,\end{aligned}\tag{6.14}$$

en que n_f corresponde al número de puntos factoriales. Note que no siempre n_c es un número entero. En el ejemplo de la tabla 6.27 se tiene $n_f = 4$ puntos factoriales, luego:

$$\alpha = 4^{1/4} = 1.414.$$

Por otro lado, en cuanto al número de puntos centrales serían:

$$n_c = 4\sqrt{n_f} + 4 - 2k = 4\sqrt{4} + 4 - (2 \times 2) = 8.$$

Dejando los tres puntos centrales, esto es, ignorando que $n_c = 8$ y usando $\alpha = 1.414$, el diseño presentado en la tabla 6.27 queda de acuerdo a la tabla 6.28.

³Usar las dos especificaciones de α y n_c dadas en (6.14) genera un diseño rotatable y ortogonal. Usar α , pero no el n_c sugerido, genera un diseño rotatable, pero no ortogonal.

Tabla 6.28. Diseño con arreglos factoriales 2^2 con tres puntos centrales y puntos axiales, $\alpha = 1.414$.

Niveles codificados		Niveles no codificados	
Factor A	Factor B	Factor A	Factor B
-1	-1	2.5	540
1	-1	5	540
-1	1	2.5	1080
1	1	5	1080
0	0	3.75	810
0	0	3.75	810
0	0	3.75	810
-1.414	0	1.98	810
1.414	0	5.52	810
0	-1.414	3.75	428.2
0	1.414	3.75	1191.8

Para relacionar los niveles codificados, x_i , con los niveles en sus unidades originales, digamos a_i , se usa [43, p. 49]:

$$x_i = \frac{a_i - \frac{\max\{a_0, a_1\} + \min\{a_0, a_1\}}{2}}{\frac{\max\{a_0, a_1\} - \min\{a_0, a_1\}}{2}}.$$

El código en R para obtener el diseño presentado en la tabla 6.28 es:

```
library(rsm)
dis = ccd(2, n0 = c(3,0), alpha = "rotatable", randomize = FALSE)
dis

## descodificando
dis = ccd(2, n0 = c(3,0), alpha = "rotatable", randomize = FALSE,
  coding=list(x1~(rela-3.75)/1.25, x2~(presion-810)/270))
dis
```

El **diseño central compuesto** (ccd) es útil para ajustar modelos de segundo orden, esto es para identificar curvatura no solo generada por las interacciones,

sino también por términos cuadráticos. Primero se va a considerar el **modelo de primer orden** o **modelo de efectos principales** dado por [43, p. 23]:

$$\begin{aligned}\eta &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &= \beta_0 + \underbrace{\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}}_{\mathbf{x}^\top} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\mathbf{b}} = \beta_0 + \mathbf{x}^\top \mathbf{b},\end{aligned}\tag{6.15}$$

en que x_1, x_2 y x_3 son factores cuantitativos, $\beta_0, \beta_1, \beta_2$ y β_3 son parámetros del modelo y $\eta = \mathbb{E}(Y) = \mu$ para el caso de un modelo lineal o $\eta = g[\mathbb{E}(Y)] = g[\mu]$ para el caso de un modelo lineal generalizado, vale resaltar que $\eta = \mathbb{E}(Y)$ es un caso particular de un modelo lineal generalizado. Si al modelo (6.15) se le agregan interacciones, se tiene un **modelo de primer orden con interacción**:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3.$$

Por otro lado, un **modelo de segundo orden** es dado por:

$$\begin{aligned}\eta &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \\ &\quad \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \\ \eta &= \beta_0 + \mathbf{x}^\top \mathbf{b} + \underbrace{\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}}_{\mathbf{x}^\top} \underbrace{\begin{pmatrix} \beta_{11} & \frac{\beta_{12}}{2} & \frac{\beta_{13}}{2} \\ \frac{\beta_{12}}{2} & \beta_{22} & \frac{\beta_{23}}{2} \\ \frac{\beta_{13}}{2} & \frac{\beta_{23}}{2} & \beta_{33} \end{pmatrix}}_{\mathbf{B}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\mathbf{x}} \\ &= \beta_0 + \mathbf{x}^\top \mathbf{b} + \mathbf{x}^\top \mathbf{B} \mathbf{x}.\end{aligned}$$

Ejemplo 6.9. Considere el experimento presentado por [42] sobre el proceso de absorción del CO_2 de un biogás en una solución alcalina en que consideraron los factores concentración inicial de la solución alcalina (relación $\text{Na}_2\text{CO}_3 / \text{NaHCO}_3$) a los niveles 1.35 y 4.04, presión inicial en el reactor a los niveles 1241.05 mbar y 1861.58 mbar y la temperatura del sistema gas-líquido a los niveles 25°C y 35°C . La variable respuesta que se va a usar es el cambio del % del CO_2 en el biogás. Los investigadores realizaron un diseño central compuesto con 4 puntos centrales. Los datos se presentan en la tabla 6.29. Ajuste el modelo y presente la superficie de respuesta estimada.

Tabla 6.29. Datos ejemplo sobre absorción del CO_2 de un biogás

Relación	Presión	Temperatura	Cambio del % de CO_2
-1	-1	-1	9.5744
1	-1	-1	14.1021
-1	1	-1	8.8388
1	1	-1	14.2295
-1	-1	1	10.1957
1	-1	1	18.0527
-1	1	1	9.5429
1	1	1	16.6059
0	0	0	11.0063
0	0	0	11.3211
0	0	0	10.5261
0	0	0	11.0952
-1.68	0	0	6.5599
1.68	0	0	14.6750
0	-1.68	0	12.1961
0	1.68	0	10.7009
0	0	-1.68	9.2532
0	0	1.68	12.1545

Solución. Inicialmente se ajustó un modelo de segundo orden, ver tabla 6.30, en que se observa que los términos cuadráticos para relación y para temperatura, y las interacciones entre relación y presión, y entre presión y temperatura tienen P-valor mayor a 0.20, por lo que se retiran del modelo, ver tabla 6.31.

Tabla 6.30. Estimación de parámetros para el modelo de superficie de respuesta ajustado a los datos del ejemplo 6.9

Término	Estimador	Error estándar	Valor t	P-valor
Intercepto	10.9037	0.6066	17.9753	<0.0001
x_1 (Relación)	2.8181	0.3288	8.5716	<0.0001
x_2 (Presión)	-0.3824	0.3288	-1.1631	0.278
x_3 (Temperatura)	0.9176	0.3288	2.7910	0.024
$x_1 : x_2$	0.0086	0.4296	0.0201	0.984
$x_1 : x_3$	0.6252	0.4296	1.4554	0.184
$x_2 : x_3$	-0.1864	0.4296	-0.4340	0.676
x_1^2	0.2424	0.3416	0.7096	0.498
x_2^2	0.5362	0.3416	1.5697	0.155
x_3^2	0.2730	0.3416	0.7991	0.447

Tabla 6.31. Estimación de parámetros para el modelo de superficie de respuesta ajustado a los datos del ejemplo 6.9

Término	Estimador	Error estándar	Valor t	P-valor
Intercepto	11.3625	0.3314	34.281	<0.0001
x_1 (Relación)	2.8181	0.2868	9.824	<0.0001
x_2 (Presión)	-0.3824	0.2868	-1.333	0.2073
x_3 (Temperatura)	0.9176	0.2868	3.199	0.0076
$x_1 : x_3$	0.6252	0.3748	1.668	0.1211
x_2^2	0.4470	0.2870	1.557	0.1454

En la figura 6.14 se presentan perspectivas de la superficie de respuesta ajustada.

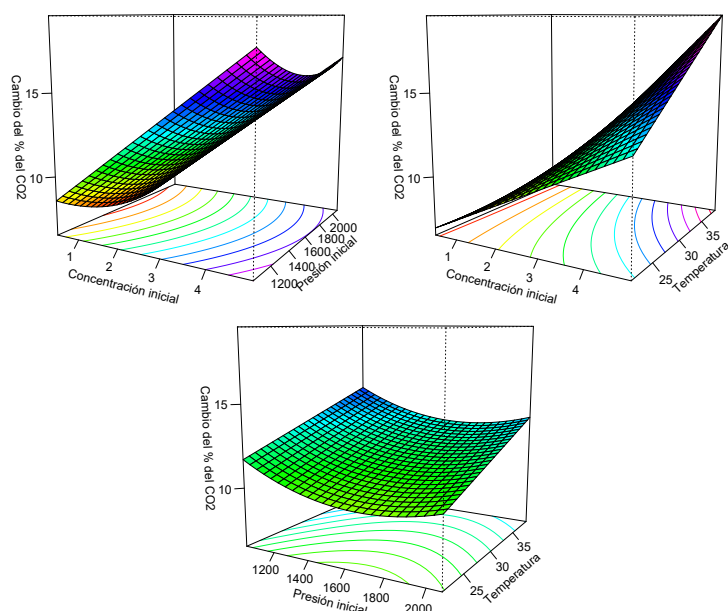


Figura 6.14. Perspectivas de la superficie de respuesta ajustada a los datos del ejemplo 6.9

Confusión y diseños con arreglos factoriales fraccionados

Las instrucciones en R usadas en el ajuste del modelo son:

```
library(rsm);library(plotly);library(ggplot2);library(data.table)
remove(list=ls())
rela=c(1.35,4.04);(rela[2]+rela[1])/2;(rela[2]-rela[1])/2
presion=c(1241.05,1861.58);(presion[2]+presion[1])/2
(presion[2]-presion[1])/2
temp=c(25,35);(temp[2]+temp[1])/2;(temp[2]-temp[1])/2
dis = ccd(3, n0 = c(4,0), alpha = "rotatable",
coding=list(x1~(rela-2.695)/1.345,x2~(presion-1551.315)/310.265,
            x3~(Temp-30)/5),randomize = FALSE)
dis
cambio=c(9.5744,14.1021,8.8388,14.2295,10.1957,18.0527,9.5429,
          16.6059,11.0063,11.3211,10.5261,11.0952,6.5599,14.6750,
          12.1961,10.7009,9.2532,12.1545)
summary(rsm(cambio~S0(x1,x2,x3),data=dis))
ajuste=rsm(cambio~F0(x1,x2,x3)+TWI(x1,x3)+PQ(x2),data=dis)
summary(ajuste)
op <- par(no.readonly = TRUE)
par(mar=c(0.8,0.8,0.8,0.8))
layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))
persp(ajuste,~x1+x2+x3,xlabs=c("Concentración inicial",
"Presión inicial","Temperatura"),
      zlab = "Cambio del % del CO2",
      contours = list(z="bottom",col="colors"), #posición y color
      at = c(summary(ajuste$canonical$xs)),theta=30,phi=10)
par(op); par(mfrow=c(1,1))
hnp::hnp(ajuste,halfnormal=F)
## Graficos 3D
par(mfrow=c(2,2))
a=contour(ajuste,~x1+x2+x3,image=TRUE)
par(mfrow=c(1,1))
## a[[1]] corresponde a Relación (x1) con Presión (x2) con
## temperatura=30 (temperatura central)
X1=a[[1]]$x; X2=a[[1]]$y; X3=a[[1]]$z
figura = plot_ly(x=~X2,y=~X1,z=~X3) %>%
add_surface(showscale=FALSE,
contours = list(z = list(show=TRUE,usecolormap=TRUE,
```

```

        project=list(z=TRUE)))
figura%>%layout(scene=list(yaxis =
list(title = "Concentración inicial",
range = c(0.3,5)),xaxis = list(title = "Presión inicial",
autorange="reversed",range = c(1000,2100)),
zaxis=list(title="Cambio del % del CO2"))
## a[[2]] corresponde a Relación (x1) con Temperaturan (x3) con
## presión=1551,3 (presión central)
X1=a[[2]]$x; X2=a[[2]]$y; X3=a[[2]]$z
figura = plot_ly(x=~X2,y=~X1,z=~X3) %>%
    add_surface(showscale=FALSE,
    contours = list(z = list(show=TRUE,usecolormap=TRUE,
project=list(z=TRUE))))
figura%>%layout(scene=list(yaxis =
    list(title = "Concentración inicial",
range = c(0.3,5)),xaxis = list(title = "Temperatura",
autorange="reversed",
range = c(21,39)),zaxis=list(title="Cambio del % del CO2"))
## a[[3]] corresponde a Presión (x2) con Temperaturan (x3) con
## relación=2,69 (relación central)
X1=a[[3]]$x; X2=a[[3]]$y; X3=a[[3]]$z
figura = plot_ly(x=~X2,y=~X1,z=~X3) %>%
    add_surface(showscale=FALSE,
contours = list(z = list(show=TRUE,usecolormap=TRUE,
project=list(z=TRUE))))
figura%>%layout(scene=list(yaxis =
    list(title = "Presión inicial"),
axis = list(title = "Temperatura",autorange="reversed",
range = c(21,39)),zaxis=list(title="Cambio del % del CO2"))

```



7. Ejercicios

- 7.1 Presentar la tabla de análisis de varianza con las columnas: causa de variación, grados de libertad y sumas de cuadrados, para el diseño pre-

sentado en la tabla 6.6, asumiendo submuestreo en que n^* es el número de unidades observacionales por unidad experimental.

- 7.2 Realizar el análisis de residuales al modelo seleccionado en el ejemplo 6.4. Incluir gráficos.
- 7.3 Para tres niveles desigualmente espaciados x_1, x_2 y x_3 de un factor cuantitativo, construya polinomios ortogonales lineales y cuadráticos comenzando con las columnas de entradas $(1,1,1)$, (x_1, x_2, x_3) , (x_1^2, x_2^2, x_3^2) y usando el procedimiento de ortonormalización de Gram-Schmidt [61, p. 244]. Ajuste los datos del ejemplo 6.6.
- 7.4 Para un diseño 3_{IV}^{6-2} con generadores del diseño ABC y AB^2D , encuentre, justificando adecuadamente, todos los efectos clear o transparentes [61, p. 247].
- 7.5 Considere un diseño $2^3 \times 3^3$. Obtenga un posible sistema de confusión para bloques de tamaño 6 [33, p. 495].
- 7.6 Construya un sistema de confusión parcial para el diseño con arreglos factoriales 3^2 en bloques de tamaño 3 tal que al menos información parcial pueda ser obtenida acerca de la interacción de dos componentes y que se tenga información completa de los efectos principales [33, p. 495].
- 7.7 Construir un diseño 3^{5-2} con $I = ABC$ e $I = CDE$ [41, p. 389]. Escribir la estructura de los alias de este diseño. ¿Cuál es la resolución de este diseño?
- 7.8 Sugiera un sistema de confusión para un experimento 3^5 en 27 bloques de tamaño 9 si todas las interacciones de dos factores y la interacción de tres factores ABE son estimables [21, p. 491].
- 7.9 Sugiera un esquema de confusión para un diseño con arreglos factoriales $2^2 \times 3^2 \times 6$ en 12 bloques de tamaño 18 [21, p. 493]. ¿Bajo qué circunstancias este diseño sería útil? Presente dos bloques del diseño.
- 7.10 Derive la relación de alias para el diseño 5^{3-1} con $C = AB$ [61].

7.11 Considere el modelo

$$Y_{isj} = \mu + \tau_i + \gamma_s + (\tau\gamma)_{is} + \beta_{j(s)} + e_{isj},$$

en que Y_{isj} denota la respuesta del individuo en el tratamiento i , localidad s , bloque j .

Si

$$\sum_{i=1}^a \tau_i = \sum_{s=1}^k \gamma_s = \sum_{i=1}^a (\tau\gamma)_{is} = \sum_{s=1}^k (\tau\gamma)_{is} = \sum_{j=1}^b \beta_{j(s)} = 0$$

entonces

μ : es la media general.

τ_i : es el efecto del tratamiento i .

γ_s : es el efecto de la localidad s .

$(\tau\gamma)_{is}$: es el efecto de la interacción tratamiento por localidad.

$\beta_{j(s)}$: es el efecto del bloque j anidado en la localidad s .

Asumiendo que τ_i y γ_s son efectos fijos, y $\beta_{j(s)}$ es un efecto aleatorio:

- Presente la matriz diseño matricialmente considerando $a = 2$, $k = 3$ y $b = 3$.
- Construya la tabla de análisis de varianza con causa de variación, grados de libertad, sumas de cuadrados, cuadrado medio, esperanza del cuadrado medio y F-calculado.

Capítulo 7

Diseño anidado en dos etapas

En este capítulo se realiza una breve introducción al diseño anidado en 2 etapas presentando tres casos, uno en que los dos factores son de efectos fijos, otro en que el factor no anidado es de efectos fijos y el factor anidado es de efectos aleatorios y el último cuando los dos factores son de efectos aleatorios.

Un factor B se dice que está **anidado** dentro de un factor A si cada uno de los niveles de A es observado solo con un nivel del factor B [21, p. 671].

Ejemplo 7.1. *Ilustrativamente, considere una empresa manufacturera de moldes de plástico que desea reemplazar el ingrediente tradicional por una alternativa más barata. Para determinar si la alternativa más barata es adecuada, diseñan un experimento en que consideran, además de los dos ingredientes, las fábricas, ya que la compañía cuenta con tres fábricas en diferentes partes del país, en las que el clima es diferente (consideran que puede afectar el producto). Adicionalmente, para tener en cuenta el factor humano, en cada fábrica el experimento será realizado por dos operadores y se replicará r veces. Ver tabla 7.1.*

Tabla 7.1. Datos de un diseño anidado en dos etapas - ejemplo

Ingrediente 1						Ingrediente 2					
Lugar 1		Lugar 2		Lugar 3		Lugar 1		Lugar 2		Lugar 3	
Operario		Operario		Operario		Operario		Operario		Operario	
1	2	3	4	5	6	7	8	9	10	11	12
y ₁₁₁₁	y ₁₁₂₁	y ₁₂₃₁	y ₁₂₄₁	y ₁₃₅₁	y ₁₃₆₁	y ₂₁₇₁	y ₂₁₈₁	y ₂₂₉₁	y _{2,2,10,1}	y _{2,3,11,1}	y _{2,3,12,1}
y ₁₁₁₂	y ₁₁₂₂	y ₁₂₃₂	y ₁₂₄₂	y ₁₃₅₂	y ₁₃₆₂	y ₂₁₇₂	y ₂₁₈₂	y ₂₂₉₂	y _{2,2,10,2}	y _{2,3,11,2}	y _{2,3,12,2}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
y _{111r}	y _{112r}	y _{123r}	y _{124r}	y _{135r}	y _{136r}	y _{217r}	y _{218r}	y _{229r}	y _{2,2,10,r}	y _{2,3,11,r}	y _{2,3,12,r}

Si el ingrediente es el factor A , el lugar el factor B y el operario el factor C , entonces los factores A y B están cruzados, y el factor C está anidado en la interacción AB .

Ejemplo 7.2. *Ilustrativamente considere un experimento con seis tipos de máquinas para medir ozono (en ppb - partes por billón) en tres estaciones de medición. Dos empresas manufacturan estas máquinas, cada una manufactura tres tipos de máquinas. Los valores de ozono recolectados se presentan en la tabla 7.2.*

Tabla 7.2. Datos ilustrativos sobre nivel de ozono (O_3 en ppb) - ejemplo 7.2

Estación	Empresa 1			Empresa 2		
	Máq. tipo 1	Máq. tipo 2	Máq. tipo 3	Máq. tipo 4	Máq. tipo 5	Máq. tipo 6
1	18	25	15	20	21	63
2	30	45	42	35	49	23
3	53	45	104	50	58	147

1. El modelo

Según [21, p. 672], sean dos factores en que el factor B está anidado en el factor A . El respectivo modelo es dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{ijk}, \quad (7.1)$$

con $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$, $k = 1, 2, \dots, r$, siendo:

- y_{ijk} es la respuesta de la unidad experimental que fue asignada a la réplica k del nivel j del factor B que está anidado en el nivel i del factor A .
- e_{ijk} es el componente asociado con el error aleatorio que se asume $e_{ijk} \sim N(0, \sigma_e^2)$.

Asumiendo las restricciones:

$$\sum_{i=1}^a \alpha_i = 0 \quad y \quad \sum_{j=1}^b \beta_{j(i)} = 0, \quad i = 1, \dots, a,$$

se tiene que:

- α_i es el efecto medio del nivel i del factor A con respecto a la media general.
- $\beta_{j(i)}$ es el efecto medio del nivel j del factor B en el nivel i del factor A con respecto a la media general.

Matricialmente

Considerando las restricciones para los α y los β se tiene que:

$$\begin{aligned}\alpha_1 + \alpha_2 + \cdots + \alpha_a &= 0 \quad \text{luego} \quad \alpha_a = -\alpha_1 - \alpha_2 - \cdots - \alpha_{a-1} \\ \beta_{1(1)} + \beta_{2(1)} + \cdots + \beta_{b(1)} &= 0 \quad \text{luego} \quad \beta_{b(1)} = -\beta_{1(1)} - \beta_{2(1)} - \cdots - \beta_{b-1(1)} \\ &\vdots \\ \beta_{1(a)} + \beta_{2(a)} + \cdots + \beta_{b(a)} &= 0 \quad \text{luego} \quad \beta_{b(a)} = -\beta_{1(a)} - \beta_{2(a)} - \cdots - \beta_{b-1(a)}.\end{aligned}$$

Así, el modelo matricialmente escrito es:

$$Y = X\beta + E$$

$$\begin{bmatrix} Y_{111} \\ \vdots \\ Y_{11r} \\ Y_{121} \\ \vdots \\ Y_{12r} \\ \vdots \\ Y_{1,b-1,1} \\ \vdots \\ Y_{1,b-1,r} \\ Y_{1b1} \\ \vdots \\ Y_{1br} \\ Y_{211} \\ \vdots \\ Y_{21r} \\ \vdots \\ Y_{2b1} \\ \vdots \\ Y_{2br} \\ \vdots \\ Y_{a11} \\ \vdots \\ Y_{a1r} \\ \vdots \\ Y_{ab1} \\ \vdots \\ Y_{abr} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & -1 & -1 & \cdots & -1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & -1 & -1 & \cdots & -1 & \cdots & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 & 0 & 0 & \cdots & 0 & \cdots & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 & 0 & 0 & \cdots & 0 & \cdots & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{a-1} \\ \beta_{1(1)} \\ \beta_{2(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{bmatrix} + e,$$

en que $\mathbf{X}^\top \mathbf{X}$ se puede escribir como:

$$\mathbf{X}^\top \mathbf{X} = \underbrace{abr \oplus br}_{(a-1) \times (a-1)} \oplus \left[\mathbf{I}_a \otimes r \underbrace{\begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{pmatrix}}_{(b-1) \times (b-1)} \right],$$

porque:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} abr & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 2br & br & \dots & br & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & br & 2br & \dots & br & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & br & br & \dots & 2br & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 2r & r & \dots & r & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & r & 2r & \dots & r & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & r & r & \dots & 2r & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 2r & r & \dots & r \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & r & 2r & \dots & r \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & r & r & \dots & 2r \end{pmatrix}.$$

Vale señalar que, en este caso, $\oplus$ indica suma directa de matrices, esto es, sean $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_r$, matrices no necesariamente del mismo orden:

$$\mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \dots \oplus \mathbf{A}_r = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_r \end{pmatrix}.$$

El estimador mínimos cuadrados de β es dado por:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Nota 7.1. La matriz inversa de la matriz $\mathbf{I}_n + \mathbf{J}_n$, siendo $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top$, es dada por:

$$(\mathbf{I}_n + \mathbf{J}_n)^{-1} = \mathbf{I}_n - \frac{1}{1+n} \mathbf{J}_n.$$

Luego:

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{abr} \oplus \frac{1}{br} \left(\mathbf{I}_{a-1} - \frac{1}{a} \mathbf{J}_{a-1} \right) \oplus \left[\frac{1}{r} \mathbf{I}_a \otimes \left(\mathbf{I}_{b-1} - \frac{1}{b} \mathbf{J}_{b-1} \right) \right].$$

Ejemplo 7.3. Continuando con el ejemplo 7.2, presente el modelo usando las restricciones de la suma y encuentre el estimador de β .

Solución. Se tiene que el modelo es dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \delta_k + e_{ijk} \quad i = 1, 2, \quad j = 1, 2, 3, \quad k = 1, 2, 3,$$

donde:

- Y_{ijk} es la respuesta en la estación k del nivel j del factor tipo de máquina que está anidado en el nivel i del factor empresa.
- e_{ijk} es el componente asociado con el error aleatorio que se asume $e_{ijk} \sim N(0, \sigma^2)$.

Asumiendo las restricciones:

$$\delta_1 + \delta_2 + \delta_3 = 0, \quad \alpha_1 + \alpha_2 + \alpha_3 = 0 \text{ y } \beta_{1(i)} + \beta_{2(i)} = 0, \quad i = 1, 2, 3,$$

en que:

- α_i es el efecto medio de la empresa i con respecto a la media general.
- $\beta_{j(i)}$ es el efecto medio del tipo de máquina j de la empresa i con respecto a la media general.
- δ_k es el efecto medio de la estación k en que se mide el ozono. Este efecto se va a considerar como efecto de bloque.

Matricialmente el modelo sería dado por:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e},$$

Diseño anidado en dos etapas

con:

$$\begin{pmatrix} 18 \\ 30 \\ 53 \\ 25 \\ 45 \\ 45 \\ 15 \\ 42 \\ 104 \\ 20 \\ 35 \\ 50 \\ 21 \\ 49 \\ 58 \\ 63 \\ 23 \\ 147 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 1 \\ 1 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 & -1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \beta_{1(1)} \\ \beta_{2(1)} \\ \beta_{1(2)} \\ \beta_{2(2)} \\ \delta_1 \\ \delta_2 \end{pmatrix} + e.$$

Para calcular el vector de parámetros estimados, $\hat{\beta}$, es necesario evaluar $(\mathbf{X}^\top \mathbf{X})^{-1}$, en que:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= 18 \oplus 18 \oplus [\mathbf{I}_2 \otimes 3(\mathbf{I}_2 + \mathbf{J}_2)] \oplus 6(\mathbf{I}_2 + \mathbf{J}_2) \\ &= \begin{pmatrix} 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 12 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 12 \end{pmatrix}, \end{aligned}$$

luego:

$$\begin{aligned}
 (\mathbf{X}^\top \mathbf{X})^{-1} &= \frac{1}{18} \oplus \frac{1}{18} \oplus \left[\frac{1}{3} \mathbf{I}_2 \otimes \left(\mathbf{I}_2 - \frac{1}{1+2} \mathbf{J}_2 \right) \right] \oplus \frac{1}{6} \left(\mathbf{I}_2 - \frac{1}{1+2} \mathbf{J}_2 \right) \\
 &= \frac{1}{3} \begin{pmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1 - \frac{1}{3}) & -\frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{3} & (1 - \frac{1}{3}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1 - \frac{1}{3}) & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & (1 - \frac{1}{3}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} (1 - \frac{1}{3}) & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{3} & \frac{1}{2} (1 - \frac{1}{3}) \end{pmatrix},
 \end{aligned}$$

y por tanto:

$$\hat{\beta} = \begin{pmatrix} 46.8333 \\ -4.9444 \\ -8.2222 \\ -3.5556 \\ -16.7778 \\ -9.1111 \\ -19.8333 \\ -9.5 \end{pmatrix}.$$

✓

2. Análisis de varianza

Para el análisis de varianza se debe tener en cuenta que las fuentes de variación son variación debida al factor A y variación debida al factor B , que está anidado en el factor A . Esto implica separar la varianza debida al factor A y la varianza debida al factor B . Por tanto, se subdivide la matriz diseño en dos partes, en que una parte corresponde a las columnas asociadas con el factor A ($\mathbf{X}_\alpha$) y la otra parte corresponde a las columnas asociadas al factor B que está anidado en el factor A ($\mathbf{X}_{\beta(A)}$), es decir:

Diseño anidado en dos etapas

$$\mathbf{X}_\alpha = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \vdots & \ddots \\ 1 & \dots & 0 \\ \vdots & \vdots & \ddots \\ 1 & \dots & 0 \\ \vdots & \vdots & \ddots \\ 1 & \dots & 0 \\ \vdots & \vdots & \ddots \\ -1 & \dots & -1 \\ \vdots & \vdots & \ddots \\ -1 & \dots & -1 \\ \vdots & \vdots & \ddots \\ -1 & \dots & -1 \\ \vdots & \vdots & \ddots \\ -1 & \dots & -1 \end{pmatrix} \quad \text{y} \quad \mathbf{X}_{\beta(A)} = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ -1 & \dots & -1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ -1 & \dots & -1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & -1 & \dots & -1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & -1 & \dots & -1 \end{pmatrix}.$$

La tabla de análisis de varianza se presenta en la tabla 7.3.

Tabla 7.3. Análisis de varianza para un diseño anidado en 2 etapas,
balanceado

Causa de variación	Grad. de libertad	Suma de cuadrados	Cuadrado medio
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{SCFA}{a - 1}$
Factor B(A)	$a(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{\beta(A)} \mathbf{Y}$	$\frac{SCFB(A)}{a(b - 1)}$
Error	$ab(r - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$	$\frac{SCE}{ab(r - 1)}$
Total	$abr - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$	
$\bar{y} = \frac{1}{abr} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r y_{ijk} \quad \mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ $\mathbf{P}_\alpha = \mathbf{X}_\alpha (\mathbf{X}_\alpha^\top \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^\top \quad \mathbf{P}_{\beta(A)} = \mathbf{X}_{\beta(A)} (\mathbf{X}_{\beta(A)}^\top \mathbf{X}_{\beta(A)})^{-1} \mathbf{X}_{\beta(A)}^\top$			

Para la interpretación del P-valor se consideran tres casos, uno en que los dos factores son de efectos fijos, otro en que el factor no anidado es de efectos fijos y el factor anidado es de efectos aleatorios y el último cuando los dos factores son de efectos aleatorios.

2.1. Casos a considerar para la interpretación del análisis de varianza

Caso 1: Los α_i y los $\beta_{j(i)}$ corresponden a efectos fijos. Las hipótesis de interés son:

- $H_{0\alpha} : \alpha_i = 0 \quad \forall i$ versus $H_{a\alpha} : \text{Al menos un } \alpha_i \neq 0$ y
- $H_{0\beta(A)} : \beta_{j(i)} = 0 \quad \forall i, j$ versus $H_{a\beta(A)} : \text{Al menos un } \beta_{j(i)} \neq 0$.

$H_{0\alpha}$ se rechaza si el P-valor asociado al factor A es menor que un nivel de significancia dado α en la tabla de análisis de varianza, tabla 7.4, y se considerará que hay diferencia estadística significativa entre el efecto medio de los niveles del factor A sobre la respuesta. Igualmente para $H_{0\beta(A)}$. Dificultades en la interpretación y análisis de este tipo de diseño son discutidos en [62]. Para deducción de la tabla de análisis de varianza, ver anexo 1.

Tabla 7.4. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i y $\beta_{j(i)}$ asociados a efectos fijos

Causa de variación	Grad. de libertad	Suma de cuadrados	F-calculado	$\mathbb{E}(CM)$
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{CMFA}{CME}$	$\sigma_e^2 + \frac{br}{a-1} \sum_{i=1}^a \alpha_i^2$
Factor $B(A)$	$a(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{\beta(A)} \mathbf{Y}$	$\frac{CMFB(A)}{CME}$	$\sigma_e^2 + \frac{r}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2$
Error	$ab(r - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$		σ_e^2
Total	$abr - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$		

Ejemplo 7.4. Continuando con el ejemplo 7.2, realice el análisis de varianza y concluya.

Solución. Se tiene que:

$$\begin{aligned} \mathbf{y}^\top \mathbf{P}_X \mathbf{y} &= 51\,746, & \mathbf{y}^\top \mathbf{P}_\alpha \mathbf{y} &= 440.0556 & \mathbf{y}^\top \mathbf{P}_{\beta(A)} \mathbf{y} &= 3761.111 \\ \mathbf{y}^\top \mathbf{P}_\delta \mathbf{y} &= 8064.3333 & \mathbf{y}^\top \mathbf{y} &= 57\,951 & y & n\bar{y}^2 = 39\,480.5. \end{aligned}$$

La tabla de análisis de varianza respectiva se presenta en la tabla 7.5 y se concluye que no hay efectos estadísticamente significativos en la media de la medición de ozono ni por tipo de máquina ni por empresa.

Tabla 7.5. Análisis de varianza para los datos el ejemplo 7.2 asumiendo que los factores tipo de máquina y empresa son de efectos fijos

Causa de variación	Grad. de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Empresa	1	440.1	440.1	0.709	0.4194
Tipo de máquina (empresa)	4	3761.1	940.3	1.515	0.2701
Lugar	2	8064.3	4032.2		
Error	10	6205	620.5		
Total	17	18 470.5			

Las instrucciones en R para el ajuste del modelo presentado en la tabla 7.5 son:

```
remove(list=ls())
empresa=factor(gl(2,9,18))
contrasts(empresa)=contr.sum(levels(empresa))
tipomaquina=factor(gl(3,3,18))
contrasts(tipomaquina)=contr.sum(levels(tipomaquina))
estacion=factor(gl(3,1,18))
contrasts(estacion)=contr.sum(levels(estacion))
ozono=c(18,30,53,25,45,45,15,42,104,20,35,50,21,49,58,63,
        23,147)
ajuste=aov(ozono~empresa+empresa/tipomaquina+estacion)
ajuste$coef; summary(ajuste)
```

```
summary(aov(ozono~empresa+estacion))
summary(aov(ozono~estacion))
```



Caso 2: Los α_i corresponden a efectos fijos y los $\beta_{j(i)}$ corresponden a efectos aleatorios.

Las hipótesis de interés son:

- $H_{0\alpha} : \alpha_i = 0 \quad \forall i$ versus $H_{a\alpha} : \text{Al menos un } \alpha_i \neq 0$ y
- $H_{0\beta} : \sigma_{\beta(A)}^2 = 0$ versus $H_{a\beta} : \sigma_{\beta(A)}^2 > 0$.

$H_{0\beta}$ se rechaza si el P-valor asociado al factor $B(A)$ es menor que un nivel de significancia dado α en la tabla de análisis de varianza, tabla 7.6, y se considerará que la varianza debida al factor $B(A)$ es estadísticamente significativa. La interpretación respectiva para $H_{0\alpha}$ es similar al caso 1.

Tabla 7.6. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i asociado a efectos fijos y $\beta_{j(i)}$ asociado a efectos aleatorios

Causa de variación	Grad. de libertad	Suma de cuadrados	F-calculado	$E(CM)$
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{CMFA}{CMFB(A)}$	$\sigma_e^2 + r\sigma_{\beta(A)}^2 + \frac{br}{a-1} \sum_{i=1}^a \alpha_i^2$
Factor $B(A)$	$a(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{\beta(A)} \mathbf{Y}$	$\frac{CMFB(A)}{CME}$	$\sigma_e^2 + r\sigma_{\beta(A)}^2$
Error	$ab(r - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$		σ_e^2
Total	$abr - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$		

Fuente: [21, p. 680].

Ejemplo 7.5. Continuando con el ejemplo 7.2, si los tipos de máquina usados en el experimentos son, simplemente, una muestra aleatoria de máquinas que fabrican las dos empresas, presente el análisis de varianza.

Solución. El análisis de varianza correspondiente presentado en la tabla 7.7 indicaría que no hay efectos significativos en la media de la medición de ozono por empresa y que la varianza debida a las máquinas tampoco es significativa.



Tabla 7.7. Análisis de varianza para los datos el ejemplo 7.2 asumiendo que el factor máquina es un efecto aleatorio y el factor empresa es un efecto fijo

Causa de variación	Grad. de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Empresa	1	440.1	440.1	0.468	0.5315
Máquina (empresa)	4	3761.1	940.3	1.515	0.2701
Lugar	2	8064.3	4032.2		
Error	10	6205	620.5		
Total	17	18 470.5			

Caso 3: Los α_i y los $\beta_{j(i)}$ corresponden a efectos aleatorios.

Las hipótesis de interés son:

- $H_{0\alpha} : \sigma_{\alpha}^2 = 0$ versus $H_{a\alpha} : \sigma_{\alpha}^2 > 0$ y
- $H_{0\beta} : \sigma_{\beta(A)}^2 = 0$ versus $H_{a\beta} : \sigma_{\beta(A)}^2 > 0$.

$H_{0\alpha}$ se rechaza si el P-valor asociado al factor A es menor que un nivel de significancia dado α en la tabla de análisis de varianza, tabla 7.8, y se considerará que la varianza debida al factor A es estadísticamente significativa. La interpretación respectiva para $H_{0\beta}$ es similar al caso 2.

Ejemplo 7.6. Continuando con el ejemplo 7.2, si se asume que los tipos de máquina usados en el experimento son una muestra aleatoria de máquinas que fabrican las empresas, y que las empresas son seleccionadas de una población de empresas que fabrican máquinas para medir ozono, presentar el análisis de varianza.

Tabla 7.8. Análisis de varianza para un diseño en anidado en 2 etapas, balanceado con α_i y $\beta_{j(i)}$ asociados a efectos aleatorios

Causa de variación	Grad. de libertad	Suma de cuadrados	F-calculado	$\mathbb{E}(CM)$
Factor A	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y}$	$\frac{CMFA}{CMFB(A)}$	$\sigma_e^2 + r\sigma_\beta^2 + rb\sigma_\alpha^2$
Factor $B(A)$	$a(b - 1)$	$\mathbf{Y}^\top \mathbf{P}_{\beta(A)} \mathbf{Y}$	$\frac{CMFB(A)}{CME}$	$\sigma_e^2 + r\sigma_\beta^2$
Error	$ab(r - 1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$		σ_e^2
Total	$abr - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$		

Fuente: [41, p. 559].

Solución. En la tabla 7.7 se presenta el análisis de varianza e indicaría que la varianza debida a la máquinas y la varianza debida a las empresas no es significativa.

Una aplicación de diseños anidados considerando todos los factores aleatorios se puede consultar en [46].



Apéndice A

Inversa

generalizada

Se presentan dos tipos de inversa generalizada:

- Inversa de Moore-Penrose.
- Inversa generalizada.

1. Inversa de Moore-Penrose

Dada una matriz $A_{p \times q}$, hay una única matriz $M_{q \times p}$ tal que, ver [30, p. 497]:

1. $A_{p \times q} M_{q \times p} A_{p \times q} = A_{p \times q}$.
2. $M_{q \times p} A_{p \times q} M_{q \times p} = M_{q \times p}$.
3. AM es simétrica.
4. MA es simétrica.

Para encontrar a M se puede usar la factorización:

$$A_{p \times q} = K_{p \times r} L_{r \times q},$$

con K , una matriz de rango columna completo r y L , una matriz de rango fila completo r . M es dada por:

$$M_{q \times p} = L_{r \times q}^{\top} \left(K_{p \times r}^{\top} A_{p \times q} L_{r \times q}^{\top} \right)^{-1} K_{p \times r}^{\top}. \quad (\text{A.1})$$

Un procedimiento, para encontrar K y L , es usar la forma canónica equivalente:

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

y a partir de P y Q se obtienen K y L :

$$P^{-1} = \begin{pmatrix} K_{p \times r} & W_{p \times (p-r)} \end{pmatrix} \quad \text{y} \quad Q^{-1} = \begin{pmatrix} L_{r \times q} \\ W_{(q-r) \times q} \end{pmatrix}.$$

Ejemplo A.1. Encontrar la inversa de Moore-Penrose de la matriz:

$$A_{3 \times 4} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{pmatrix}.$$

Inversa generalizada

Solución. Se tiene $p = 3$ y $q = 4$. Para encontrar P se realizan operaciones entre las filas de A hasta obtener una matriz triangular superior. Estas operaciones también se realizan sobre una matriz idéntica de orden $p \times p$. La matriz resultante es P . Veamos:

$$\begin{array}{cccc|ccc}
 1 & 2 & 4 & 3 & 1 & 0 & 0 \\
 3 & -1 & 2 & -2 & 0 & 1 & 0 \\
 5 & -4 & 0 & -7 & 0 & 0 & 1
 \end{array} \Rightarrow \begin{array}{c} \text{fila } 1 \times -3 \\ \text{fila } 1 \times -5 \end{array} \left(\begin{array}{cccc|ccc}
 -3 & -6 & -12 & -9 & -3 & 0 & 0 \\
 -5 & -10 & -20 & -15 & -5 & 0 & 0
 \end{array} \right)$$

$$\begin{array}{cccc|ccc}
 1 & 2 & 4 & 3 & 1 & 0 & 0 \\
 0 & -7 & -10 & -11 & -3 & 1 & 0 \\
 0 & -14 & -20 & -22 & -5 & 0 & 1
 \end{array} \Rightarrow \begin{array}{c} \text{fila } 1 \times -3 + \text{fila } 2 \\ \text{fila } 1 \times -5 + \text{fila } 3 \end{array} \left(\begin{array}{cccc|ccc}
 0 & 14 & 20 & 22 & 6 & -2 & 0
 \end{array} \right)$$

$$\begin{array}{cccc|ccc}
 1 & 2 & 4 & 3 & 1 & 0 & 0 \\
 0 & -7 & -10 & -11 & -3 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1 & -2 & 1
 \end{array} \Rightarrow \begin{array}{c} \text{fila } 2 \times -2 + \text{fila } 3 \end{array} \underbrace{\left(\begin{array}{cccc|ccc}
 0 & -7 & -10 & -11 & -3 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1 & -2 & 1
 \end{array} \right)}_P$$

por tanto $r = 2$ y:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}, P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 2 & 1 \end{bmatrix}, K_{3 \times 2} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ 5 & 2 \end{bmatrix} \text{ y } W_{3 \times 1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Para encontrar Q se realizan operaciones entre las columnas de A hasta obtener una matriz triangular inferior. Estas operaciones también se realizan sobre una matriz idéntica de orden $q \times q$. La matriz resultante es Q . Veamos:

$$\begin{array}{cccc}
 1 & 2 & 4 & 3 \\
 3 & -1 & 2 & -2 \\
 5 & -4 & 0 & -7
 \end{array} \Rightarrow \text{columna } 1 \times -2 = \begin{pmatrix} -2 \\ -6 \\ -10 \\ -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\text{columna } 1 \times -4 = \begin{pmatrix} -4 \\ -12 \\ -20 \\ \hline -4 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{columna } 1 \times -3 = \begin{pmatrix} -3 \\ -9 \\ -15 \\ \hline -3 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

sumando, respectivamente, se tiene:

$$\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & -7 & -10 & -11 \\ 5 & -14 & -20 & -22 \\ \hline 1 & -2 & -4 & -3 \end{array} \Rightarrow \text{columna } 2 \times \frac{-10}{7} = \begin{pmatrix} 0 \\ 10 \\ 20 \\ \hline 20/7 \\ -10/7 \\ 0 \\ 0 \end{pmatrix},$$

$$\text{columna } 2 \times \frac{-11}{7} = \begin{pmatrix} 0 \\ 11 \\ 22 \\ \hline 22/7 \\ -11/7 \\ 0 \\ 0 \end{pmatrix},$$

sumando, respectivamente, se tiene que:

$$\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & -7 & 0 & 0 \\ 5 & -14 & 0 & 0 \\ \hline 1 & -2 & -8/7 & 1/7 \\ 0 & 1 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array},$$

Inversa generalizada

dividiendo la columna 2 por -7:

$$\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 5 & 2 & 0 & 0 \\ \hline 1 & 2/7 & -8/7 & 1/7 \\ 0 & -1/7 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}$$

$\underbrace{\hspace{10em}}_Q$

Luego:

$$Q = \begin{pmatrix} 1 & 2/7 & -8/7 & 1/7 \\ 0 & -1/7 & -10/7 & -11/7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad Q^{-1} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Por tanto:

$$L_{2 \times 4} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \end{pmatrix} \quad \text{y} \quad Z_{2 \times 4} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Usando (A.1):

$$\begin{aligned} M &= \begin{pmatrix} 1 & 0 \\ 2 & -7 \\ 4 & -10 \\ 3 & -11 \end{pmatrix} \left[\begin{pmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -7 \\ 4 & -10 \\ 3 & -11 \end{pmatrix} \right]^{-1} \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \end{bmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 2 & -7 \\ 4 & -10 \\ 3 & -11 \end{pmatrix} \begin{pmatrix} -81 & 465 \\ -45 & 219 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \end{pmatrix} \\ &= \frac{1}{3186} \begin{pmatrix} 1 & 0 \\ 2 & -7 \\ 4 & -10 \\ 3 & -11 \end{pmatrix} \begin{pmatrix} 219 & -465 \\ 45 & -81 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \end{pmatrix} = \frac{1}{3186} \begin{pmatrix} 219 & 192 & 165 \\ 123 & 6 & -111 \\ 426 & 228 & 30 \\ 162 & -18 & -198 \end{pmatrix}. \end{aligned}$$

✓

Ejercicio A.1. Verificar que $PAQ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ en el ejemplo A.1.

Ejercicio A.2. Verificar que M es la inversa de Moore-Penrose de A en el ejemplo A.1.

2. Inversa generalizada G

Dada una matriz $A_{p \times q}$, si $G_{q \times p}$ es una inversa generalizada de A , se tiene que G satisface el punto 1. de la inversa de Moore-Penrose, esto es, $AGA = A$. Un método para obtener G dada una matriz $A_{p \times q}$ de rango r es:

1. Obtener una submatriz A_{11} de rango r :

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

2. Calcular A_{11}^{-1} .
3. Sustituir en $A^\top$ los elementos de $A_{11}^\top$ por sus correspondientes A_{11}^{-1} y hacer todos los otros elementos iguales a cero.

Ejemplo A.2. Encontrar la inversa generalizada de la matriz A dada en el ejemplo A.1.

Solución. El rango de la matriz A dada en el ejemplo A.1 es $r = 2$. Los pasos para obtener G son:

1. Obtener una submatriz A_{11} de rango 2: $A_{11} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$.
2. Sustituir en $A^\top$ los elementos de $A_{11}^\top$ por sus correspondientes A_{11}^{-1} y hacer todos los otros elementos iguales a cero:

$$A_{11}^{-1} = \begin{pmatrix} 1/7 & 2/7 \\ 3/7 & -1/7 \end{pmatrix}$$

$$G = \begin{pmatrix} 1/7 & 2/7 & 0 \\ 3/7 & -1/7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ejercicio A.3. Verificar que G obtenido en el ejemplo A.2 es la inversa generalizada de A dada en el ejemplo A.1.

Nota A.1 (La matriz $X^\top X$). Si G es una inversa generalizada de $X^\top X$, $G = [X^\top X]^-$, se tienen las siguientes propiedades:

1. $G^\top$ también es una inversa generalizada de $X^\top X$.
2. $XGX^\top X = X$, esto es, $GX^\top$ es una inversa generalizada de X .
3. $XGX^\top$ es invariante a G .
4. $XGX^\top$ es simétrica.

Veamos: 1. Se sabe que $X^\top XGX^\top X = X^\top X$, luego:

$$\begin{aligned} X^\top X (X^\top X)^- X^\top X &= X^\top X \\ \left[X^\top X (X^\top X)^- X^\top X \right]^\top &= (X^\top X)^\top \\ X^\top X \left[(X^\top X)^- \right]^\top X^\top X &= X^\top X. \end{aligned}$$

Entonces $\left[(X^\top X)^- \right]^\top = G^\top$ es una inversa generalizada de $X^\top X$.

2. Teniendo en cuenta que [30, p. 52]:

- Para cualquier matriz $A_{m \times n}$, se puede probar que $A = 0$ si y solo si $A^\top A = 0$.
- Para cualesquieras matrices $A_{m \times n}$, $B_{n \times p}$ y $C_{n \times p}$, se puede probar que $AB = AC$ si y solo si $A^\top AB = A^\top AC$.

Entonces:

$$\begin{aligned} X^\top X (X^\top X)^- X^\top X &= X^\top X \\ \underbrace{X^\top}_{A^\top} \underbrace{X}_A \underbrace{(X^\top X)^-}_{B} X^\top X &= \underbrace{X^\top}_{A^\top} \underbrace{X}_A \underbrace{I}_C \\ X \underbrace{(X^\top X)^-}_{\substack{\text{Inversa} \\ \text{generalizada}}} X^\top X &= X, \end{aligned}$$

por tanto, $(X^\top X)^- X^\top$ es una inversa generalizada de X .

3. Usando el resultado de [30, p. 119], sean las matrices $A_{m \times n}$, $B_{p \times n}$ y $C_{m \times q}$. Si $\mathcal{R}(B) \subset \mathcal{R}(A)$ y $\mathcal{C}(C) \subset \mathcal{C}(A)$ ¹ se tiene que BA^-C es invariante a la escogencia de la inversa generalizada A^- , sea:

$$BA^-C = \underbrace{B}_B \underbrace{(X^\top X)^-}_{A^-} \underbrace{X^\top}_C,$$

luego se necesita encontrar L y R tal que:

$$B = LA$$

$$C = AR.$$

Sea:

$$X = LX^\top X$$

$$X^\top = X^\top XR,$$

de la propiedad 2. se tiene que:

$$\begin{aligned} X(X^\top X)^-X^\top X &= X \Rightarrow L = X(X^\top X)^- \\ X^\top X \left[(X^\top X)^- \right]^\top X^\top &= X^\top \Rightarrow R = \left[X(X^\top X)^- \right]^\top. \end{aligned}$$

Así $X(X^\top X)^-X^\top$ es invariante a $(X^\top X)^-$.

4. Usando que para cualquier matriz A , $(A^-)^\top$ es una inversa generalizada de $A^\top$ [30, p. 117]. En este caso, si $A^- = (X^\top X)^-$ es la inversa generalizada de $A = X^\top X$, $(A^-)^\top = \left[(X^\top X)^- \right]^\top$ es una inversa generalizada de $(X^\top X)^\top = X^\top X$, y entonces:

$$\left[X(X^\top X)^-X^\top \right]^\top = X \left[(X^\top X)^- \right]^\top X^\top,$$

¹El espacio columna de $A_{m \times n}$, denotado por $\mathcal{C}(A)$, es el conjunto de todos los vectores columna m -dimensionales tales que se pueden expresar como combinaciones lineales de las n columnas de A :

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{A} \mathbf{x},$$

con x_i escalar y $\mathbf{a}_i$ la i -ésima columna de A [30, p. 27].

la cual por la propiedad 3. es invariante a G , entonces:

$$\mathbf{X} \left[\left(\mathbf{X}^\top \mathbf{X} \right)^- \right]^\top \mathbf{X}^\top = \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^- \mathbf{X}^\top.$$

✓

Nota A.2. Sea $\mathbf{A}^-$ una inversa generalizada de $\mathbf{A}$. Entonces:

i) La ecuación $\mathbf{A}\mathbf{x} = \mathbf{d}$ es consistente si $\mathbf{A}\mathbf{A}^-\mathbf{d} = \mathbf{d}$ [30, p. 121].

ii) Cualquier solución es de la forma, ver [30, p. 141]:

$$\mathbf{x} = \mathbf{A}^-\mathbf{d} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}, \quad (\text{A.2})$$

para algún vector columna $\mathbf{z}$.

Retomando la ecuación (2.4), si $\mathbf{X}^\top \mathbf{X}$ es singular, $\mathbf{A} = \mathbf{X}^\top \mathbf{X}$, $\mathbf{x} = \hat{\beta}$ y $\mathbf{d} = \mathbf{X}^\top \mathbf{y}$ en (A.2):

$$\underbrace{\mathbf{X}^\top \mathbf{X}}_{\mathbf{A}} \underbrace{\hat{\beta}}_{\mathbf{x}} = \underbrace{\mathbf{X}^\top \mathbf{y}}_{\mathbf{d}},$$

se puede determinar si es una ecuación consistente y encontrar una solución.

Veamos: i)

$$\mathbf{X}^\top \underbrace{\mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^-}_{\text{inversa gener. de } \mathbf{X}^\top} \mathbf{X}^\top \mathbf{Y} = \mathbf{X}^\top \mathbf{Y} = \mathbf{d},$$

luego $\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}$ es una ecuación consistente.

ii) Se tiene una solución para β y es dada por:

$$\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X} \right)^- \mathbf{X}^\top \mathbf{Y} + \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X} \right)^- \mathbf{X}^\top \mathbf{X} \right] \mathbf{z}, \quad (\text{A.3})$$

para algún $\mathbf{z}$ y, por tanto, se tendrán infinitas soluciones. En particular:

$$\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X} \right)^- \mathbf{X}^\top \mathbf{y}.$$

✓



Apéndice B

Deducción del

modelo DCA

1. Variables aleatorias involucradas en el diseño completamente aleatorizado

El proceso de aleatorización de las unidades experimentales a los tratamientos o viceversa puede ser expresado matemáticamente por [33, p.155]:

$$\delta_{ij}^k = \begin{cases} 1 & \text{Si la unidad experimental } k \text{ está asociada con el nivel } ij \\ 0 & \text{en otro caso,} \end{cases}$$

con $k = 1, \dots, n$; $i = 1, \dots, a$; $j = 1, \dots, r$. Por tanto:

$$\begin{aligned} P\left(\delta_{ij}^k = 1\right) &= \frac{1}{n} \\ \mathbb{E}\left(\delta_{ij}^k\right) &= p\left(\delta_{ij}^k = 1\right) = \frac{1}{n} \\ \mathbb{E}\left[\left(\delta_{ij}^k\right)^2\right] &= 1^2 p\left(\delta_{ij}^k = 1\right) = \frac{1}{n} \\ \mathbb{V}\left(\delta_{ij}^k\right) &= \frac{1}{n} - \frac{1}{n^2} = \frac{1}{n} \left(1 - \frac{1}{n}\right) \\ p\left(\delta_{ij}^k = 1, \delta_{i'j'}^{k'} = 1\right) &= \begin{cases} \frac{1}{n} \frac{1}{n-1} & \text{Si } k \neq k', ij \neq i'j' \\ 0 & \text{en otro caso} \end{cases} \\ \mathbb{E}\left(\delta_{ij}^k \delta_{i'j'}^{k'}\right) &= \begin{cases} \frac{1}{n} & \text{Si } k = k', ij = i'j' \\ \frac{1}{n} \frac{1}{n-1} & \text{Si } k \neq k', ij \neq i'j' \\ 0 & \text{en otro caso} \end{cases} \\ \mathbb{Cov}\left(\delta_{ij}^k, \delta_{i'j'}^{k'}\right) &= \mathbb{E}\left(\delta_{ij}^k \delta_{i'j'}^{k'}\right) - \mathbb{E}\left(\delta_{ij}^k\right) \mathbb{E}\left(\delta_{i'j'}^{k'}\right) \\ &= \begin{cases} \frac{1}{n} - \frac{1}{n^2} & \text{Si } k = k', ij = i'j' \\ \frac{1}{n} \frac{1}{n-1} - \frac{1}{n^2} & \text{Si } k \neq k', ij \neq i'j' \\ -\frac{1}{n^2} & \text{Si } k = k', ij \neq i'j' \\ -\frac{1}{n^2} & \text{Si } k \neq k', ij = i'j' \end{cases} \end{aligned}$$

Se resalta que los δ_{ij}^k son variables aleatorias Bernoulli e idénticamente distribuidas, pero NO son independientes.

1.1. Respuestas y observaciones conceptuales

Suponga que el tratamiento i es aplicado a la unidad experimental k . La verdadera (o conceptual) respuesta es un número, el cual se denotará por T_{ik} . Sin embargo, no se pueden observar todos los T_{ik} porque el tratamiento i es aplicado solo a algunos k . Para asociar lo que se observa y los T_{ik} , se usa δ_{ij}^k :

$$Y_{ij} = \sum_{k=1}^n \delta_{ij}^k T_{ik} \quad (\text{B.1})$$

donde δ_{ij}^k es aleatorio y T_{ik} es fijo. Así, si la unidad k está asociada con el nivel ij (tratamiento i , réplica j), se observa T_{ik} .

Ejemplo B.1. Suponga $a = 2$, $r = 2$ y $N = 4$, entonces los δ_{ij}^k se pueden colocar en un arreglo como:

$$\begin{array}{cccc} \delta_{11}^1 & \delta_{12}^1 & \delta_{21}^1 & \delta_{22}^1 \\ \delta_{11}^2 & \delta_{12}^2 & \delta_{21}^2 & \delta_{22}^2 \\ \delta_{11}^3 & \delta_{12}^3 & \delta_{21}^3 & \delta_{22}^3 \\ \delta_{11}^4 & \delta_{12}^4 & \delta_{21}^4 & \delta_{22}^4 \end{array}$$

Luego $\delta_{12}^3 = 1$ significa que la unidad experimental 3 recibe el nivel 12 (tratamiento 1, réplica 2). Esto significa que $\delta_{11}^3 = \delta_{21}^3 = \delta_{22}^3 = 0$ y $\delta_{12}^1 = \delta_{12}^2 = \delta_{12}^4 = 0$ y además $y_{12} = T_{13}$. $\square$

Nota B.1. En general $\sum_{k=1}^n \delta_{ij}^k = 1$ y $\sum_{i=1}^a \sum_{j=1}^r \delta_{ij}^k = 1$.

Por otro lado, T_{ik} se puede escribir como:

$$T_{ik} = T_{ik} + \bar{T}_{i*} - \bar{T}_{i*} + \bar{T}_{*k} - \bar{T}_{*k} + \bar{T}_{**} - \bar{T}_{**} + \bar{T}_{**} - \bar{T}_{**}, \quad (\text{B.2})$$

en que:

$$\bar{T}_{i*} = \frac{1}{n} \sum_{k=1}^n T_{ik}, \quad \bar{T}_{*k} = \frac{1}{a} \sum_{i=1}^a T_{ik} \quad \text{y} \quad \bar{T}_{**} = \frac{1}{na} \sum_{k=1}^n \sum_{i=1}^a T_{ik}.$$

Reordenando T_{ik} queda:

$$T_{ik} = \bar{T}_{**} + (\bar{T}_{i*} - \bar{T}_{**}) + (\bar{T}_{*k} - \bar{T}_{**}) + (T_{ik} - \bar{T}_{i*} - \bar{T}_{*k} + \bar{T}_{**}). \quad (\text{B.3})$$

Además, si se asume que:

$$T_{ik} = T_i + U_k,$$

esto es, que T_{ik} es el resultado de la respuesta debida al tratamiento más la respuesta debida a la unidad experimental, se tiene que:

$$\begin{aligned}\bar{T}_{i*} &= \frac{1}{n} \sum_{k=1}^n T_{ik} = \frac{1}{n} \sum_{k=1}^n (T_i + U_k) = T_i + \bar{U} \\ \bar{T}_{*k} &= \frac{1}{a} \sum_{i=1}^a T_{ik} = \frac{1}{a} \sum_{i=1}^a (T_i + U_k) = \bar{T} + U_k \\ \bar{T}_{**} &= \frac{1}{na} \sum_{k=1}^n \sum_{i=1}^a T_{ik} = \frac{1}{na} \sum_{k=1}^n \sum_{i=1}^a (T_i + U_k) \\ &= \frac{1}{na} \left(\sum_{k=1}^n \sum_{i=1}^a T_i + \sum_{k=1}^n \sum_{i=1}^a U_k \right) = \frac{1}{na} \left(n \sum_{i=1}^a T_i + a \sum_{k=1}^n U_k \right) \\ &= \bar{T} + \bar{U}.\end{aligned}$$

Notando que el último término de (B.3) es cero, esto es:

$$T_{ik} - \bar{T}_{i*} - \bar{T}_{*k} + \bar{T}_{**} = \cancel{T_i} + \cancel{U_k} - \cancel{T_i} - \cancel{U_k} - \cancel{\bar{T}} - \cancel{\bar{U}} + \cancel{\bar{T}} + \cancel{\bar{U}} = 0,$$

T_{ik} se puede escribir como:

$$\begin{aligned}T_{ik} &= \underbrace{\bar{T}_{**}}_{\bar{T} + \bar{U}} + \underbrace{(\bar{T}_{i*} - \bar{T}_{**})}_{T_i + \bar{U} - \bar{T} - \bar{U}} + \underbrace{(\bar{T}_{*k} - \bar{T}_{**})}_{\bar{T} + U_k - \bar{T} - \bar{U}} + \underbrace{(T_{ik} - \bar{T}_{i*} - \bar{T}_{*k} + \bar{T}_{**})}_0 \\ &= \underbrace{(\bar{T} + \bar{U})}_{\mu} + \underbrace{(T_i - \bar{T})}_{\tau_i} + \underbrace{(U_k - \bar{U})}_{u_k} \\ T_{ik} &= \mu + \tau_i + u_k.\end{aligned}\tag{B.4}$$

Note que de (B.4) sigue que:

$$\begin{aligned}\sum_{i=1}^a \tau_i &= 0 \\ \sum_{k=1}^n u_k &= 0.\end{aligned}\tag{B.5}$$

Sustituyendo (B.4) en (B.1) se tiene que:

$$\begin{aligned}
 Y_{ij} &= \sum_{k=1}^n \delta_{ij}^k T_{ik} = \sum_{k=1}^n \delta_{ij}^k (\mu + \tau_i + u_k) = \underbrace{\mu \sum_{k=1}^n \delta_{ij}^k}_1 + \tau_i \underbrace{\sum_{k=1}^n \delta_{ij}^k}_1 + \sum_{k=1}^n u_k \delta_{ij}^k \\
 &= \mu + \tau_i + \underbrace{\sum_{k=1}^n u_k \delta_{ij}^k}_{\omega_{ij}} \\
 Y_{ij} &= \mu + \tau_i + \omega_{ij}, \tag{B.6}
 \end{aligned}$$

en que (B.6) es el modelo lineal derivado asociado al **diseño completamente aleatorizado**. Además, se puede mostrar que:

$$\begin{aligned}
 \mathbb{E}(\omega_{ij}) &= 0 \\
 \mathbb{V}(\omega_{ij}) &= \left(1 - \frac{1}{n}\right) \sigma_u^2 \quad \text{con} \quad \sigma_u^2 = \frac{1}{n-1} \sum_{k=1}^n u_k^2 \\
 \mathbb{Cov}(\omega_{ij}, \omega_{i'j'}) &= -\frac{1}{n} \sigma_u^2 \\
 \mathbb{E}(Y_{ij}) &= \mu + \tau_i \\
 \mathbb{V}(Y_{ij}) &= \left(1 - \frac{1}{n}\right) \sigma_u^2 \\
 \mathbb{Cov}(Y_{ij}, Y_{i'j'}) &= -\frac{1}{n} \sigma_u^2,
 \end{aligned}$$

luego las réplicas dentro de los tratamientos son correlacionadas. Note que σ_u^2 se puede interpretar como una medida de la variabilidad entre las unidades experimentales (UE) [33, p. 160]. De lo anterior se tiene que:

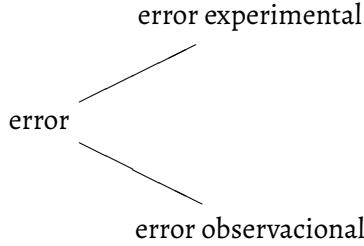
$$\mathbb{E}(\bar{Y}_{i*}) = \mathbb{E}\left(\frac{1}{r} \sum_{j=1}^r Y_{ij}\right) = \frac{1}{r} \sum_{j=1}^r (\mu + \tau_i) = \mu + \tau_i$$

y

$$\begin{aligned}\mathbb{V}(\bar{Y}_{i*}) &= \mathbb{V}\left(\frac{1}{r} \sum_{j=1}^r Y_{ij}\right) = \frac{1}{r^2} \left(\sum_{j=1}^r \mathbb{V}(Y_{ij}) + \sum_{j \neq j'=1}^r \mathbb{Cov}(Y_{ij}, Y_{i'j'}) \right) \\ &= \frac{1}{r^2} \left[r \left(1 - \frac{1}{n}\right) \sigma_u^2 - \frac{r(r-1)}{n} \sigma_u^2 \right] = \frac{1}{r} \left[\frac{n-1}{n} - \frac{r-1}{n} \right] \sigma_u^2 \\ &= \frac{1}{r} \left(\frac{n-r}{n} \right) \sigma_u^2 = \frac{1}{r} \left(1 - \frac{r}{n} \right) \sigma_u^2.\end{aligned}$$

1.2. Considerando otros errores en el modelo en un sentido amplio

Se va a considerar un tipo de error adicional al error experimental ya antes considerado, que es el **error observacional**.



Luego la respuesta ya no es T_{ik} , sino T_{ik} más el error observacional:

$$Y_{ik} = T_{ik} + M_{ik} = T_i + U_k + M_{ik}, \quad (\text{B.7})$$

en que Y_{ik} denota la observación de la k -ésima unidad experimental a la cual el i -ésimo tratamiento ha sido aplicado. La componente M_{ik} expresa lo que se llama error observacional. Como parte del error experimental [33, p. 161], están:

- i) El error de aplicación del tratamiento, debido a la falta de habilidad para replicar un tratamiento.
- ii) El error de estado, debido a los cambios aleatorios en el estado físico de una unidad experimental.

- iii) El error de selección, debido a la selección aleatoria de las unidades experimentales para el experimento.

Como parte del error observacional, están:

- i) El error de medición, debido a la imprecisión en la medición o procedimiento de medida E_{ik} (aleatorio).
 ii) El error de muestreo, debido a la selección aleatoria de las unidades observacionales O_{ik} (aleatorio).

Así, M_{ik} se puede escribir como:

$$M_{ik} = E_{ik} + O_{ik}. \quad (\text{B.8})$$

Reemplazando (B.8) en (B.7), pero incluyendo adicionalmente en U_k el error de aplicación y el error de estado:

$$\begin{aligned} Y_{ik} &= T_i + U_k + E_{ik} + O_{ik} \\ &= \mu + \tau_i + u_k + E_{ik} + O_{ik}. \end{aligned}$$

Realizando el mismo procedimiento que en (B.1), se tiene que:

$$\begin{aligned} Y_{ij} &= \sum_{k=1}^n \delta_{ij}^k Y_{ik} = \mu + \tau_i + \underbrace{\sum_{k=1}^n \delta_{ij}^k u_k}_{\omega_{ij}} + \underbrace{\sum_{k=1}^n \delta_{ij}^k E_{ik}}_{\nu_{ij}} + \underbrace{\sum_{k=1}^n \delta_{ij}^k O_{ik}}_{\eta_{ij}} \\ Y_{ij} &= \mu + \tau_i + \omega_{ij} + \nu_{ij} + \eta_{ij}. \end{aligned} \quad (\text{B.9})$$

Uniando el error experimental, ω_{ij} , con el error de medición, ν_{ij} , (B.9) queda:

$$Y_{ij} = \mu + \tau_i + \underbrace{\omega_{ij} + \nu_{ij}}_{\varepsilon_{ij}} + \eta_{ij}$$

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} + \eta_{ij}.$$

Asumiendo que δ_{ij}^k , E_{ik} y O_{ik} son estadísticamente independientes, que:

$$\begin{aligned} \mathbb{E}(\nu_{ij}) &= 0 & \mathbb{E}(\eta_{ij}) &= 0 \\ \mathbb{V}(\nu_{ij}) &= \sigma_\nu^2 & \mathbb{V}(\eta_{ij}) &= \sigma_\eta^2, \end{aligned}$$

que los ε_{ij} son independientes e idénticamente distribuidos [33, p. 163], y denotando:

$$\mathbb{V}(\varepsilon_{ij}) = \mathbb{V}(\omega_{ij} + \nu_{ij}) = \sigma_\varepsilon^2,$$

se tiene que:

$$\begin{aligned}\mathbb{E}(Y_{ij}) &= \mathbb{E}(\mu + \tau_i + \omega_{ij} + \nu_{ij} + \eta_{ij}) = \mu + \tau_i \\ \mathbb{V}(Y_{ij}) &= \mathbb{V}(\mu + \tau_i + \omega_{ij} + \nu_{ij} + \eta_{ij}) \\ &= \mathbb{V}(\omega_{ij} + \nu_{ij} + \eta_{ij}) \\ &= \sigma_\varepsilon^2 + \sigma_\eta^2.\end{aligned}$$

Nota B.2. A σ_ε^2 se le conoce como componente de varianza del error experimental y a σ_η^2 se le conoce como el componente de varianza del error de muestreo (u observacional).

Nota B.3. Como ε_{ij} y η_{ij} tienen los mismos subíndices, se pueden condensar en:

$$e_{ij} = \varepsilon_{ij} + \eta_{ij},$$

con $\mathbb{E}(e_{ij}) = 0$ y $\mathbb{V}(e_{ij}) = \sigma_e^2 = \sigma_\varepsilon^2 + \sigma_\eta^2$. Reemplazando en (B.9) el modelo final es dado por:

$$Y_{ij} = \mu + \tau_i + e_{ij}. \quad (\text{B.10})$$

2. Análisis de varianza - DCA

Se parte del modelo dado en (B.10), que matricialmente se puede escribir como:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{a,r-1} \\ Y_{ar} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_a \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{a,r-1} \\ e_{ar} \end{bmatrix},$$

y que $\mathbf{Y}^\top \mathbf{Y}$ es igual a:

$$\mathbf{Y}^\top \mathbf{Y} = \mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} + \mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y},$$

con:

$$\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top,$$

cuya tabla de análisis de varianza simple se presentó en la tabla 2.2.

Al restar de la suma de cuadrados total, $\mathbf{Y}^\top \mathbf{Y}$, la cantidad $n\bar{Y}^2$, siendo:

$$\bar{Y} = \frac{1}{ar} \sum_{i=1}^a \sum_{j=1}^r Y_{ij},$$

se puede construir una tabla de análisis de varianza corregida (por la media), ver tabla B.1.

Tabla B.1. Análisis de varianza corregida por la media

Causa de variación	Grados de libertad	Suma de cuadrados
$\mathbf{X}$	$\text{rango}(\mathbf{X}) - 1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - \frac{1}{n} \mathbf{Y}^\top \mathbf{1} \mathbf{1}^\top \mathbf{Y}$
Residual	$n - \text{rango}(\mathbf{X})$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - \frac{1}{n} \mathbf{Y}^\top \mathbf{1} \mathbf{1}^\top \mathbf{Y}$

2.1. Distribución y esperanza de las sumas de cuadrados

2.1.1. Suma de cuadrados del modelo (SCM)

$$\begin{aligned} SCM &= \mathbf{Y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - \frac{1}{n} \mathbf{Y}^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{Y} \\ &= \mathbf{Y}^\top \underbrace{\left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)}_{\mathbf{A}} \mathbf{Y} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}. \end{aligned}$$

Para encontrar la distribución de la suma de cuadrados del modelo, SCM, note que:

- La matriz $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ es simétrica.

- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma_e^2 \mathbf{I}_n)$ se tiene $\mathbf{V} = \sigma_e^2 \mathbf{I}_n$. Sea:

$$\mathbf{Y}^* = \frac{1}{\sigma_e} \mathbf{Y} \sim N\left(\frac{1}{\sigma_e} \boldsymbol{\mu}, \mathbf{I}_n\right), \quad (\text{B.11})$$

entonces $\mathbf{V}^* = \mathbf{I}$ y $\boldsymbol{\mu}^* = \frac{1}{\sigma_e} \boldsymbol{\mu}$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{A}\mathbf{V}^*$:

$$\begin{aligned} \mathbf{A}\mathbf{V}^* \mathbf{A}\mathbf{V}^* &= \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \\ &= \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{X}} - \frac{1}{n} \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_n \mathbf{1}_n^\top}_{\mathbf{1}_n} \\ &\quad - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{1}_n^\top} + \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \\ &= \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{1}} - \cancel{\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top} - \cancel{\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top} + \cancel{\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top = \mathbf{A}\mathbf{V}^*. \end{aligned}$$

Por tanto, $\mathbf{A}\mathbf{V}^*$ es idempotente y entonces:

$$\mathbf{Y}^{*\top} \mathbf{A}\mathbf{Y}^* = \frac{\mathbf{Y}^\top \mathbf{A}\mathbf{Y}}{\sigma_e^2} = \frac{SCM}{\sigma_e^2} \sim \chi^2(\text{rango}(\mathbf{A}), \lambda),$$

con $\lambda = \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^*$, luego:

$$\frac{SCM}{\sigma_e^2} \sim \chi^2\left(\text{rango}(\mathbf{A}), \frac{1}{2} \frac{1}{\sigma_e} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \frac{1}{\sigma_e}\right).$$

¹ $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} = \mathbf{X}$ por la nota A.1 propiedad 2.
 $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{1}_n = \mathbf{1}_n$ porque la columna $\mathbf{1}_n$ es parte de las columnas de $\mathbf{X}$ y $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ es la matriz de proyección de las columnas de $\mathbf{X}$.

Siendo que:

$$\begin{aligned}
 \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} &= \boldsymbol{\beta}^\top \mathbf{X}^\top \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{X} \boldsymbol{\beta} \\
 &= \boldsymbol{\beta}^\top \mathbf{X}^\top \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{X}} \mathbf{X} \boldsymbol{\beta} - \frac{1}{n} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{X} \boldsymbol{\beta} \\
 &= \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \frac{1}{n} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{X} \boldsymbol{\beta} \quad y \\
 rango(\mathbf{A}) &= rango \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \\
 &\stackrel{\text{idempotencia}}{=} traza \left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) = traza(\mathbf{P}_x) - traza \left(\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \\
 &= rango(\mathbf{P}_x) - traza \left(\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \\
 &= rango(\mathbf{X}) - 1 = a - 1,
 \end{aligned}$$

se tiene que:

$$\frac{SCM}{\sigma_e^2} \sim \chi^2 \left(a - 1, \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{X}^\top \left[\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \mathbf{X} \boldsymbol{\beta} \right),$$

esto es, $\lambda = \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{X}^\top \left[\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \mathbf{X} \boldsymbol{\beta}$.

2.1.2. Suma de cuadrados del error (SCE)

$$SCE = \mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}.$$

Como $\mathbf{I} - \mathbf{P}_x$ es simétrica e idempotente, $(\mathbf{I} - \mathbf{P}_x) \mathbf{V}^*$ es idempotente, en que $\mathbf{V}^*$ es la matriz de varianza-covarianza de $\mathbf{Y}^*$, dada en (B.11). Por tanto:

$$\begin{aligned}
 \mathbf{Y}^{*\top} (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}^* &= \frac{\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}}{\sigma_e^2} \\
 &= \frac{SCE}{\sigma_e^2} \sim \chi^2 \left(rango(\mathbf{I} - \mathbf{P}_x), \frac{1}{2} \boldsymbol{\mu}^{*\top} (\mathbf{I} - \mathbf{P}_x) \boldsymbol{\mu}^* \right) \\
 \frac{SCE}{\sigma_e^2} &\sim \chi^2 \left(rango(\mathbf{I} - \mathbf{P}_x), \frac{1}{2} \frac{1}{\sigma_e} \boldsymbol{\mu}^\top (\mathbf{I} - \mathbf{P}_x) \boldsymbol{\mu} \frac{1}{\sigma_e} \right).
 \end{aligned}$$

Siendo que:

$$\begin{aligned}
 \boldsymbol{\mu}^\top (\mathbf{I} - \mathbf{P}_x) \boldsymbol{\mu} &= \boldsymbol{\beta}^\top \mathbf{X}^\top \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \mathbf{X} \boldsymbol{\beta} \\
 &= \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \underbrace{\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}}_{\mathbf{X}} \\
 &= \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} = 0 \quad y \\
 rango(\mathbf{I} - \mathbf{P}_x) &= ar - rango(\mathbf{P}_x) = ar - rango(\mathbf{X}) \\
 &= ar - a = a(r - 1),
 \end{aligned}$$

se tiene que:

$$\frac{SCE}{\sigma_e^2} \sim \chi_{a(r-1)}^2. \quad (\text{B.12})$$

2.1.3. Esperanza de las sumas de cuadrados

$$\begin{aligned}
 \mathbb{E}(SCM) &= \mathbb{E} \left(\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{Y}^2 \right) \\
 &= \mathbb{E} \left[\mathbf{Y}^\top \underbrace{\left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)}_{\mathbf{A}} \mathbf{Y} \right] = \mathbb{E} \left[\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \right] \\
 &\stackrel{\text{ver (2.10)}}{=} traza(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu},
 \end{aligned}$$

con $\mathbf{V} = \sigma_e^2 \mathbf{I}_n$ y $\boldsymbol{\mu} = \mathbf{X} \boldsymbol{\beta}$.

Reemplazando:

$$\begin{aligned}
 \mathbb{E}(SCM) &= traza \left(\left[\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \sigma_e^2 \mathbf{I}_n \right) + \\
 &\quad \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \frac{1}{n} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{X} \boldsymbol{\beta} \\
 &= (a - 1) \sigma_e^2 + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \frac{1}{n} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{X} \boldsymbol{\beta} \\
 &= (a - 1) \sigma_e^2 + \sum_{i=1}^a (\mu + \tau_i)^2 r - \frac{1}{n} \left(\sum_{i=1}^a (\mu + \tau_i) r \right)^2
 \end{aligned}$$

Deducción del modelo DCA

$$\begin{aligned}
 &= (a-1)\sigma_e^2 + \cancel{r\mu^2} + r \sum_{i=1}^a \tau_i^2 - \cancel{\frac{1}{ar}a^2r^2\mu^2} \\
 &= (a-1)\sigma_e^2 + r \sum_{i=1}^a \tau_i^2, \\
 \mathbb{E}(SCE) &= \mathbb{E} \left(\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y} \right) \\
 &= \text{traza} \{ (\mathbf{I} - \mathbf{P}_x) \sigma_e^2 \mathbf{I} \} + \boldsymbol{\mu}^\top (\mathbf{I} - \mathbf{P}_x) \boldsymbol{\mu} \\
 &= \sigma_e^2 a(r-1) + \boldsymbol{\beta}^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{X} \boldsymbol{\beta} \\
 &= \sigma_e^2 a(r-1) + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \underbrace{\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}}_{\mathbf{X}} \\
 &= \sigma_e^2 a(r-1) + \cancel{\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}} - \cancel{\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}} = \sigma_e^2 a(r-1).
 \end{aligned}$$

2.2. Cuadrados medios (CM)

Los **cuadrados medios** (CM), se obtienen al dividir la suma de cuadrados entre los grados de libertad, ver tabla B.2.

Tabla B.2. Análisis de varianza corregida por la media

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio
Modelo	$\text{rango}(\mathbf{X}) - 1 = a - 1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{Y}^2$	$\frac{SCM}{glm}$
Error	$n - \text{rango}(\mathbf{X}) = a(r-1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$	$\frac{SCE}{gle}$
Total	$n - 1 = ar - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{Y}^2$	

La esperanza de los cuadrados medios del modelo y del error es:

$$\begin{aligned}
 \mathbb{E}(CMM) &= \mathbb{E} \left(\frac{SCM}{glm} \right) = \mathbb{E} \left(\frac{SCM}{a-1} \right) = \frac{1}{a-1} \mathbb{E}(SCM) \\
 &= \frac{1}{a-1} \left((a-1)\sigma_e^2 + r \sum_{i=1}^a \tau_i^2 \right) = \sigma_e^2 + \frac{r}{a-1} \sum_{i=1}^a \tau_i^2
 \end{aligned}$$

$$\begin{aligned}\mathbb{E}(CME) &= \mathbb{E}\left(\frac{SCE}{gle}\right) = \mathbb{E}\left(\frac{SCE}{a(r-1)}\right) = \frac{1}{a(r-1)}\mathbb{E}(SCE) \\ &= \frac{\sigma_e^2 \cancel{a(r-1)}}{\cancel{a(r-1)}} = \sigma_e^2.\end{aligned}\quad (\text{B.13})$$

2.2.1. Cociente de cuadrados medios

Para obtener la distribución del cociente del cuadrado medio del modelo entre el cuadrado medio del error, primero se verifica que las dos formas cuadráticas asociadas a los cuadrados medios sean independientes, para esto se usa el teorema 2.3.

$$\begin{aligned}A_1 V A_2 &= \underbrace{\left(X(X^\top X)^{-1}X^\top - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)}_{A_1} \underbrace{I_n}_{V^*} \underbrace{\left(I_n - X(X^\top X)^{-1}X^\top\right)}_{A_2} \\ &= \cancel{P_x} - \underbrace{P_x P_x}_{\cancel{P_x}} - \cancel{\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top} + \underbrace{\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top P_x}_{\cancel{\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top}} = 0.\end{aligned}$$

Por tanto, las formas cuadráticas:

$$Q_1 = Y^{*\top} \left(X(X^\top X)^{-1}X^\top - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \right) Y^*$$

y

$$Q_2 = Y^{*\top} \left(I_n - X(X^\top X)^{-1}X^\top \right) Y^*,$$

son independientes. Entonces el cociente, ver definición 2.6:

$$\frac{\frac{SCM}{\cancel{\sigma_e^2}}}{\frac{\cancel{a-1}}{\frac{SCE}{\cancel{\sigma_e^2}}}} = \frac{\frac{SCM}{a-1}}{\frac{SCE}{a(r-1)}} = \frac{CMM}{CME} \sim F(a-1, a(r-1), \lambda),$$

en que:

$$\lambda = \frac{1}{2\sigma_e^2} \beta^\top X^\top \left(I - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \right) X \beta = \frac{1}{2\sigma_e^2} \left(r \sum_{i=1}^a \tau_i^2 \right). \quad (\text{B.14})$$

La tabla B.3 presenta de forma reducida los resultados obtenidos de las esperanzas de los cuadrados medios y su distribución.

Tabla B.3. Análisis de varianza incluyendo esperanza de los cuadrados medios

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	$\mathbb{E}(CM)$	
Modelo	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{Y}^2$	$\frac{SCM}{a-1}$	$\sigma_e^2 + \frac{r}{a-1} \sum \tau_i^2$	$\frac{SCM}{\sigma_e^2} \sim \chi^2(a-1, \lambda)$
Error	$a(r-1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$	$\frac{SCE}{a(r-1)}$	σ_e^2	$\frac{SCE}{\sigma_e^2} \sim \chi^2_{a(r-1)}$
Total	$ar - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{Y}^2$			
con $\lambda = \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{X}^\top (\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top) \mathbf{X} \boldsymbol{\beta} = \frac{1}{2\sigma_e^2} r \sum_{i=1}^a \tau_i^2$					

El cociente de los cuadrados medios sigue una distribución F central, esto es:

$$\frac{CMM}{CME} \sim F(a-1, a(r-1)),$$

solo si $\lambda = 0$, esto es:

$$\sum_{i=1}^a \tau_i^2 = 0,$$

en otras palabras, solo si $\tau_i = 0$, para todo i . Por tanto bajo la hipotesis nula:

$$H_0 : \tau_i = 0 \quad \text{para todo } i,$$

$$\frac{CMM}{CME} \sim F(a-1, a(r-1)).$$

Ejemplo B.2. Ilustrativamente, asuma que se tienen los datos presentados en la tabla B.4. Presente el modelo y el respectivo análisis de varianza.

Tabla B.4. Datos ilustrativos

Tratamiento 1	Tratamiento 2	Tratamiento 3
4	3	10
6	5	15
9	2	11

Solución. El modelo asociado es:

$$y_{11} = 4 = \mu + \tau_1 + \underbrace{e_{11}}_{\text{una realización de } e_{11}}$$

$$y_{12} = 6 = \mu + \tau_1 + \underbrace{e_{12}}_{\text{una realización de } e_{12}}$$

$$y_{13} = 9 = \mu + \tau_1 + \underbrace{e_{13}}_{\text{una realización de } e_{13}}$$

$$y_{21} = 3 = \mu + \tau_2 + e_{21}$$

$$y_{22} = 5 = \mu + \tau_2 + e_{22}$$

$$y_{23} = 2 = \mu + \tau_2 + e_{23}$$

$$y_{31} = 10 = \mu + \tau_3 + e_{31}$$

$$y_{32} = 15 = \mu + \tau_3 + e_{32}$$

$$y_{33} = 11 = \mu + \tau_3 + \underbrace{e_{33}}_{\text{una realización de } e_{33}}.$$

Usando la restricción $\tau_1 + \tau_2 + \tau_3 = 0$, ver (B.5), se tiene $\tau_3 = -\tau_1 - \tau_2$, que se puede usar para realizar una reparametrización de rango completo.

Matricialmente:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$$

$$\mathbf{y} = \begin{pmatrix} 4 \\ 6 \\ 9 \\ 3 \\ 5 \\ 2 \\ 10 \\ 15 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \underbrace{\begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ e_{33} \end{pmatrix}}_{\text{una realización de } \mathbf{E}}$$

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 6 & 3 \\ 0 & 3 & 6 \end{pmatrix} \quad (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{9} & 0 & 0 \\ 0 & \frac{2}{9} & -\frac{1}{9} \\ 0 & -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$$

y

$$\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} = \frac{1}{3} \begin{pmatrix} \mathbf{J}_3 & 0 & 0 \\ 0 & \mathbf{J}_3 & 0 \\ 0 & 0 & \mathbf{J}_3 \end{pmatrix},$$

siendo $\mathbf{J}_a = \mathbf{1}_a \mathbf{1}_a^\top$. Dado que:

$$\begin{aligned} \mathbf{y}^\top \mathbf{y} &= 617, & \bar{y} &= 7.2\bar{2}, & n\bar{y}^2 &= 469.4\bar{4}, \\ \mathbf{y}^\top \mathbf{P}_x \mathbf{y} &= 585.6\bar{6}, & \hat{\sigma}_e &= 5.2\bar{2}, \end{aligned}$$

en la tabla B.5 se presenta el análisis de varianza, concluyendo que al menos un $\tau_i \neq 0$ con un P-valor de 0.0096.

Tabla B.5. Análisis de varianza para el ejemplo ilustrativo B.2

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	2	116.2 $\bar{2}$	58.1 $\bar{1}$	11.1277	0.0096
Error	6	31.3 $\bar{3}$	5.2 $\bar{2}$		
Total	8	147.5 $\bar{5}$			

Las instrucciones en R para los datos del ejemplo B.2 son:

```
remove(list=ls())
tratamientos=gl(3,3); Rta=c(4,6,9,3,5,2,10,15,11)
tratamientos=as.factor(tratamientos)
contrasts(tratamientos)=contr.sum(levels(tratamientos))
ajuste = aov(Rta ~ tratamientos)
ajuste$coefficients # Imprime los coeficientes estimados (beta)
summary(ajuste) # Imprime el análisis de varianza
```



3. Análisis de varianza con número desigual de replicaciones

Si al tratamiento asociado con el efecto τ_i se replica r_i veces, con r_i no necesariamente igual a $r_{i'}, i \neq i'$, se tiene el modelo:

$$Y_{ij} = \mu + \tau_i + e_{ij},$$

con $i = 1, 2, \dots, a$ y $j = 1, 2, \dots, r_i, \forall i$. La tabla B.6 presenta el análisis de varianza asociado a este modelo.

Tabla B.6. Análisis de varianza para un diseño a una vía de clasificación no balanceado

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado
Tratamiento	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{Y}^2$	$\frac{SCM}{glm}$	$\frac{CMM}{CME}$
Error	$\sum_{i=1}^a r_i - a$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$	$\frac{SCE}{gle}$	
Total	$\sum_{i=1}^a r_i - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{Y}^2$		

Bajo la hipótesis $H_0 : \tau_i = 0$, para todo i , se tiene que:

$$\frac{CMM}{CME} \sim F \left(a - 1, \sum_{i=1}^a r_i - a \right).$$

Ejemplo B.3. Considere los datos presentados en la tabla B.7. Presente el respectivo análisis de varianza.

Tabla B.7. Datos ilustrativos

Tratamiento 1	Tratamiento 2	Tratamiento 3
101	84	32
105	88	
94		

Solución Las respectivas matrices $\mathbf{y}$, $\mathbf{X}$, β y $\mathbf{E}$ son dadas por:

$$\mathbf{y} = \begin{pmatrix} 101 \\ 105 \\ 94 \\ 84 \\ 88 \\ 32 \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} \text{ y } \mathbf{E} = \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{pmatrix},$$

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, (\mathbf{X}^\top \mathbf{X})^- = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 504 \\ 300 \\ 172 \\ 32 \end{pmatrix}.$$

Entonces:

$$\mathbf{P}_x = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top$$

$$= \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{y}^\top \mathbf{P}_x \mathbf{y} = 45\,816$$

$$\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{y} = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}_x \mathbf{y} = 45\,886 - 45\,816 = 70$$

$$n\bar{y}^2 = 42\,336$$

$$\mathbf{y}^\top \mathbf{P}_x \mathbf{y} - n\bar{y}^2 = 3480, \quad \mathbf{y}^\top \mathbf{y} - n\bar{y}^2 = 3550.$$

El análisis de varianza respectivo se presenta en la tabla B.8, concluyendo que al menos un $\tau_i \neq 0$ con un P-valor de 0.0028.

Tabla B.8. Análisis de varianza para los datos ilustrativos presentados en la tabla B.7

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	F-calculado	P-valor
Tratam.	2	3480	1740	74.57	0.0028
Error	3	70	23.3		
Total	5	3550			

Las instrucciones en R con los datos del ejemplo B.3 son:

```
remove(list=ls())
tratamientos=c(1,1,1,2,2,3); Rta=c(101,105,94,84,88,32)
tratamientos=as.factor(tratamientos)
contrasts(tratamientos)=contr.sum(levels(tratamientos))
ajuste = aov(Rta ~ tratamientos)
ajuste$coefficients; summary(ajuste)
```



4. Tamaño de muestra

El cálculo del tamaño de muestra se puede obtener a partir del parámetro de no centralidad, λ , de la distribución F dado en (B.14):

$$\lambda = \frac{1}{2\sigma_e^2} \left(r \sum_{i=1}^a \tau_i^2 \right) = \frac{r}{2\sigma_e^2} \left(\tau_{\text{máx}}^2 + \tau_{\text{mín}}^2 + \sum_{i \neq \text{máx}, i \neq \text{mín}} \tau_i^2 \right). \quad (\text{B.15})$$

Por otro lado, se sabe de (B.5) que:

$$\sum_{i=1}^a \tau_i = 0$$

$$\tau_{\text{máx}} + \tau_{\text{mín}} + \sum_{i \neq \text{máx}, i \neq \text{mín}} \tau_i = 0$$

$$(a-2) \underbrace{\left(\frac{1}{a-2} \sum_{i \neq \text{máx}, \text{mín}} \tau_i \right)}_{\tau^*} = -(\tau_{\text{máx}} + \tau_{\text{mín}})$$

$$\tau^* = \frac{-(\tau_{\text{máx}} + \tau_{\text{mín}})}{a-2},$$

note que $\tau_{\text{mín}} \leq \tau^* \leq \tau_{\text{máx}}$. Si la diferencia que sea desea detectar es $\Delta = \tau_{\text{máx}} - \tau_{\text{mín}}$, se puede asumir $\tau^* = 0$ y entonces [33, p.184]:

$$0 = \frac{-(\tau_{\text{máx}} + \tau_{\text{mín}})}{a-2}$$

$$\tau_{\text{máx}} = -\tau_{\text{mín}}$$

$$\Delta = \tau_{\text{máx}} - \tau_{\text{mín}} = \tau_{\text{máx}} + \tau_{\text{máx}} = 2\tau_{\text{máx}},$$

y λ en (B.15) quedaría:

$$\lambda = \frac{r}{2\sigma_e^2} (\tau_{\text{mín}}^2 + \tau_{\text{máx}}^2 + (a-2)\tau^{*2}) = \frac{r}{2\sigma_e^2} (\tau_{\text{máx}}^2 + \tau_{\text{máx}}^2)$$

$$\lambda = \frac{r\Delta^2}{4\sigma_e^2}. \quad (\text{B.16})$$

Por otro lado, sea ϕ^2 el cociente entre la varianza de $\{\tau_1, \tau_2, \tau_3, \dots, \tau_a\}$:

$$\mathbb{V}(\tau' s) = \frac{\sum_{i=1}^a \tau_i^2}{a},$$

y la varianza de la media muestral de tratamiento, $\mathbb{V}(\bar{Y}_{i*}) = \frac{\sigma_e^2}{r}$, esto es:

$$\phi^2 = \frac{\mathbb{V}(\tau' s)}{\mathbb{V}(\bar{Y}_{i*})} = \frac{\frac{\sum_{i=1}^a \tau_i^2}{a}}{\frac{\sigma_e^2}{r}}.$$

Despejando $\sum_{i=1}^a \tau_i^2$ y reemplazando en (B.15) se tiene que:

$$\sum_{i=1}^a \tau_i^2 = \frac{a\phi^2\sigma_e^2}{r}$$

$$\lambda = \frac{r}{2\sigma_e^2} \left(\sum_{i=1}^a \tau_i^2 \right) = \frac{r}{2\sigma_e^2} \left(\frac{a\phi^2\sigma_e^2}{r} \right) = \frac{a\phi^2}{2}. \quad (\text{B.17})$$

Igualando (B.16) y (B.17):

$$\frac{r\Delta^2}{4\sigma_e^2} = \frac{a\phi^2}{2}$$

$$r = \frac{2a\sigma_e^2\phi^2}{\Delta^2}.$$

4.1. Programa básico para tablas sobre tamaño de muestra

```
#####
### Programa Básico para tablas de tamaño de muestra ###
#####
remove(list=ls())
a=6 # Número de tratamientos
alpha=0.05 # nivel de significancia
phi=seq(0.10,3,0.05)
replicas=c(2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,250)
A2=length(replicas);A3=length(phi)
v1=a-1 # modificar de acuerdo al diseño (grados de
# libertad del numerador)
v2=a*(replicas-1) # modificar según los grados de libertad del
# error (gle) del diseño

tablaSampleSize=NULL
for(i2 in 1:A2)
{ potencia=matrix(99,1,A3)
  for(i3 in 1:A3)
  {
    power=pf(qf(alpha,v1,v2[i2],lower.tail=FALSE,log.p=FALSE),
             v1,v2[i2],a*(phi[i3]^2),lower.tail=FALSE,log.p=FALSE)
    potencia[1,i3]=round(power,2)
  }
  salida=cbind(alpha,replicas[i2],v2[i2],potencia)
  names.col=c("alpha","r","v2",phi); colnames(salida)=names.col
  tablaSampleSize=rbind(tablaSampleSize,salida)
}
tablaSampleSize
```

5. Análisis de varianza - submuestreo

Se parte del modelo dado en (3.10), y para facilitar los cálculos no se considerará ninguna restricción sobre los τ_i . El modelo dado en (3.10) se puede escribir matricialmente:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta},$$

con $\mathbf{Y}_{n \times 1}$, $\boldsymbol{\eta}_{n \times 1}$, $\mathbf{Z}_{n \times ar'}$ y $\boldsymbol{\varepsilon}_{ar' \times 1}$ definidos en (3.11):

$$\boldsymbol{\beta}_{(a+1) \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_a \end{pmatrix} \quad \text{y} \quad \mathbf{X}_{n \times (a+1)} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix},$$

en que una forma reducida de $\mathbf{X}$ y $\mathbf{Z}$ es dada por²:

$$\begin{aligned} \mathbf{X} &= \left(\mathbf{1}_{ar'n^*} \mid \mathbf{I}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \right)_{ar'n^* \times (a+1)} \\ \mathbf{Z} &= \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*} \right)_{ar'n^* \times ar'}, \end{aligned} \tag{B.18}$$

²En la notación se va a usar $\otimes$ para indicar producto Kronecker. El producto Kronecker de dos matrices, $\mathbf{A}$ de orden $m \times n$ y $\mathbf{B}$ de orden $p \times q$, se denota por $\mathbf{A} \otimes \mathbf{B}$ y es definido como una matriz de orden $mp \times nq$ tal que:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}_{mp \times nq},$$

ver [30, p. 337].

al igual que $\mathbb{V}(\boldsymbol{\varepsilon})$ y $\mathbb{V}(\boldsymbol{\beta})$:

$$\begin{aligned}\mathbb{V}(\boldsymbol{\varepsilon}) &= \sigma_{\varepsilon}^2 \mathbf{I}_{ar'} \\ \mathbb{V}(\boldsymbol{\eta}) &= \sigma_{\eta}^2 \mathbf{I}_{ar'n^*},\end{aligned}$$

por tanto:

$$\begin{aligned}\boldsymbol{\mu} &= \mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \\ \mathbf{V} &= \mathbb{V}(\mathbf{Y}) = \mathbb{V}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta}) = \mathbb{V}(\mathbf{Z}\boldsymbol{\varepsilon} + \boldsymbol{\eta}) \\ &= \mathbb{V}(\mathbf{Z}\boldsymbol{\varepsilon}) + \mathbb{V}(\boldsymbol{\eta}) = \sigma_{\varepsilon}^2 \mathbf{Z}\mathbf{Z}^{\top} + \sigma_{\eta}^2 \mathbf{I}_{ar'n^*}.\end{aligned}$$

5.1. Distribución y esperanza de las sumas de cuadrados

5.1.1. Suma de cuadrados del modelo (SCM)

$$\begin{aligned}SCM &= \mathbf{Y}^{\top} \mathbf{P}_x \mathbf{Y} - n\bar{y}^2 \\ &= \mathbf{Y}^{\top} \underbrace{\left[\mathbf{X}(\mathbf{X}^{\top} \mathbf{X})^{-1} \mathbf{X}^{\top} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\top} \right]}_{\mathbf{A}} \mathbf{Y} = \mathbf{Y}^{\top} \mathbf{A} \mathbf{Y},\end{aligned}$$

esto es:

$$\mathbf{A} = \mathbf{X}(\mathbf{X}^{\top} \mathbf{X})^{-1} \mathbf{X}^{\top} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\top} = \mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\top}. \quad (\text{B.19})$$

Para encontrar la distribución de la suma de cuadrados del modelo, SCM , note que:

- La matriz $\mathbf{A}$ es simétrica.
- Asumiendo que:

$$\mathbf{Y} \sim N\left(\boldsymbol{\mu}, \sigma_{\varepsilon}^2 \mathbf{Z}\mathbf{Z}^{\top} + \sigma_{\eta}^2 \mathbf{I}_{ar'n^*}\right),$$

se tiene $\mathbf{V} = \sigma_{\varepsilon}^2 \mathbf{Z}\mathbf{Z}^{\top} + \sigma_{\eta}^2 \mathbf{I}_{ar'n^*}$. Sea:

$$\mathbf{Y}^* = \frac{1}{\sqrt{n^* \sigma_{\varepsilon}^2 + \sigma_{\eta}^2}} \mathbf{Y},$$

con:

$$\begin{aligned}\boldsymbol{\mu}^* &= \mathbb{E}(\mathbf{Y}^*) = \frac{1}{\sqrt{n^*\sigma_\varepsilon^2 + \sigma_\eta^2}}(\mathbf{X}\boldsymbol{\beta}) = \frac{1}{\sqrt{n^*\sigma_\varepsilon^2 + \sigma_\eta^2}}\boldsymbol{\mu} \quad \text{y} \\ \mathbf{V}^* &= \mathbb{V}(\mathbf{Y}^*) = \frac{1}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \left(\sigma_\varepsilon^2 \mathbf{Z}\mathbf{Z}^\top + \sigma_\eta^2 \mathbf{I}_{ar'n^*} \right) = \frac{1}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \mathbf{V},\end{aligned}\tag{B.20}$$

tal que $\frac{1}{\sqrt{n^*\sigma_\varepsilon^2 + \sigma_\eta^2}}\mathbf{Y} \sim N\left(\frac{1}{\sqrt{n^*\sigma_\varepsilon^2 + \sigma_\eta^2}}\boldsymbol{\mu}, \mathbf{V}^*\right)$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{A}\mathbf{V}^*$:

$$\mathbf{A}\mathbf{V}^* \mathbf{A}\mathbf{V}^* = \frac{1}{(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)^2} \mathbf{A}\mathbf{V} \mathbf{A}\mathbf{V},$$

en que:

$$\mathbf{A}\mathbf{V} = \left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \left(\sigma_\varepsilon^2 \mathbf{Z}\mathbf{Z}^\top + \sigma_\eta^2 \mathbf{I}_n \right).$$

Resolviendo por partes:

$$\begin{aligned}\mathbf{Z}\mathbf{Z}^\top &= \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*} \right) \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*}^\top \right) = \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) \\ \mathbf{1}_n \mathbf{1}_n^\top \mathbf{Z}\mathbf{Z}^\top &= \left(\mathbf{1}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \right) \left(\mathbf{1}_a^\top \otimes \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}^\top \right) \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) \\ &= n^* \left(\mathbf{1}_a \mathbf{1}_a^\top \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right).\end{aligned}$$

$$\begin{aligned}\mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} \mathbf{1}_{ar'n^*}^\top \\ \mathbf{I}_a \otimes \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}^\top \end{pmatrix} \begin{pmatrix} \mathbf{1}_{ar'n^*} & \mathbf{I}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \end{pmatrix} \\ &= \begin{pmatrix} ar'n^* & r'n^* \mathbf{1}_a^\top \\ r'n^* \mathbf{1}_a & r'n^* \mathbf{I}_a \end{pmatrix} = r'n^* \begin{pmatrix} a & \mathbf{1}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix}\end{aligned}$$

$$(\mathbf{X}^\top \mathbf{X})^- = \frac{1}{r'n^*} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \quad \text{es una de las posibles inversas}$$

generalizadas

$$\begin{aligned}
 \mathbf{P}_x &= \begin{bmatrix} \mathbf{1}_{ar'n^*} & | & \mathbf{I}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \end{bmatrix} \frac{1}{r'n^*} \begin{bmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{bmatrix} \begin{bmatrix} \mathbf{1}_{ar'n^*}^\top \\ \mathbf{I}_a \otimes \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}^\top \end{bmatrix} \\
 &= \frac{1}{r'n^*} \begin{bmatrix} \mathbf{0} & | & \mathbf{I}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{ar'n^*}^\top \\ \mathbf{I}_a \otimes \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}^\top \end{bmatrix} \\
 &= \frac{1}{r'n^*} \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) \\
 \mathbf{P}_x \mathbf{Z} \mathbf{Z}^\top &= \frac{1}{r'n^*} \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) \left(\mathbf{I}_a \otimes \mathbf{I}_{r'} \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) \\
 &= \frac{1}{r'} \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right)
 \end{aligned}$$

Entonces:

$$\begin{aligned}
 \mathbf{A} \mathbf{V} &= \sigma_\varepsilon^2 \mathbf{P}_x \mathbf{Z} \mathbf{Z}^\top - \frac{1}{n} \sigma_\varepsilon^2 \mathbf{1}_n \mathbf{1}_n^\top \mathbf{Z} \mathbf{Z}^\top + \sigma_\eta^2 \left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \\
 &= \sigma_\varepsilon^2 \frac{1}{r'} \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{n} \sigma_\varepsilon^2 n^* \left(\mathbf{1}_a \mathbf{1}_a^\top \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) + \\
 &\quad \sigma_\eta^2 \frac{1}{r'n^*} \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{n} \sigma_\eta^2 \mathbf{1}_n \mathbf{1}_n^\top \quad \text{note que } n = ar'n^* \\
 &= \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r'n^*} \right) \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{n} \right) \mathbf{1}_n \mathbf{1}_n^\top \\
 &= \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r'n^*} \right) \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n \mathbf{1}_n^\top \right], \\
 \mathbf{A} \mathbf{V} \mathbf{A} \mathbf{V} &= \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r'n^*} \right)^2 \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n \mathbf{1}_n^\top \right] \times \\
 &\quad \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n \mathbf{1}_n^\top \right] = \\
 &\quad \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r'n^*} \right)^2 \left[r'n^* \left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{2r'n^*}{a} \mathbf{1}_n \mathbf{1}_n^\top + \frac{r'n^*}{a} \mathbf{1}_n \mathbf{1}_n^\top \right] \\
 &= \frac{(n^* \sigma_\varepsilon^2 + \sigma_\eta^2)^2}{r'n^*} \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n \mathbf{1}_n^\top \right]
 \end{aligned}$$

y

$$\begin{aligned}
 \mathbf{A}\mathbf{V}^*\mathbf{A}\mathbf{V}^* &= \frac{1}{(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)^2} \mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V} \\
 &= \frac{1}{\cancel{(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)^2}} \frac{(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)^2}{r'n^*} \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'}\mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}\mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n\mathbf{1}_n^\top \right] \\
 &= \frac{1}{r'n^*} \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'}\mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*}\mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n\mathbf{1}_n^\top \right] = \mathbf{A}\mathbf{V}^*.
 \end{aligned}$$

Por tanto $\mathbf{A}\mathbf{V}^*$ es idempotente y entonces:

$$\frac{SCM}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(rango(\mathbf{A}), \lambda),$$

en que:

$$\begin{aligned}
 rango(\mathbf{A}) &= rango\left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n\mathbf{1}_n^\top\right) = a - 1 \text{ y} \\
 \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* = \frac{1}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \\
 &= \frac{1}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \boldsymbol{\beta}^\top \mathbf{X}^\top \left\{ \mathbf{P}_x - \frac{1}{n} \mathbf{1}_n\mathbf{1}_n^\top \right\} \mathbf{X} \boldsymbol{\beta} \\
 \lambda &= \frac{1}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \left(\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \frac{1}{n} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{1}_n\mathbf{1}_n^\top \mathbf{X} \boldsymbol{\beta} \right),
 \end{aligned}$$

como

$$\begin{aligned}
 \mathbf{X} \boldsymbol{\beta} &= \left(\mathbf{1}_{ar'n^*} \mid \mathbf{I}_a \otimes \mathbf{1}_{r'} \otimes \mathbf{1}_{n^*} \right) \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_a \end{pmatrix} \\
 &= \left[\underbrace{\mu + \tau_1 \quad \dots \quad \mu + \tau_1}_{r'n^*} \quad \dots \quad \underbrace{\mu + \tau_a \quad \dots \quad \mu + \tau_a}_{r'n^*} \right]^\top,
 \end{aligned}$$

entonces:

$$\begin{aligned}
 \lambda &= \frac{1}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \left(\sum_{i=1}^a r' n^* (\mu + \tau_i)^2 - \frac{1}{n} \left[\sum_{i=1}^a r' n^* (\mu + \tau_i) \right]^2 \right) \\
 &= \frac{1}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \underbrace{\left(r' n^* a \mu^2 + r' n^* \sum_{i=1}^a \tau_i^2 - \frac{1}{a r' n^*} [r' n^* a \mu]^2 \right)}_{\mu^\top \mathbf{A} \mu} \\
 &= \frac{r' n^*}{2(n^*\sigma_\varepsilon^2 + \sigma_\eta^2)} \left(\sum_{i=1}^a \tau_i^2 \right) = \frac{r'}{2 \left(\sigma_\varepsilon^2 + \frac{\sigma_\eta^2}{n^*} \right)} \sum_{i=1}^a \tau_i^2, \tag{B.21}
 \end{aligned}$$

y

$$\frac{SCM}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2 \left(a - 1, \frac{r'}{2 \left(\sigma_\varepsilon^2 + \frac{\sigma_\eta^2}{n^*} \right)} \sum_{i=1}^a \tau_i^2 \right).$$

5.1.2. Suma de cuadrados del error experimental (SCEE)

$$SCEE = \mathbf{Y}^\top \underbrace{(\mathbf{P}_z - \mathbf{P}_x)}_{\mathbf{A}_1} \mathbf{Y} = \mathbf{Y}^\top \mathbf{A}_1 \mathbf{Y}, \tag{B.22}$$

esto es:

$$\mathbf{A}_1 = \mathbf{P}_z - \mathbf{P}_x = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top. \tag{B.23}$$

Para encontrar la distribución de la suma de cuadrados del error experimental, notado por $SCEE$, note que:

- $\mathbf{A}_1$ es simétrica.
- Se puede mostrar que $\mathbf{A}_1 \mathbf{V}^*$ es idempotente, ver ejercicio 5.1. Aplicando el teorema 2.2:

$$\frac{\mathbf{Y}^\top (\mathbf{P}_z - \mathbf{P}_x) \mathbf{Y}}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(\text{rango}(\mathbf{A}_1), \lambda_1), \tag{B.24}$$

en que $\lambda_1 = 0$, ver ejercicio 5.2, y $\text{rango}(\mathbf{A}_1) = a(r' - 1)$, esto es:

$$\frac{\mathbf{Y}^\top (\mathbf{P}_z - \mathbf{P}_x) \mathbf{Y}}{n^*\sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(a(r' - 1)).$$

5.1.3. Suma de cuadrados del error observacional (SCEO)

$$SCEO = \mathbf{Y}^\top \underbrace{(\mathbf{I} - \mathbf{P}_z)}_{\mathbf{A}_2} \mathbf{Y} = \mathbf{Y}^\top \mathbf{A}_2 \mathbf{Y}. \quad (\text{B.25})$$

Para encontrar la distribución de la suma de cuadrados del error observacional, notado por $SCEO$, note que:

- $\mathbf{A}_2 = \mathbf{I} - \mathbf{P}_z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top$ es simétrica.
- Se puede mostrar que $\mathbf{A}_2 \mathbf{V}^*$ es idempotente, ver ejercicio 5.5. Aplicando el teorema 2.2:

$$\frac{\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_z) \mathbf{Y}}{n^* \sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(\text{rango}(\mathbf{A}_2), \lambda_2), \quad (\text{B.26})$$

con $\text{rango}(\mathbf{A}_2) = ar'(n^* - 1)$, ver ejercicio 5.5.

5.2. Esperanza de los cuadrados medios

Los **cuadrados medios** (CM), se obtienen al dividir la suma de cuadrados entre los grados de libertad, ver tabla B.9. La esperanza del cuadrado medio del modelo, del cuadrado medio del error experimental y del cuadrado medio del error observacional son:

$$\begin{aligned} \mathbb{E}(CMM) &= \mathbb{E} \left(\frac{SCM}{a-1} \right) = \frac{1}{a-1} \mathbb{E} \left[\mathbf{Y}^\top \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{Y} \right] \\ &= \frac{1}{a-1} [\text{traza}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}] \\ &= \frac{1}{a-1} \left[\text{traza} \left\{ \left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r' n^*} \right) \left[\left(\mathbf{I}_a \otimes \mathbf{1}_{r'} \mathbf{1}_{r'}^\top \otimes \mathbf{1}_{n^*} \mathbf{1}_{n^*}^\top \right) - \frac{1}{a} \mathbf{1}_n \mathbf{1}_n^\top \right] \right\} \right. \\ &\quad \left. + r' n^* \sum_{i=1}^a \tau_i^2 \right] \\ &= \frac{1}{a-1} \left[\left(\frac{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}{r' n^*} \right) [ar' n^* - r' n^*] + r' n^* \sum_{i=1}^a \tau_i^2 \right] \\ &= n^* \sigma_\varepsilon^2 + \sigma_\eta^2 + \frac{r' n^*}{a-1} \sum_{i=1}^a \tau_i^2 \end{aligned}$$

$$\begin{aligned}\mathbb{E}(CMEE) & \underbrace{=}_{\text{ejercicio 5.7}} n^* \sigma_\varepsilon^2 + \sigma_\eta^2 \\ \mathbb{E}(CMEO) & \underbrace{=}_{\text{ejercicio 5.8}} \sigma_\eta^2.\end{aligned}\tag{B.27}$$

5.3. Cociente de cuadrados medios

Se puede mostrar que las formas cuadráticas $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ y $\mathbf{Y}^\top \mathbf{A}_1 \mathbf{Y}$ son independientes, ver ejercicio 5.9, entonces:

$$\frac{\frac{SCM}{n^* \sigma_\varepsilon^2 + \sigma_\eta^2}}{a-1} = \frac{\frac{SCM}{a-1}}{\frac{SC EE}{a(r'-1)}} = \frac{CMM}{CMEE} \sim F(a-1, a(r'-1), \lambda),$$

con λ definido en (B.21). La tabla B.9 presenta de forma reducida los resultados obtenidos de las esperanzas de los cuadrados medios y su distribución.

Tabla B.9. Análisis de varianza incluyendo esperanza de los cuadrados medios

Causa de variac.	Grados de libertad	Suma de cuadrados	Cuadr. medio	$\mathbb{E}(CM)$	
Modelo	$a-1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{y}^2$	$\frac{SCT}{a-1}$	$\sigma_\eta^2 + n^* \sigma_\varepsilon^2 + \frac{r' n^*}{a-1} \sum_{i=1}^a \tau_i^2$	$\frac{SCM}{n^* \sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(a-1, \lambda)$
Error ex- perim.	$a(r'-1)$	$\mathbf{Y}^\top \mathbf{P}_z \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_x \mathbf{Y}$	$\frac{SC EE}{a(r'-1)}$	$\sigma_\eta^2 + n^* \sigma_\varepsilon^2$	$\frac{SC EE}{n^* \sigma_\varepsilon^2 + \sigma_\eta^2} \sim \chi^2(a[r'-1])$
Error ob- serv.	$ar'(n^*-1)$	$\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_z \mathbf{Y}$	$\frac{SC EO}{ar'(n^*-1)}$	σ_η^2	
Total	$ar' n^* - 1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			

Finalmente, si $\lambda = 0$, el cociente:

$$\frac{CMM}{CMEE} \sim F(a-1, a(r'-1)),$$

implicando que:

$$\sum_{i=1}^a \tau_i^2 = 0,$$

y por tanto, $\tau_i = 0$ para todo i . Luego, bajo la hipótesis nula:

$$H_0 : \tau_i = 0 \quad \text{para todo } i$$

se tiene que:

$$\frac{CMM}{CME E} \sim F(a-1, a[r'-1]).$$

5.4. Ejercicios

5.1 Mostrar que $\mathbf{A}_1 \mathbf{V}^*$ es idempotente, en que $\mathbf{A}_1$ fue definida en (B.23) y $\mathbf{V}^*$ en (B.20), también mostrar que $\text{rango}(\mathbf{A}_1) = \text{rango}(\mathbf{P}_z - \mathbf{P}_x) = a(r' - 1)$.

5.2 En (B.24), mostrar que $\lambda_1 = 0$.

5.3 Teniendo en cuenta la definición de $\mathbf{Z}$ dada (B.18), mostrar que:

$$\mathbb{E}(\mathbf{Y}^\top \mathbf{P}_z \mathbf{Y}) = \sigma_\varepsilon^2 ar' n^* + \sigma_\eta^2 ar' + ar' n^* \mu^2 + r' n^* \sum_{i=1}^a \tau_i^2.$$

5.4 Teniendo en cuenta la definición de $\mathbf{X}$ dada (B.18), mostrar que:

$$\mathbb{E}(\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y}) = \sigma_\varepsilon^2 an^* + \sigma_\eta^2 a + ar' n^* \mu^2 + r' n^* \sum_{i=1}^a \tau_i^2.$$

5.5 Con base en la definición dada de $\mathbf{A}_2$ en (B.25), mostrar que $\mathbf{A}_2 \mathbf{V}^*$ es idempotente y que $\text{rango}(\mathbf{A}_2) = \text{rango}(\mathbf{I} - \mathbf{P}_z) = ar'(n^* - 1)$.

5.6 En (B.26), encontrar λ_2 .

5.7 Mostrar que $\mathbb{E}(CME E) = \mathbb{E}\left(\frac{SCEE}{glee}\right) = n^* \sigma_\varepsilon^2 + \sigma_\eta^2$, en que $SCEE$ fue definido en (B.22) y $glee$ corresponde a los grados de libertad del error experimental.

5.8 Mostrar que $\mathbb{E}(CME O) = \mathbb{E}\left(\frac{SCEO}{gleo}\right) = \sigma_\eta^2$, en que $SCEO$ fue definido en (B.25) y $gleo$ corresponde a los grados de libertad del error observacional.

5.9 Mostrar que $\mathbf{A}_1 \mathbf{A} = (\mathbf{P}_z - \mathbf{P}_x)(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top) = 0$, en que $\mathbf{A}_1$ fue definida en (B.23) y $\mathbf{A}$ en (B.19).

5.10 Encontrar $\mathbf{A}_1 \mathbf{A}_2$ en que $\mathbf{A}_1$ fue definida en (B.23) y $\mathbf{A}_2$ en (B.25).

6. Análisis de varianza - modelo de efectos aleatorios

El modelo (3.12) se puede escribir matricialmente como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E},$$

con:

$$\mathbf{X} = \left(\mathbf{1}_n \mid \mathbf{I}_a \otimes \mathbf{1}_r \right), \boldsymbol{\beta} = \left(\mu \quad \tau_1 \quad \tau_2 \quad \dots \quad \tau_a \right)^\top \text{ y}$$

$$\mathbf{E} = \left(e_{11} \quad e_{12} \quad \dots \quad e_{ar} \right)^\top.$$

Por tanto:

$$\begin{aligned} \mathbb{E}(\mathbf{Y}) &= \mathbb{E} \left(\mu \mathbf{1}_n + (\mathbf{I}_a \otimes \mathbf{1}_r) \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_a \end{pmatrix} + \mathbf{E} \right) \\ &= \mu \mathbf{1}_n + (\mathbf{I}_a \otimes \mathbf{1}_r) \mathbb{E} \left[\begin{pmatrix} \tau_1 \\ \vdots \\ \tau_a \end{pmatrix} \right] + \mathbb{E}(\mathbf{E}) = \mu \mathbf{1}_n \\ \mathbb{V}(\mathbf{Y}) &= \mathbb{V}(\mathbf{X}\boldsymbol{\beta} + \mathbf{E}) = \mathbb{V} \left(\mu \mathbf{1}_n + (\mathbf{I}_a \otimes \mathbf{1}_r) \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_a \end{pmatrix} + \mathbf{E} \right) \\ &= (\mathbf{I}_a \otimes \mathbf{1}_r) \mathbb{V} \left[\begin{pmatrix} \tau_1 \\ \vdots \\ \tau_a \end{pmatrix} \right] (\mathbf{I}_a \otimes \mathbf{1}_r)^\top + \mathbb{V}(\mathbf{E}) \\ &= \sigma_\tau^2 (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) + \sigma_e^2 \mathbf{I}_n = \mathbf{V}. \end{aligned}$$

6.1. Distribución y esperanza de las sumas de cuadrados

6.1.1. Suma de cuadrados del modelo

$$\begin{aligned} SCM &= \mathbf{Y}^\top \underbrace{\left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)}_{\mathbf{A}} \mathbf{Y} = \mathbf{Y}^\top \left(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-} \mathbf{X}^\top - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{Y} \\ &= \mathbf{Y}^\top \mathbf{A} \mathbf{Y}, \end{aligned}$$

en que:

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix}^\top \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix} \quad (\text{B.28}) \\ &= \begin{pmatrix} ra & r\mathbf{1}_a^\top \\ r\mathbf{1}_a & r\mathbf{I}_a \end{pmatrix} = r \begin{pmatrix} a & \mathbf{1}_a^\top \\ \mathbf{1}_a & \mathbf{I}_a \end{pmatrix} \\ (\mathbf{X}^\top \mathbf{X})^{-} &= \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \text{ es una de las posibles inversas} \end{aligned}$$

generalizadas

$$\begin{aligned} \mathbf{P}_x &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-} \mathbf{X}^\top \quad (\text{B.29}) \\ &= \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix} \frac{1}{r} \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} \mathbf{1}_n & | & \mathbf{I}_a \otimes \mathbf{1}_r \end{pmatrix}^\top \\ &= \frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top. \quad (\text{B.30}) \end{aligned}$$

Para encontrar la distribución de la suma de cuadrados del modelo (SCM), note que:

- La matriz $\mathbf{A} = \mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ es simétrica.
- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ en que $\boldsymbol{\mu} = \mu \mathbf{1}_n$ y que:

$$\mathbf{V} = \sigma_\tau^2 \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) + \sigma_e^2 \mathbf{I}_n,$$

sea:

$$\mathbf{Y}^* = \frac{1}{\sqrt{r\sigma_\tau^2 + \sigma_e^2}} \mathbf{Y},$$

con:

$$\begin{aligned}\boldsymbol{\mu}^* &= \mathbb{E}(\mathbf{Y}^*) = \frac{1}{\sqrt{r\sigma_\tau^2 + \sigma_e^2}} \mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \mathbf{E}) = \frac{1}{\sqrt{r\sigma_\tau^2 + \sigma_e^2}} \boldsymbol{\mu} \mathbf{1}_n \quad \text{y} \\ \mathbf{V}^* &= \mathbb{V}(\mathbf{Y}^*) = \frac{1}{r\sigma_\tau^2 + \sigma_e^2} \left[\sigma_\tau^2 \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) + \sigma_e^2 \mathbf{I}_n \right] \\ &= \frac{1}{r\sigma_\tau^2 + \sigma_e^2} \mathbf{V},\end{aligned}$$

esto es, $\frac{1}{\sqrt{r\sigma_\tau^2 + \sigma_e^2}} \mathbf{Y} \sim N \left(\frac{\boldsymbol{\mu}}{\sqrt{r\sigma_\tau^2 + \sigma_e^2}} \mathbf{1}, \frac{1}{r\sigma_\tau^2 + \sigma_e^2} \mathbf{V} \right)$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{AV}^*$:

$$\mathbf{AV}^* \mathbf{AV}^* = \frac{1}{(r\sigma_\tau^2 + \sigma_e^2)^2} \mathbf{AV} \mathbf{AV},$$

en que:

$$\mathbf{AV} = \left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \left[\sigma_\tau^2 \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) + \sigma_e^2 \mathbf{I}_n \right].$$

Usando (B.30) y $n = ar$,

$$\begin{aligned}\mathbf{AV} &= \left(\frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \left(\sigma_\tau^2 \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top + \sigma_e^2 \mathbf{I}_{ar} \right) \\ &= \frac{\sigma_\tau^2}{r} \mathbf{I}_a \otimes r \mathbf{1}_r \mathbf{1}_r^\top + \frac{\sigma_e^2}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{\sigma_\tau^2}{ar} r \mathbf{1}_{ar} \mathbf{1}_{ar}^\top - \frac{\sigma_e^2}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \\ &= \left(\sigma_\tau^2 + \frac{\sigma_e^2}{r} \right) \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \left(\sigma_\tau^2 + \frac{\sigma_e^2}{r} \right) \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \\ \mathbf{AV} &= (r\sigma_\tau^2 + \sigma_e^2) \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \tag{B.31} \\ \mathbf{AV}^* &= \frac{1}{(r\sigma_\tau^2 + \sigma_e^2)} (r\sigma_\tau^2 + \sigma_e^2) \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \\ &= \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \\ \mathbf{AV}^* \mathbf{AV}^* &= \frac{1}{r^2} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \\ &= \frac{1}{r^2} \left(\mathbf{I}_a \otimes r \mathbf{1}_r \mathbf{1}_r^\top - \frac{2r}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top + \frac{ar}{a^2} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \\ &= \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) = \mathbf{AV}^*,\end{aligned}$$

por tanto $\mathbf{A}\mathbf{V}^*$ es idempotente y

$$\mathbf{Y}^{*\top} \mathbf{A} \mathbf{Y}^* = \frac{\mathbf{Y}^\top \mathbf{A} \mathbf{Y}}{r\sigma_\tau^2 + \sigma_e^2} = \frac{SCM}{r\sigma_\tau^2 + \sigma_e^2} \sim \chi^2(rango(\mathbf{A}), \lambda),$$

con:

$$\begin{aligned} rango(\mathbf{A}) &= rango(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top) = a - 1, \quad y \\ \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* = \frac{1}{2} \times \frac{1}{(\sqrt{r\sigma_\tau^2 + \sigma_e^2})^2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \\ &= \frac{1}{2} \frac{\mu^2}{r\sigma_\tau^2 + \sigma_e^2} \mathbf{1}_n^\top \left[\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right] \mathbf{1}_n \\ \lambda &= \frac{1}{2} \frac{\mu^2}{r\sigma_\tau^2 + \sigma_e^2} \left[\underbrace{\mathbf{1}_n^\top \mathbf{P}_x \mathbf{1}_n}_{\substack{\text{1}_n \\ n}} - \underbrace{\frac{1}{n} \mathbf{1}_n^\top \mathbf{1}_n \mathbf{1}_n^\top \mathbf{1}_n}_{\substack{\cancel{\mathcal{N}} \\ n}} \right] = 0, \end{aligned} \quad (\text{B.32})$$

luego:

$$\frac{SCM}{r\sigma_\tau^2 + \sigma_e^2} \sim \chi^2(a - 1).$$

6.1.2. Suma de cuadrados del error

$$SCE = \mathbf{Y}^\top \underbrace{(\mathbf{I} - \mathbf{P}_x)}_{\mathbf{A}_1} \mathbf{Y} = \mathbf{Y}^\top \mathbf{A}_1 \mathbf{Y}.$$

Para encontrar la distribución de la suma de cuadrados del error (SCE), note que:

- La matriz $\mathbf{A}_1 = \mathbf{I}_{ar} - \mathbf{P}_x$ es simétrica.
- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ en que $\boldsymbol{\mu} = \mu \mathbf{1}_n$ y

$$\mathbf{V} = \sigma_\tau^2 \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) + \sigma_e^2 \mathbf{I}_n,$$

sea:

$$\mathbf{Y}^{**} = \frac{1}{\sigma_e} \mathbf{Y},$$

con:

$$\begin{aligned} \boldsymbol{\mu}^{**} &= \mathbb{E}(\mathbf{Y}^{**}) = \frac{1}{\sigma_e} (\mathbf{X}\boldsymbol{\beta} + \mathbf{E}) = \frac{1}{\sigma_e} \mu \mathbf{1}_n \quad y \\ \mathbf{V}^{**} &= \mathbb{V}(\mathbf{Y}^{**}) = \frac{1}{\sigma_e^2} \left[\sigma_\tau^2 \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) + \sigma_e^2 \mathbf{I}_n \right] = \frac{1}{\sigma_e^2} \mathbf{V}, \end{aligned}$$

tal que $\frac{1}{\sigma_e} \mathbf{Y} \sim N\left(\frac{\mu}{\sigma_e} \mathbf{1}, \frac{1}{\sigma_e^2} \mathbf{V}\right)$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{A}_1 \mathbf{V}^{**}$:

$$\begin{aligned}
 \mathbf{A}_1 \mathbf{V} &= (\mathbf{I}_{ar} - \mathbf{P}_x) (\sigma_\tau^2 (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) + \sigma_e^2 \mathbf{I}_n) \\
 &= \left(\mathbf{I}_{ar} - \frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) [\sigma_\tau^2 (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) + \sigma_e^2 \mathbf{I}_n] \\
 &= \sigma_\tau^2 \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top + \sigma_e^2 \mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) (\sigma_\tau^2 \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top + \sigma_e^2 \mathbf{I}_{ar}) \\
 &= \cancel{\sigma_\tau^2 \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top} + \sigma_e^2 \mathbf{I}_{ar} - \cancel{\frac{\sigma_\tau^2}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top} - \frac{\sigma_e^2}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \\
 \mathbf{A}_1 \mathbf{V} &= \sigma_e^2 \left[\mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \right] \tag{B.33} \\
 \mathbf{A}_1 \mathbf{V}^{**} &= \frac{1}{\sigma_e^2} \mathbf{A}_1 \mathbf{V} = \mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \\
 \mathbf{A}_1 \mathbf{V}^{**} \mathbf{A}_1 \mathbf{V}^{**} &= \left[\mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \right] \left[\mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \right] \\
 &= \mathbf{I}_{ar} - \frac{2}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) + \frac{r}{r^2} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \\
 &= \mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) = \mathbf{A}_1 \mathbf{V}^{**}.
 \end{aligned}$$

Por tanto $\mathbf{A}_1 \mathbf{V}^{**}$ es idempotente y se tiene que:

$$\mathbf{Y}^{**\top} \mathbf{A}_1 \mathbf{Y}^{**} = \frac{\mathbf{Y}^\top \mathbf{A}_1 \mathbf{Y}}{\sigma_e^2} = \frac{SCE}{\sigma_e^2} \sim \chi^2(\text{rango}(\mathbf{A}_1), \lambda_1),$$

con:

$$\begin{aligned}
 \text{rango}(\mathbf{A}_1) &= \text{rango}(\mathbf{I}_{ar} - \mathbf{P}_x) = \text{rango} \left(\mathbf{I}_{ar} - \frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) \\
 &= a(r-1) \\
 \lambda_1 &= \frac{1}{2} \boldsymbol{\mu}^{**\top} \mathbf{A}_1 \boldsymbol{\mu}^{**} = \frac{1}{2} \frac{\mu^2}{\sigma_e^2} \mathbf{1}_{ar}^\top \left(\mathbf{I}_{ar} - \frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) \underbrace{\mathbf{1}_{ar}}_{\mathbf{1}_a \otimes \mathbf{1}_r} \\
 &= \frac{1}{2} \frac{\mu^2}{\sigma_e^2} \left[\underbrace{\mathbf{1}_{ar}^\top \mathbf{I}_{ar} \mathbf{1}_{ar}}_{ar} - \frac{1}{r} \underbrace{(\mathbf{1}_a \otimes \mathbf{1}_r)^\top (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) (\mathbf{1}_a \otimes \mathbf{1}_r)}_{\mathbf{1}_a^\top \mathbf{I}_a \mathbf{1}_a \otimes \mathbf{1}_r^\top \mathbf{1}_r \mathbf{1}_r^\top \mathbf{1}_r} \right] = 0, \tag{B.34}
 \end{aligned}$$

luego:

$$\frac{SCE}{\sigma_e^2} \sim \chi^2(a[r-1]).$$

6.1.3. Esperanza de las sumas de cuadrados

Usando $\mathbf{A}\mathbf{V}$ dado en (B.31) y que $\boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu} = 0$ dado en (B.32):

$$\begin{aligned}\mathbb{E}(SCM) &= \mathbb{E} \left(\mathbf{Y}^\top \underbrace{\left(\mathbf{P}_x - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right)}_{\mathbf{A}} \mathbf{Y} \right) \\ &= \text{traza}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu} \\ &= \text{traza} \left[\left(r\sigma_\tau^2 + \sigma_e^2 \right) \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \right] + 0 \\ &= \frac{r\sigma_\tau^2 + \sigma_e^2}{r} [ar - r] = (r\sigma_\tau^2 + \sigma_e^2) (a - 1) .\end{aligned}$$

Usando $\mathbf{A}_1\mathbf{V}$ dado en (B.33) y que $\boldsymbol{\mu}^\top \mathbf{A}_1\boldsymbol{\mu} = 0$ dado en (B.34):

$$\begin{aligned}\mathbb{E}(SCE) &= \mathbb{E} \left(\mathbf{Y}^\top (\mathbf{I}_{ar} - \mathbf{P}_x) \mathbf{Y} \right) = \text{traza}(\mathbf{A}_1\mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A}_1\boldsymbol{\mu} \\ &= \sigma_e^2 \text{traza} \left(\mathbf{I}_{ar} - \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \right) + 0 \\ &= \sigma_e^2 a(r - 1) .\end{aligned}$$

6.2. Cuadrados medios

Los cuadrados medios, CM , se obtienen al dividir la suma de cuadrados entre los grados de libertad, ver tabla B.10. La esperanza de los cuadrados medios del modelo y del error es dada por:

$$\begin{aligned}\mathbb{E}(CMM) &= \mathbb{E} \left(\frac{SCM}{glm} \right) = \mathbb{E} \left(\frac{SCM}{a - 1} \right) = \frac{(r\sigma_\tau^2 + \sigma_e^2) (a - 1)}{a - 1} \\ &= r\sigma_\tau^2 + \sigma_e^2 \tag{B.35} \\ \mathbb{E}(CME) &= \mathbb{E} \left(\frac{SCE}{gle} \right) = \mathbb{E} \left(\frac{SCE}{a(r - 1)} \right) = \frac{\sigma_e^2 a(r - 1)}{a(r - 1)} = \sigma_e^2 .\end{aligned}$$

6.2.1. Cociente de cuadrados medios

$$\frac{\frac{\frac{SCM}{r\sigma_\tau^2 + \sigma_e^2}}{a - 1}}{\frac{\frac{SCE}{\sigma_e^2}}{a(r - 1)}} = \frac{\frac{1}{r\sigma_\tau^2 + \sigma_e^2} \times \frac{SCM}{a - 1}}{\frac{1}{\sigma_e^2} \times \frac{SCE}{a(r - 1)}} = \frac{\sigma_e^2}{r\sigma_\tau^2 + \sigma_e^2} \frac{CMM}{CME} .$$

Usando que $\mathbf{AV} = (r\sigma_\tau^2 + \sigma_e^2) \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top)$, ver (B.31), se tiene que:

$$\begin{aligned} \mathbf{AV} \mathbf{A}_1 &= (r\sigma_\tau^2 + \sigma_e^2) \frac{1}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) (\mathbf{I}_{ar} - \mathbf{P}_x) \\ &= \frac{r\sigma_\tau^2 + \sigma_e^2}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \left(\mathbf{I}_{ar} - \frac{1}{r} \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top \right) \\ &= \frac{r\sigma_\tau^2 + \sigma_e^2}{r} \left(\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top - \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top - \mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top + \frac{1}{a} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) = 0, \end{aligned}$$

por tanto:

$$\frac{\sigma_e^2}{r\sigma_\tau^2 + \sigma_e^2} \frac{CMM}{CME} \sim F(a-1, a[r-1]). \quad (\text{B.36})$$

La tabla B.10 presenta de forma reducida los resultados obtenidos de las esperanzas de los cuadrados medios y su distribución.

Tabla B.10. Análisis de varianza incluyendo esperanza de los cuadrados medios

Causa de variación	Grados de libertad	Suma de cuadrados	Cuadrado medio	$\mathbb{E}(CM)$	
Tratam.	$a-1$	$\mathbf{Y}^\top \mathbf{P}_x \mathbf{Y} - n\bar{y}^2$	$\frac{SCM}{a-1}$	$r\sigma_\tau^2 + \sigma_e^2$	$\frac{SCM}{r\sigma_\tau^2 + \sigma_e^2} \sim \chi^2(a-1)$
Error	$a(r-1)$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_x) \mathbf{Y}$	$\frac{SCE}{a(r-1)}$	σ_e^2	$\frac{SCE}{\sigma_e^2} \sim \chi^2(a[r-1])$
Total	$ar-1$	$\mathbf{Y}^\top \mathbf{Y} - n\bar{y}^2$			

Solo si $\sigma_\tau^2 = 0$ se tiene que:

$$\frac{CMM}{CME} \sim F(a-1, a[r-1]),$$

esto es, bajo la hipótesis nula:

$$H_0 : \sigma_\tau^2 = 0,$$

el cociente:

$$\frac{CMM}{CME} \sim F(a-1, a[r-1]).$$

7. Construcción envelope

En la construcción del envelope se busca generar la distribución empírica del residuo. Una vez realizado esto, se grafican $t_{s_k}^*$ contra los percentiles de la distribución normal estándar. El procedimiento para construir estas bandas consiste en:

1. Definir m , el número de muestras a generar (simulaciones) para la construcción de las bandas del envelope, por ejemplo $m = 100$.
2. Generar m vectores aleatorios n -dimensionales de la distribución $N_n(\mathbf{0}, \mathbf{I}_n)$, que corresponderan a los errores (e_j), $j = 1, 2, 3, \dots, m$. Por tanto en $\sigma_e^2 = 1$ en las simulaciones.
3. Para cada e_j generado, calcular los residuales bajo el modelo que se quiere revisar:

$$\begin{aligned}
 r_j &= Y_j - \hat{Y} = (\mathbf{X}\beta + e_j) - \mathbf{X}\hat{\beta} \\
 &= \mathbf{X}\beta + e_j - \mathbf{X} \left\{ \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top Y_j + \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{X} \right] z \right\} \\
 &= \mathbf{X}\beta + e_j - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top (\mathbf{X}\beta + e_j) + \\
 &\quad \underbrace{\mathbf{X} \left\{ \left[\mathbf{I} - \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{X} \right] z \right\}}_0 \\
 &= \cancel{\mathbf{X}\beta} + e_j - \underbrace{\mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \cancel{\mathbf{X}\beta}}_{\mathbf{X}} + \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top e_j \\
 &= e_j - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top e_j = (\mathbf{I} - \mathbf{P}_x) e_j,
 \end{aligned}$$

con $j = 1, 2, 3, \dots, m$.

4. Para cada r_j , $j = 1, 2, 3, \dots, m$, se calculan los residuos estandarizados, t_{r_i} , teniendo en cuenta que $\mathbb{V}(Y_j) = \mathbb{V}(\mathbf{X}\beta + e_j) = \mathbb{V}(e_j) = \underbrace{\sigma_e^2}_1 \mathbf{I} = \mathbf{I}$:

$$t_{r_{ij}} = \frac{r_{ij}}{\sigma \sqrt{1 - h_{ii}}} = \frac{r_{ij}}{\sqrt{1 - h_{ii}}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m,$$

y se ordenan de menor a mayor, esto es, $t_{r(1)j} \leq t_{r(2)j} \leq \dots \leq t_{r(n)j}$, ver tabla B.11.

Tabla B.11. Organización de los residuales estandarizados de las m simulaciones

Simulación						
1	$t_{r(1)1}$	$\leq$	$t_{r(2)1}$	$\leq$	$\dots$	$\leq t_{r(n)1}$
2	$t_{r(1)2}$	$\leq$	$t_{r(2)2}$	$\leq$	$\dots$	$\leq t_{r(n)2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	$t_{r(1)m}$	$\leq$	$t_{r(2)m}$	$\leq$	$\dots$	$\leq t_{r(n)m}$
Percentil 5:	$P_{5(1)}$		$P_{5(2)}$		$\dots$	$P_{5(n)}$
Percentil 95:	$P_{95(1)}$		$P_{95(2)}$		$\dots$	$P_{95(n)}$

5. En cada columna de la tabla B.11, calcular los percentiles 5 y 95. Así se tendran n percentiles 5 y n percentiles 95, y construir sobre un mismo gráfico un Q-Q plot para los n percentiles 5 y un Q-Q plot para los n percentiles 95.
6. Sobre el gráfico anterior, realizar un Q-Q plot para los residuales estandarizados originales, ver (3.15), esto es, los obtenidos de los datos observados, $t_{s_i}^*$.

8. Análisis de varianza - polinomios ortogonales

8.1. Sumas de cuadrados de γ_l , $l = 1, 2, \dots, a - 1$

De (3.30) se tiene:

$$SC(\hat{\gamma}_l) = r\hat{\gamma}_l^2,$$

luego:

$$SC(\hat{\gamma}_l) = \underbrace{\mathbf{Y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}_{\mathbf{A}}.$$

Para encontrar la distribución de la suma de cuadrados de γ_l , note que:

- La matriz $\mathbf{A} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ es simétrica.
- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ con $\mathbf{V} = \sigma_e^2 \mathbf{I}$, sea $\mathbf{Y}^* = \frac{1}{\sigma_e} \mathbf{Y}$, por tanto:

$$\begin{aligned} \boldsymbol{\mu}^* &= \mathbb{E}(\mathbf{Y}^*) = \frac{1}{\sigma_e} (\mathbf{X}\boldsymbol{\beta}) = \frac{1}{\sigma_e} \boldsymbol{\mu} \quad y \\ \mathbf{V}^* &= \mathbb{V}(\mathbf{Y}^*) = \frac{1}{\sigma_e^2} \mathbf{V} = \mathbf{I}, \end{aligned} \tag{B.37}$$

es decir, $\frac{1}{\sigma_e} \mathbf{Y} \sim N\left(\frac{1}{\sigma_e} \boldsymbol{\mu}, \mathbf{I}\right)$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{A}\mathbf{V}^*$:

$$\begin{aligned} \mathbf{A}\mathbf{V}^* \mathbf{A}\mathbf{V}^* &= \\ \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top &= \\ \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top &= \\ = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top &= \\ = \mathbf{A}\mathbf{V}^*. \end{aligned}$$

Como $\mathbf{AV}^*$ es idempotente se tiene que:

$$\frac{SC(\hat{\gamma}_l)}{\sigma_e^2} \sim \chi^2(\text{rango}(\mathbf{A}), \lambda)$$

en que:

$$\begin{aligned} \text{rango}(\mathbf{A}) &= \text{rango}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \\ &= \text{rango} \left\{ \underbrace{\boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l}_{\frac{1}{r}} \left[\frac{1}{r} \right]^{-1} \right\} = 1 \end{aligned}$$

y

$$\begin{aligned} \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* \\ &= \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} \\ &= \frac{r}{2\sigma_e^2} \boldsymbol{\beta}^\top \boldsymbol{\lambda}_l \boldsymbol{\lambda}_l^\top \boldsymbol{\beta} = \frac{r\gamma_l^2}{2\sigma_e^2}, \end{aligned}$$

luego:

$$\frac{SC(\hat{\gamma}_l)}{\sigma_e^2} \sim \chi^2 \left(1, \frac{r\gamma_l^2}{2\sigma_e^2} \right).$$

8.1.1. Esperanza de la suma de cuadrados de γ_l , $l = 1, \dots, a - 1$

$$\begin{aligned} \mathbb{E}(SC(\hat{\gamma}_l)) &= \mathbb{E} \left\{ \mathbf{Y}^\top \underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{\lambda}_l \left[\frac{1}{r} \right]^{-1} \boldsymbol{\lambda}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{A}} \mathbf{Y} \right\} \\ &= \text{traza}(\mathbf{AV}) + \underbrace{\boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}}_1 = \sigma_e^2 \text{traza}(\mathbf{A}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} = \sigma_e^2 + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} \\ &= \sigma_e^2 + r\gamma_l^2. \end{aligned}$$

8.2. Suma de cuadrados del error (SCE)

Sea $\mathbf{A}_2 = \mathbf{I} - \mathbf{P}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$, y sea:

$$SCE = \mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_X) \mathbf{Y} = \mathbf{Y}^\top \mathbf{A}_2 \mathbf{Y}.$$

Como $I - P_X$ tiene la misma estructura de la sección 2.1.2, se tiene:

$$\frac{SCE}{\sigma_e^2} \sim \chi_{a(r-1)}^2.$$

Por otro lado:

$$\begin{aligned} AV^*A_2 &= AA_2 = A - AP_X \\ &= A - X(X^\top X)^{-1}\lambda_l \left[\frac{1}{r} \right]^{-1} \lambda_l^\top \cancel{(X^\top X)^{-1} X^\top X} (X^\top X)^{-1} X^\top \\ &= A - A = 0, \end{aligned}$$

lo cual implica independencia entre SCE y $SC(\hat{\gamma}_l)$, y

$$\frac{\frac{SC(\hat{\gamma}_l)}{\cancel{\sigma_e^2}}}{\frac{CME}{\cancel{\sigma_e^2}}} = \frac{SC(\hat{\gamma}_l)}{CME} \sim F(1, a(r-1), \lambda),$$

con:

$$\lambda = \frac{r\gamma_l^2}{2\sigma_e^2}.$$

9. Análisis de varianza - modelo con una co-variable

Se considera el modelo (3.33), sin considerar ninguna restricción, escrito matricialmente:

$$Y = X\beta + E,$$

con $\mathbf{Y}$ y $\mathbf{E}$ definidos en (3.34):

$$\boldsymbol{\beta}_{(a+2) \times 1} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_a \\ \gamma \end{pmatrix} \quad \text{y} \quad \mathbf{X}_{n \times (a+2)} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & x_{11} - \bar{x} \\ 1 & 1 & 0 & \dots & 0 & x_{12} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 0 & \dots & 0 & x_{1r} - \bar{x} \\ 1 & 0 & 1 & \dots & 0 & x_{21} - \bar{x} \\ 1 & 0 & 1 & \dots & 0 & x_{22} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 1 & \dots & 0 & x_{2r} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{a1} - \bar{x} \\ 1 & 0 & 0 & \dots & 1 & x_{a2} - \bar{x} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{ar} - \bar{x} \end{pmatrix}.$$

La matriz diseño, $\mathbf{X}$, se puede escribir como una matriz particionada en que una parte corresponde a los efectos de tratamiento, $\mathbf{X}_1$, y la otra parte corresponde al efecto de la covariable, $\mathbf{X}_2$, a $\mathbf{X}$ se renombra como $\mathbf{X}_{12}$, es decir:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 1 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} x_{11} - \bar{x} \\ x_{12} - \bar{x} \\ \vdots \\ x_{1r} - \bar{x} \\ \vdots \\ x_{a1} - \bar{x} \\ \vdots \\ x_{ar} - \bar{x} \end{pmatrix} \quad \text{y} \quad \mathbf{X} = \mathbf{X}_{12} = (\mathbf{X}_1 | \mathbf{X}_2). \quad (\text{B.38})$$

De forma reducida $\mathbf{X}_1$ y $\mathbf{X}_2$ se pueden escribir:

$$\mathbf{X}_1 = (\mathbf{1}_{ar} | \mathbf{I}_a \otimes \mathbf{1}_r)_{ar \times (a+1)} \quad \text{y} \quad \mathbf{X}_2 = (\mathbf{x} - \bar{x} \mathbf{1}_{ar})_{ar \times 1},$$

con $\mathbf{x} = (x_{11} \ x_{12} \ \dots \ x_{1r} \ x_{21} \ \dots \ x_{a1} \ \dots \ x_{ar})^\top$ y $\bar{x}$ definida en (3.33).

9.1. Distribución y esperanza de las sumas de cuadrados

Para el calculo de la distribución y esperanza de las sumas de cuadrados, se usan las matrices de proyección asociadas a las matrices $\mathbf{X}_1$, $\mathbf{X}_2$ y $\mathbf{X}_{12}$:

$$\begin{aligned}
 P_1 &= \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \\
 &= (\mathbf{1}_{ar} | \mathbf{I}_a \otimes \mathbf{1}_r) \left[\begin{pmatrix} ar & r\mathbf{1}_a^\top \\ r\mathbf{1}_a & r\mathbf{I}_a \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{1}_{ar}^\top \\ \mathbf{I}_a \otimes \mathbf{1}_r^\top \end{pmatrix} \\
 &= \frac{1}{r} (\mathbf{1}_{ar} | \mathbf{I}_a \otimes \mathbf{1}_r) \begin{pmatrix} 0 & \mathbf{0}_a^\top \\ \mathbf{0}_a & \mathbf{I}_a \end{pmatrix} \begin{pmatrix} \mathbf{1}_{ar}^\top \\ \mathbf{I}_a \otimes \mathbf{1}_r^\top \end{pmatrix} = \frac{1}{r} (\mathbf{I}_a \otimes \mathbf{1}_r \mathbf{1}_r^\top) \\
 P_2 &= \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top = \mathbf{X}_2 \left(\sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x})^2 \right)^{-1} \mathbf{X}_2^\top \\
 &= \frac{1}{E_{xx}} \begin{pmatrix} (x_{11} - \bar{x})(x_{11} - \bar{x}) & (x_{11} - \bar{x})(x_{12} - \bar{x}) & \dots & (x_{11} - \bar{x})(x_{ar} - \bar{x}) \\ (x_{12} - \bar{x})(x_{11} - \bar{x}) & (x_{12} - \bar{x})(x_{12} - \bar{x}) & \dots & (x_{12} - \bar{x})(x_{ar} - \bar{x}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{ar} - \bar{x})(x_{11} - \bar{x}) & (x_{ar} - \bar{x})(x_{12} - \bar{x}) & \dots & (x_{ar} - \bar{x})(x_{ar} - \bar{x}) \end{pmatrix} \\
 P_{12} &= \mathbf{X}_{12}(\mathbf{X}_{12}^\top \mathbf{X}_{12})^{-1} \mathbf{X}_{12}^\top = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top,
 \end{aligned} \tag{B.39}$$

en que:

$$\begin{aligned}
 E_{xx} &= \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x})^2 = \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*} + \bar{x}_{i*} - \bar{x})^2 \\
 &= \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2 + 2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})(\bar{x}_{i*} - \bar{x}) + \sum_{i=1}^a \sum_{j=1}^r (\bar{x}_{i*} - \bar{x})^2 \\
 &= \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2 + r \sum_{i=1}^a (\bar{x}_{i*} - \bar{x})^2.
 \end{aligned} \tag{B.40}$$

Para encontrar P_{12} , resolviendo por partes, se tiene que:

$$\begin{aligned} \mathbf{X}_{12}^\top \mathbf{X}_{12} &= \begin{pmatrix} ar & r\mathbf{1}_a^\top & 0 \\ r\mathbf{1}_a & r\mathbf{I}_a & (\mathbf{I}_a \otimes \mathbf{1}_r^\top)(\mathbf{x} - \bar{x}\mathbf{1}_{ar}) \\ 0 & (\mathbf{x} - \bar{x}\mathbf{1}_{ar})^\top (\mathbf{I}_a \otimes \mathbf{1}_r) & \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x})^2 \end{pmatrix} \\ &= r \begin{pmatrix} a & \mathbf{1}_a^\top & 0 \\ \mathbf{1}_a & \mathbf{I}_a & \bar{x} - \bar{x}\mathbf{1}_a \\ 0 & (\bar{x} - \bar{x}\mathbf{1}_a)^\top & \frac{E_{xx}}{r} \end{pmatrix}, \end{aligned}$$

con $\bar{x} = (\bar{x}_{1*} \quad \bar{x}_{2*} \quad \dots \quad \bar{x}_{a*})^\top$ en que $\bar{x}_{i*} = \frac{1}{r} \sum_{j=1}^r x_{ij}$, luego:

$$(\mathbf{X}_{12}^\top \mathbf{X}_{12})^{-1} = \frac{1}{r} \begin{pmatrix} a & \mathbf{1}_a^\top & 0 \\ \mathbf{1}_a & \mathbf{I}_a & \bar{x} - \bar{x}\mathbf{1}_a \\ 0 & (\bar{x} - \bar{x}\mathbf{1}_a)^\top & \frac{E_{xx}}{r} \end{pmatrix}^{-1}.$$

Nota B.4. Inversa de matrices particionadas. Sea $\mathbf{A}_{m \times m}$, $\mathbf{W}_{m \times n}$, $\mathbf{V}_{n \times m}$ y $\mathbf{D}_{n \times n}$. Suponga que $\mathbf{A}$ es no singular. Entonces: $\begin{pmatrix} \mathbf{A} & \mathbf{W} \\ \mathbf{V} & \mathbf{D} \end{pmatrix}$ es no singular si y solo si $\mathbf{Q} = \mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W}$ es no singular [30, p. 98], y

$$\begin{aligned} \begin{pmatrix} \mathbf{A} & \mathbf{W} \\ \mathbf{V} & \mathbf{D} \end{pmatrix}^{-1} &= \\ &\begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{W}(\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W})^{-1}\mathbf{V}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{W}(\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W})^{-1} \\ -(\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W})^{-1}\mathbf{V}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W})^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} -\mathbf{A}^{-1}\mathbf{W} \\ \mathbf{I}_n \end{pmatrix} (\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W})^{-1} \begin{pmatrix} -\mathbf{V}\mathbf{A}^{-1} & \mathbf{I}_n \end{pmatrix}. \end{aligned}$$

Usando la nota B.4, sea $\mathbf{A} = \mathbf{I}_a$, $\mathbf{W} = \bar{x} - \bar{x}\mathbf{1}_a$, $\mathbf{V} = \mathbf{W}^\top$ y $\mathbf{D} = \frac{E_{xx}}{r}$, se tiene que:

$$\begin{aligned} \mathbf{Q} = \mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{W} &= \frac{1}{r}E_{xx} - (\bar{x} - \bar{x}\mathbf{1}_a)^\top \mathbf{I}_a (\bar{x} - \bar{x}\mathbf{1}_a) \\ &= \frac{1}{r}E_{xx} - \sum_{i=1}^a (\bar{x}_{i*} - \bar{x})^2 = A^*. \end{aligned} \quad (\text{B.41})$$

Reemplazando (B.40) en (B.41) se tiene que:

$$\begin{aligned}
 A^* &= \frac{1}{r} E_{xx} - \sum_{i=1}^a (\bar{x}_{i*} - \bar{x})^2 \\
 &= \frac{1}{r} \left[\sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2 + r \sum_{i=1}^a (\bar{x}_{i*} - \bar{x})^2 \right] - \sum_{i=1}^a (\bar{x}_{i*} - \bar{x})^2 \\
 A^* &= \frac{1}{r} \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2. \tag{B.42}
 \end{aligned}$$

Retomando:

$$\begin{aligned}
 \begin{pmatrix} \mathbf{A} & \mathbf{W} \\ \mathbf{V} & \mathbf{D} \end{pmatrix}^{-1} &= \begin{pmatrix} \mathbf{I}_a & \bar{\mathbf{x}} - \bar{x} \mathbf{1}_a \\ \bar{\mathbf{x}}^\top - \bar{x} \mathbf{1}_a^\top & \frac{1}{r} E_{xx} \end{pmatrix}^{-1} \\
 &= \begin{pmatrix} \mathbf{I}_a & \mathbf{0}_{a \times 1} \\ \mathbf{0}_{1 \times a} & 0 \end{pmatrix} + \begin{pmatrix} -(\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) \\ 1 \end{pmatrix} \frac{1}{A^*} \begin{pmatrix} -(\bar{\mathbf{x}}^\top - \bar{x} \mathbf{1}_a^\top) & 1 \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{I}_a + \frac{1}{A^*} (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & -\frac{1}{A^*} (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) \\ -\frac{1}{A^*} (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & \frac{1}{A^*} \end{pmatrix}.
 \end{aligned}$$

Por tanto:

$$\left(\mathbf{X}_{12}^\top \mathbf{X}_{12} \right)^- = \frac{1}{r A^*} \begin{pmatrix} 0 & \mathbf{0}_{1 \times a} & 0 \\ \mathbf{0}_{a \times 1} & A^* \mathbf{I}_a + (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & -(\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) \\ 0 & -(\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & 1 \end{pmatrix}$$

y

$$\begin{aligned}
 \mathbf{P}_{12} &= \mathbf{X}_{12} \left(\mathbf{X}_{12}^\top \mathbf{X}_{12} \right)^- \mathbf{X}_{12}^\top = [(\mathbf{1}_{ar} | \mathbf{I}_a \otimes \mathbf{1}_r) | (\mathbf{x} - \bar{x} \mathbf{1}_{ar})] \times \\
 &\quad \frac{1}{r A^*} \begin{pmatrix} 0 & \mathbf{0}_{1 \times a} & 0 \\ \mathbf{0}_{a \times 1} & A^* \mathbf{I}_a + (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) (\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & -(\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a) \\ 0 & -(\bar{\mathbf{x}} - \bar{x} \mathbf{1}_a)^\top & 1 \end{pmatrix} \times \\
 &\quad [(\mathbf{1}_{ar} | \mathbf{I}_a \otimes \mathbf{1}_r) | (\mathbf{x} - \bar{x} \mathbf{1}_{ar})]^\top,
 \end{aligned}$$

con A^* definida en (B.42).

9.1.1. Suma de cuadrados asociada a los tratamientos

Para definir la suma de cuadrados asociada a los tratamientos se usará la tabla 2.3 en que se presentó el análisis de varianza para un modelo en dos partes:

$$SC(\mathbf{X}_1 | \mathbf{X}_2) = SCTrat = \mathbf{Y}^\top \underbrace{\left(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right)}_{\mathbf{A}} \mathbf{Y} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y},$$

esto es:

$$\mathbf{A} = \mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top. \quad (\text{B.43})$$

Para encontrar la distribución de $SCTrat$ note que:

- La matriz $\mathbf{A}$ es simétrica.
- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma_e^2 \mathbf{I}_n)$ se tiene $\mathbf{V} = \sigma_e^2 \mathbf{I}_n$. Sea:

$$\mathbf{Y}^* = \frac{1}{\sigma_e} \mathbf{Y} \sim N\left(\frac{1}{\sigma_e} \boldsymbol{\mu}, \mathbf{I}_n\right),$$

entonces $\mathbf{V}^* = \mathbf{I}$ y $\boldsymbol{\mu}^* = \frac{1}{\sigma_e} \boldsymbol{\mu}$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{A} \mathbf{V}^*$ (ejercicio 9.1).

Una vez verificada la idempotencia y la simetría se tiene que:

$$\frac{SCTrat}{\sigma_e^2} \sim \chi^2(\text{rango}(\mathbf{A}), \lambda),$$

con:

$$\begin{aligned} \text{rango}(\mathbf{A}) &= \text{rango}\left(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top\right) \\ &= \text{traza}\left(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top\right) \\ &= \text{traza}(\mathbf{P}_{12}) - \text{traza}(\mathbf{P}_2) - \underbrace{\text{traza}\left(\frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top\right)}_{\frac{ar}{ar}} \\ &= \text{traza}(\mathbf{X}_{12}(\mathbf{X}_{12}^\top \mathbf{X}_{12})^{-1} \mathbf{X}_{12}^\top) - \underbrace{\text{traza}(\mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top)}_{\text{traza}((\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_2)} - 1 \end{aligned}$$

Deducción del modelo DCA

$$= \text{traza}(\mathbf{X}_{12}(\mathbf{X}_{12}^\top \mathbf{X}_{12})^{-1} \mathbf{X}_{12}^\top) - \underbrace{\text{traza}(\mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top)}_{\text{traza}((\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_2)} - 1$$

$$\stackrel{\text{ejercicio 9.2}}{=} a + 1 - 1 - 1 = a - 1$$

$$\begin{aligned} \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A} \boldsymbol{\mu}^* = \frac{1}{2\sigma_e^2} \boldsymbol{\mu}^\top \left(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \right) \boldsymbol{\mu} \\ &= \frac{1}{2\sigma_e^2} (\boldsymbol{\mu}^\top \mathbf{P}_{12} \boldsymbol{\mu}) - \frac{1}{2\sigma_e^2} (\boldsymbol{\mu}^\top \mathbf{P}_2 \boldsymbol{\mu}) - \frac{1}{2\sigma_e^2} \left(\boldsymbol{\mu}^\top \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \boldsymbol{\mu} \right) \\ &\stackrel{\text{ejercicio 9.3}}{=} \frac{r}{2\sigma_e^2} \left[\sum_{i=1}^a \tau_i^2 - \frac{r}{E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2 \right], \end{aligned}$$

luego:

$$\frac{SCTrat}{\sigma_e^2} \sim \chi^2 \left(a - 1, \frac{r}{2\sigma_e^2} \left[\sum_{i=1}^a \tau_i^2 - \frac{r}{E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2 \right] \right).$$

Por otro lado:

$$\begin{aligned} \mathbb{E}(SCTrat) &= \text{traza}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} = \sigma_e^2(a - 1) + \\ &\quad r \left[\sum_{i=1}^a \tau_i^2 - \frac{r}{E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2 \right]. \end{aligned}$$

9.1.2. Suma de cuadrados asociada a la regresión

Para definir la suma de cuadrados asociada a la regresión se usará la tabla 2.3, en que se presentó el análisis de varianza para un modelo en dos partes.

$$SC(\mathbf{X}_2 | \mathbf{X}_1) = SCReg = \mathbf{Y}^\top \underbrace{(\mathbf{P}_{12} - \mathbf{P}_1)}_{\mathbf{A}_R} \mathbf{Y},$$

esto es:

$$\mathbf{A}_R = \mathbf{P}_{12} - \mathbf{P}_1. \quad (\text{B.44})$$

Para encontrar la distribución de $SCReg$ note que:

- La matriz $\mathbf{A}_R$ es simétrica.
- $\mathbf{A}_R \mathbf{V}^*$ es idempotente (ejercicio 9.4).

Teniendo la idempotencia y la simetría se tiene que:

$$\frac{SCReg}{\sigma_e^2} \sim \chi^2(rango(\mathbf{A}_R), \lambda_R),$$

con:

$$\begin{aligned} rango(\mathbf{A}_R) &\stackrel{\text{ejercicio 9.4}}{=} 1 \\ \lambda_R &= \frac{1}{2} \boldsymbol{\mu}^{*\top} \mathbf{A}_R \boldsymbol{\mu}^* = \frac{1}{2\sigma_e^2} \boldsymbol{\mu}^\top [\mathbf{P}_{12} - \mathbf{P}_1] \boldsymbol{\mu} = \frac{1}{2\sigma_e^2} \boldsymbol{\mu}^\top \mathbf{P}_{12} \boldsymbol{\mu} - \frac{1}{2\sigma_e^2} \boldsymbol{\mu}^\top \mathbf{P}_1 \boldsymbol{\mu} \\ &\stackrel{\text{ejercicio 9.5}}{=} \frac{1}{2\sigma_e^2} \gamma^2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2. \end{aligned}$$

Por otro lado,

$$\mathbb{E}(SCReg) = traza(\mathbf{A}_R \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A}_R \boldsymbol{\mu} = \sigma_e^2 + \gamma^2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2.$$

La tabla B.12 presenta de forma reducida los resultados obtenidos de las esperanzas de los cuadrados medios y su distribución [33, p. 249].

Finalmente, si $\lambda = 0$, es decir si:

$$\sum_{i=1}^a \tau_i^2 - \frac{r}{E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2 = 0,$$

se tiene que el cociente:

$$\frac{\frac{SC(\mathbf{X}_1|\mathbf{X}_2)}{\sigma_e^2(a-1)}}{\frac{SCE}{\sigma_e^2(ar-a-1)}} = \frac{CMTrat}{CME} \sim F_{a-1, ar-a-1}.$$

Tabla B.12. Análisis de varianza incluyendo esperanza de los cuadrados medios

Causa de variación	Grados de libertad	Sumas de cuadrados	Cuadr. medio	$\mathbb{E}(CM)$	
Regr. $\mathbf{X}_2 \mathbf{X}_1$	1	$\mathbf{Y}^\top \mathbf{P}_{12} \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_1 \mathbf{Y}$	$SCReg$	$\sigma_e^2 + \gamma^2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2$	$\frac{SCReg}{\sigma_e^2} \sim \chi_{1,\lambda_R}^2$
Tratam. $\mathbf{X}_1 \mathbf{X}_2$	$a - 1$	$\mathbf{Y}^\top \mathbf{P}_{12} \mathbf{Y} - \mathbf{Y}^\top \mathbf{P}_2 \mathbf{Y} - \frac{1}{ar} \mathbf{Y}^\top \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \mathbf{Y}$	$\frac{SCTrat}{a-1}$	$\sigma_e^2 + \frac{r}{a-1} \sum_{i=1}^a \tau_i^2 - \frac{r^2}{(a-1)E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2$	$\frac{SCTrat}{\sigma_e^2} \sim \chi_{a-1,\lambda}^2$
Error	$n - a - 1$	$\mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_{12}) \mathbf{Y}$	$\frac{SCE}{n-a-1}$	σ_e^2	
Total	$n - 1$	$\mathbf{Y}^\top \mathbf{Y} - \frac{1}{ar} \mathbf{Y}^\top \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \mathbf{Y}$			
con $\lambda_R = \frac{1}{2\sigma_e^2} \gamma^2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2$, $\lambda = \frac{r}{2\sigma_e^2} \left[\sum_{i=1}^a \tau_i^2 - \frac{r}{E_{xx}} \left(\sum_{i=1}^a \tau_i \bar{x}_{i*} \right)^2 \right]$ y $E_{xx} = \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x})^2$.					

En particular si $\tau_1 = \tau_2 = \dots = \tau_a = 0$, se tendrá $\lambda = 0$. Por otro lado, se tiene que el cociente:

$$\frac{\frac{SC(\mathbf{X}_2|\mathbf{X}_1)}{\sigma_e^2(1)}}{\frac{SCE}{\sigma_e^2(ar-a-1)}} = \frac{CMReg}{CME} \sim F_{1,ar-a-1},$$

solo si $\lambda_R = 0$, es decir:

$$\gamma^2 \sum_{i=1}^a \sum_{j=1}^r (x_{ij} - \bar{x}_{i*})^2 = 0.$$

Salvo el caso en que todos los x_{ij} son iguales a $\bar{x}_{i*}$, se tendrá que $\lambda_R = 0$ si $\gamma = 0$. Por tanto bajo la hipótesis nula:

$$H_0 : \gamma = 0,$$

se tiene que:

$$\frac{CMReg}{CME} \sim F_{1,ar-a-1}.$$

9.2. Ejercicios

9.1 Mostrar que $\mathbf{A}\mathbf{V}^*$ es idempotente en que $\mathbf{A}$ fue definida en (B.43) y $\mathbf{V}^* = \mathbf{I}$.

9.2 Mostrar que $\text{traza}(\mathbf{X}_{12}(\mathbf{X}_{12}^\top \mathbf{X}_{12})^{-1} \mathbf{X}_{12}^\top) = a + 1$ en que $\mathbf{X}_{12}$ fue definida en (B.38).

9.3 Teniendo en cuenta que $\mathbf{P}_1$, $\mathbf{P}_2$ y $\mathbf{P}_{12}$ fueron definidas en (B.39), encontrar:

- $\boldsymbol{\mu}^\top \mathbf{P}_{12} \boldsymbol{\mu}$.
- $\boldsymbol{\mu}^\top \mathbf{P}_2 \boldsymbol{\mu}$.
- $\frac{1}{ar} \boldsymbol{\mu}^\top \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \boldsymbol{\mu}$.
- $\mathbb{E}(\mathbf{Y}^\top \mathbf{P}_2 \mathbf{Y})$. ¿De qué parámetros depende?
- $\mathbf{P}_2(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top)$. ¿Hay independencia?
- Mostrar que $(\mathbf{P}_{12} - \mathbf{P}_2 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top)(\mathbf{I} - \mathbf{P}_{12}) = \mathbf{0}$.

9.4 Mostrar que $\mathbf{A}_R \mathbf{V}^*$ es idempotente, en que $\mathbf{A}_R$ fue definida en (B.44) y $\mathbf{V}^* = \mathbf{I}$, y mostrar que $\text{rango}(\mathbf{A}_R) = 1$.

9.5 Teniendo en cuenta que $\mathbf{P}_1$, $\mathbf{P}_2$ y $\mathbf{P}_{12}$ fueron definidas en (B.39):

- Encuentre $\boldsymbol{\mu}^\top \mathbf{P}_1 \boldsymbol{\mu}$.
- Mostrar que $(\mathbf{P}_{12} - \mathbf{P}_1)(\mathbf{I} - \mathbf{P}_{12}) = \mathbf{0}$.
- Encuentre $\mathbb{E}(\mathbf{Y}^\top \mathbf{P}_1 \mathbf{Y} - \frac{1}{ar} \mathbf{Y}^\top \mathbf{1}_{ar} \mathbf{1}_{ar}^\top \mathbf{Y})$. ¿De qué parámetros depende?
- Encuentre $(\mathbf{P}_1 - \frac{1}{ar} \mathbf{1}_{ar} \mathbf{1}_{ar}^\top)(\mathbf{P}_{12} - \mathbf{P}_1)$. ¿Hay independencia?



Apéndice C

Análisis de

varianza - diseño

anidado

1. Análisis de varianza - anidado - caso 1

Se parte del modelo (7.1) en que:

$$\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{y} \quad V(\mathbf{Y}) = \sigma_e^2 \mathbf{I}_{abr}.$$

Para la suma de cuadrados se usa el resultado presentado en la sección 1.1.

1.1. Hipótesis lineal general

Sean las hipótesis [56, p. 188]:

$$\begin{aligned} H_0 : \mathbf{K}^\top \boldsymbol{\beta} &= \mathbf{m} \quad \text{o} \quad H_0 : \mathbf{K}^\top \boldsymbol{\beta} - \mathbf{m} = \mathbf{0} \\ H_a : \mathbf{K}^\top \boldsymbol{\beta} &\neq \mathbf{m} \quad \text{o} \quad H_a : \mathbf{K}^\top \boldsymbol{\beta} - \mathbf{m} \neq \mathbf{0}, \end{aligned}$$

en que $\boldsymbol{\beta}$ es un vector de parámetros de orden p , $\mathbf{K}$ matriz conocida de orden $s \times p$ de rango s , y $\mathbf{m}$ vector de constantes conocidas de orden s .

Nota C.1. Como $\mathbf{K}^\top$ es de rango fila completo se tiene que las hipótesis deben ser funciones de $\boldsymbol{\beta}$ linealmente independientes.

Si $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_e^2 \mathbf{I}_n)$, las esperanza y varianza son dadas por:

$$\begin{aligned} \mathbb{E}(\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m}) &= \mathbf{K}^\top \mathbb{E}(\hat{\boldsymbol{\beta}}) - \mathbf{m} = \mathbf{K}^\top \boldsymbol{\beta} - \mathbf{m} \\ \mathbb{V}(\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m}) &= \mathbb{V}(\mathbf{K}^\top \hat{\boldsymbol{\beta}}) = \mathbf{K}^\top \mathbb{V}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}) \mathbf{K} \\ &= \mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{V}(\mathbf{e}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K} \\ &= \sigma_e^2 \mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}, \end{aligned}$$

y entonces:

$$\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m} \sim N(\mathbf{K}^\top \boldsymbol{\beta} - \mathbf{m}, \sigma_e^2 \mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}).$$

Nota C.2. La suma de cuadrados de $\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m}$ está dada por:

$$\begin{aligned} SC(\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m}) &= \\ (\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m})^\top &\left[\mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K} \right]^{-1} (\mathbf{K}^\top \hat{\boldsymbol{\beta}} - \mathbf{m}). \end{aligned} \quad (\text{C.1})$$

1.2. Distribución y esperanza de las sumas de cuadrados

Usando (C.1) con $\mathbf{m} = \mathbf{0}$, y las respectivas restricciones, se tiene la suma de cuadrados de $\mathbf{K}^\top \hat{\beta}$:

$$\begin{aligned} SC(\mathbf{K}^\top \hat{\beta}) &= \hat{\beta}^\top \mathbf{K} \left[\mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K} \right]^{-1} \mathbf{K}^\top \hat{\beta} \\ &= \mathbf{Y}^\top \underbrace{\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K} \left[\mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K} \right]^{-1} \mathbf{K}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}_{\mathbf{P}_K} \\ SC(\mathbf{K}^\top \hat{\beta}) &= \mathbf{Y}^\top \mathbf{P}_K \mathbf{Y}. \end{aligned} \quad (\text{C.2})$$

1.2.1. Sumas de cuadrados - factor A

Sea:

$$SCFA = \mathbf{Y}^\top \mathbf{P}_\alpha \mathbf{Y} = \mathbf{Y}^\top \mathbf{X}_\alpha \left(\mathbf{X}_\alpha^\top \mathbf{X}_\alpha \right)^{-1} \mathbf{X}_\alpha^\top \mathbf{Y}, \quad (\text{C.3})$$

con:

$$\mathbf{K}_\alpha^\top = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} = (\mathbf{0}_{a-1 \times 1} | \mathbf{I}_{a-1} | \mathbf{0}_{a-1 \times a(b-1)}), \quad (\text{C.4})$$

y $\mathbf{P}_\alpha = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \left[\mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \right]^{-1} \mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. Para encontrar la distribución de la suma de cuadrados asociada al factor A note que:

- La matriz $\mathbf{P}_\alpha$ es simétrica.
- Asumiendo que $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma_e^2 \mathbf{I}_{abr})$ se tiene $\mathbf{V} = \sigma_e^2 \mathbf{I}_{abr}$. Sea:

$$\mathbf{Y}^* = \frac{1}{\sigma_e} \mathbf{Y} \sim N\left(\frac{1}{\sigma_e} \boldsymbol{\mu}, \mathbf{I}_{abr}\right), \quad (\text{C.5})$$

entonces $\mathbf{V}^* = \mathbf{I}_{abr}$ y $\boldsymbol{\mu}^* = \frac{1}{\sigma_e} \boldsymbol{\mu} = \frac{1}{\sigma_e} \mathbf{X} \boldsymbol{\beta}$.

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $P_\alpha V^*$:

$$\begin{aligned}
 P_\alpha V^* P_\alpha V^* &= P_\alpha P_\alpha \\
 &= X(X^\top X)^{-1} K_\alpha \left[K_\alpha^\top (X^\top X)^{-1} K_\alpha \right]^{-1} K_\alpha^\top (X^\top X)^{-1} X^\top X \times \\
 &\quad (X^\top X)^{-1} K_\alpha \left[K_\alpha^\top (X^\top X)^{-1} K_\alpha \right]^{-1} K_\alpha^\top (X^\top X)^{-1} X \\
 &= X(X^\top X)^{-1} K_\alpha \left[K_\alpha^\top (X^\top X)^{-1} K_\alpha \right]^{-1} K_\alpha^\top (X^\top X)^{-1} X^\top = P_\alpha,
 \end{aligned}$$

por tanto $P_\alpha V^*$ es idempotente y se tiene que:

$$\frac{SCFA}{\sigma_e^2} \sim \chi^2(\text{rango}(P_\alpha), \lambda),$$

con:

$$\begin{aligned}
 \text{rango}(P_\alpha) &= \text{rango}(X_\alpha) = a - 1 \text{ y} \\
 \lambda &= \frac{1}{2} \mu^{*t} P_\alpha \mu^* \\
 &= \frac{1}{2\sigma_e^2} \beta^\top \cancel{X^\top X (X^\top X)^{-1}} K_\alpha \left[K_\alpha^\top (X^\top X)^{-1} K_\alpha \right]^{-1} K_\alpha^\top \times \\
 &\quad \cancel{(X^\top X)^{-1} X^\top X} \beta \\
 &= \frac{1}{2\sigma_e^2} \beta^\top K_\alpha \left[K_\alpha^\top (X^\top X)^{-1} K_\alpha \right]^{-1} K_\alpha^\top \beta.
 \end{aligned}$$

Notando que:

$$\beta^\top K_\alpha = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_{a-1} \\ \beta_{1(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{pmatrix}^\top \begin{pmatrix} \mathbf{0}_{1 \times a-1} \\ I_{a-1} \\ \mathbf{0}_{a(b-1) \times a-1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \dots & \alpha_{a-1} \end{pmatrix},$$

y que:

$$\begin{aligned} \mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha &= (\mathbf{0}_{a-1 \times 1} | \mathbf{I}_{a-1} | \mathbf{0}_{a-1 \times a(b-1)}) \times \\ &\left\{ \frac{1}{abr} \oplus \frac{1}{br} \left(\mathbf{I}_{a-1} - \frac{1}{a} \mathbf{J}_{a-1} \right) \oplus \left[\frac{1}{r} \mathbf{I}_a \otimes \left(\mathbf{I}_{b-1} - \frac{1}{b} \mathbf{J}_{b-1} \right) \right] \right\} \begin{pmatrix} \mathbf{0}_{1 \times a-1} \\ \mathbf{I}_{a-1} \\ \mathbf{0}_{a(b-1) \times a-1} \end{pmatrix} \\ &= \frac{1}{br} \left(\mathbf{I}_{a-1} - \frac{1}{a} \mathbf{J}_{a-1} \right), \end{aligned}$$

se tiene que:

$$\begin{aligned} \lambda &= \frac{1}{2\sigma_e^2} \begin{pmatrix} \alpha_1 & \dots & \alpha_{a-1} \end{pmatrix} \left[\frac{1}{br} \left(\mathbf{I}_{a-1} - \frac{1}{a} \mathbf{J}_{a-1} \right) \right]^{-1} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{a-1} \end{pmatrix} \\ &\stackrel{\text{Nota 7.1}}{=} \frac{1}{2\sigma_e^2} \begin{pmatrix} \alpha_1 & \dots & \alpha_{a-1} \end{pmatrix} [br (\mathbf{I}_{a-1} + \mathbf{J}_{a-1})] \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{a-1} \end{pmatrix} \\ &= \frac{br}{2\sigma_e^2} \left[\sum_{i=1}^{a-1} \alpha_i^2 + \left(\sum_{i=1}^{a-1} \alpha_i \right)^2 \right] = \frac{br}{2\sigma_e^2} \left[\sum_{i=1}^{a-1} \alpha_i^2 + (-\alpha_a)^2 \right] \\ \lambda &= \frac{br}{2\sigma_e^2} \sum_{i=1}^a \alpha_i^2. \end{aligned}$$

Por tanto:

$$\frac{SCFA}{\sigma_e^2} \sim \chi^2 \left(a-1, \frac{br}{2\sigma_e^2} \sum_{i=1}^a \alpha_i^2 \right),$$

y

$$\begin{aligned} \mathbb{E}(SCFA) &= \text{traza}(\mathbf{P}_\alpha \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{P}_\alpha \boldsymbol{\mu} \\ &= \sigma_e^2 \text{traza} \left[\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \left[\mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \right]^{-1} \mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right] + br \sum_{i=1}^a \alpha_i^2 \\ &= \sigma_e^2 \text{traza} \left[\cancel{\mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \left[\mathbf{K}_\alpha^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_\alpha \right]^{-1} \right] + br \sum_{i=1}^a \alpha_i^2 \\ &= \sigma_e^2 \text{traza} [\mathbf{I}_{a-1}] + br \sum_{i=1}^a \alpha_i^2 = \sigma_e^2 (a-1) + br \sum_{i=1}^a \alpha_i^2. \end{aligned}$$

1.2.2. Suma de cuadrados - factor $B(A)$

Sea:

$$SCFB(A) = \mathbf{Y}^\top \mathbf{P}_\beta \mathbf{Y} = \mathbf{Y}^\top \mathbf{X}_{\beta(A)} \left(\mathbf{X}_{\beta(A)}^\top \mathbf{X}_{\beta(A)} \right)^{-1} \mathbf{X}_{\beta(A)}^\top \mathbf{Y},$$

en que:

$$\mathbf{K}_{\beta(A)}^\top = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{pmatrix} = (\mathbf{0}_{a(b-1) \times a} | \mathbf{I}_{a(b-1)}),$$

y

$$\mathbf{P}_{\beta(A)} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top.$$

Para encontrar la distribución de la suma de cuadrados asociada al factor $B(A)$ note que:

- La matriz $\mathbf{P}_{\beta(A)}$ es simétrica.
- Sea $\mathbf{V}^*$ definida en (C.5).

Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $\mathbf{P}_{\beta(A)} \mathbf{V}^*$:

$$\begin{aligned} \mathbf{P}_{\beta(A)} \mathbf{V}^* \mathbf{P}_{\beta(A)} \mathbf{V}^* &= \mathbf{P}_{\beta(A)} \mathbf{P}_{\beta(A)} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \times \\ &\quad \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{P}_{\beta(A)}, \end{aligned}$$

por tanto $\mathbf{P}_{\beta(A)} \mathbf{V}^*$ es idempotente y

$$\frac{SCFB(A)}{\sigma_e^2} \sim \chi^2(\text{rango}(\mathbf{P}_{\beta(A)}), \lambda),$$

con:

$$\begin{aligned}
 \text{rango}(\mathbf{P}_{\beta(A)}) &= \text{rango}(\mathbf{X}_{\beta(A)}) = a(b-1) \text{ y} \\
 \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*t} \mathbf{P}_{\beta(A)} \boldsymbol{\mu}^* \\
 &= \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \times \\
 &\quad \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \bar{\mathbf{X}} \boldsymbol{\beta} \\
 &= \frac{1}{2\sigma_e^2} \boldsymbol{\beta}^\top \mathbf{K}_{\beta(A)} \left[\mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} \right]^{-1} \mathbf{K}_{\beta(A)}^\top \boldsymbol{\beta}.
 \end{aligned}$$

Notando que:

$$\boldsymbol{\beta}^\top \mathbf{K}_{\beta(A)} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_{a-1} \\ \beta_{1(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{pmatrix}^\top \begin{pmatrix} \mathbf{0}_{a \times a(b-1)} \\ \mathbf{I}_{a(b-1)} \end{pmatrix} = \begin{pmatrix} \beta_{1(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{pmatrix}$$

y

$$\begin{aligned}
 \mathbf{K}_{\beta(A)}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta(A)} &= \\
 (\mathbf{0}_{a(b-1) \times a} | \mathbf{I}_{a(b-1)}) &\times \\
 \left\{ \frac{1}{abr} \oplus \frac{1}{br} \left(\mathbf{I}_{a-1} - \frac{1}{a} \mathbf{J}_{a-1} \right) \oplus \left[\frac{1}{r} \mathbf{I}_a \otimes \left(\mathbf{I}_{b-1} - \frac{1}{b} \mathbf{J}_{b-1} \right) \right] \right\} &\begin{pmatrix} \mathbf{0}_{a \times a(b-1)} \\ \mathbf{I}_{a(b-1)} \end{pmatrix} \\
 = \frac{1}{r} \mathbf{I}_a \otimes \left(\mathbf{I}_{b-1} - \frac{1}{b} \mathbf{J}_{b-1} \right), &
 \end{aligned}$$

se tiene que:

$$\lambda = \frac{1}{2\sigma_e^2} \begin{pmatrix} \beta_{1(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{pmatrix}^\top [r\mathbf{I}_a \otimes (\mathbf{I}_{b-1} + \mathbf{J}_{b-1})] \begin{pmatrix} \beta_{1(1)} \\ \vdots \\ \beta_{b-1(1)} \\ \vdots \\ \beta_{1(a)} \\ \vdots \\ \beta_{b-1(a)} \end{pmatrix}$$

$$\lambda = \frac{r}{2\sigma_e^2} \left(\sum_{i=1}^a \sum_{j=1}^{b-1} \beta_{j(i)}^2 + \sum_{i=1}^a \left[\sum_{j=1}^{b-1} \beta_{j(i)} \right]^2 \right)$$

$$\lambda = \frac{r}{2\sigma_e^2} \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2,$$

entonces:

$$\frac{SCFB(A)}{\sigma_e^2} \sim \chi^2 \left(a(b-1), \frac{r}{2\sigma_e^2} \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2 \right)$$

y

$$\begin{aligned} \mathbb{E}(CMFB(A)) &= \text{traza}(\mathbf{P}_{\beta(A)}\mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{P}_{\beta}\boldsymbol{\mu} \\ &= \sigma_e^2 \text{traza} \left[\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta} \left[\mathbf{K}_{\beta}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{K}_{\beta} \right]^{-1} \mathbf{K}_{\beta}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right] \\ &\quad + r \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2 \\ &= \sigma_e^2 \text{traza} [\mathbf{I}_{a(b-1)}] + r \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2 \\ &= \sigma_e^2 a(b-1) + r \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2. \end{aligned}$$

1.2.3. Suma de cuadrados - error experimental

Sea:

$$SCE = \mathbf{Y}^\top (\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{Y}.$$

Para encontrar la distribución de la suma de cuadrados asociada al error experimental note que:

- La matriz $\mathbf{I}_{abr} - \mathbf{P}_X$ es simétrica.
- Sea $\mathbf{Y}^*$ definida en (C.5). Para aplicar el teorema 2.2 es necesario verificar la idempotencia de $(\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{V}^*$:

$$\begin{aligned} (\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{V}^* (\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{V}^* &= (\mathbf{I}_{abr} - \mathbf{P}_X) (\mathbf{I}_{abr} - \mathbf{P}_X) \\ &= \mathbf{I}_{abr} - \mathbf{P}_X - \mathbf{P}_X + \mathbf{P}_X \mathbf{P}_X = \mathbf{I}_{abr} - \mathbf{P}_X, \end{aligned}$$

por tanto $(\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{V}^*$ es idempotente y entonces:

$$\frac{SCE}{\sigma_e^2} \sim \chi^2(\text{rango}((\mathbf{I}_{abr} - \mathbf{P}_X), \lambda),$$

con:

$$\begin{aligned} \text{rango}(\mathbf{I}_{abr} - \mathbf{P}_X) &= \text{traza}(\mathbf{I}_{abr} - \mathbf{P}_X) = \text{traza}(\mathbf{I}_{abr}) - \text{traza}(\mathbf{P}_X) \\ &= abr - \text{traza}(\mathbf{I}_{1+a-1+a(b-1)}) = abr - ab = ab(r-1) \\ \lambda &= \frac{1}{2} \boldsymbol{\mu}^{*\top} (\mathbf{I}_{abr} - \mathbf{P}_X) \boldsymbol{\mu}^* = \frac{1}{2\sigma_e^2} \left(\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_X \mathbf{X} \boldsymbol{\beta} \right) \\ \lambda &= 0, \end{aligned}$$

entonces:

$$\frac{SCE}{\sigma_e^2} \sim \chi_{ab(r-1)}^2$$

y

$$\begin{aligned} \mathbb{E}(SCE) &= \text{traza}[(\mathbf{I}_{abr} - \mathbf{P}_X) \mathbf{V}] + \boldsymbol{\mu}^\top (\mathbf{I}_{abr} - \mathbf{P}_X) \boldsymbol{\mu} \\ &= \sigma_e^2 \text{traza}[\mathbf{I}_{abr}] - \sigma_e^2 \text{traza}[\mathbf{P}_X] = ab(r-1) \sigma_e^2. \end{aligned}$$

1.3. Independencia entre las sumas de cuadrados consideradas

1.3.1. Independencia entre $SCFA$ y SCE

$$\begin{aligned}
 P_{\alpha}(I - P_X) &= P_{\alpha} - X(X^{\top}X)^{-1}K_{\alpha} \times \\
 &\quad \left[K_{\alpha}^{\top}(X^{\top}X)^{-1}K_{\alpha} \right]^{-1} \cancel{K_{\alpha}^{\top}(X^{\top}X)^{-1}X^{\top}\bar{X}}(X^{\top}X)^{-1}X^{\top} \\
 &= P_{\alpha} - \underbrace{X(X^{\top}X)^{-1}K_{\alpha} \left[K_{\alpha}^{\top}(X^{\top}X)^{-1}K_{\alpha} \right]^{-1} K_{\alpha}^{\top}(X^{\top}X)^{-1}X^{\top}}_{P_{\alpha}} \\
 &= P_{\alpha} - P_{\alpha} = 0.
 \end{aligned}$$

1.3.2. Independencia entre $SCFB(A)$ y SCE

$$\begin{aligned}
 P_{\beta(A)}(I - P_X) &= P_{\beta(A)} - \\
 &\quad X(X^{\top}X)^{-1}K_{\beta(A)} \left[K_{\beta(A)}^{\top}(X^{\top}X)^{-1}K_{\beta(A)} \right]^{-1} \cancel{K_{\beta(A)}^{\top}(X^{\top}X)^{-1}X^{\top}\bar{X}} \times \\
 &\quad (X^{\top}X)^{-1}X^{\top} \\
 &= P_{\beta(A)} - \\
 &\quad \underbrace{X(X^{\top}X)^{-1}K_{\beta(A)} \left[K_{\beta(A)}^{\top}(X^{\top}X)^{-1}K_{\beta(A)} \right]^{-1} K_{\beta(A)}^{\top}(X^{\top}X)^{-1}X^{\top}}_{P_{\beta(A)}} \\
 &= P_{\beta(A)} - P_{\beta(A)} = 0.
 \end{aligned}$$

1.3.3. Distribución cociente de cuadrados medios

$$\frac{\frac{SCFA}{\cancel{\sigma_e^2(a-1)}}}{\frac{SCE}{\cancel{\sigma_e^2 ab(r-1)}}} = \frac{CMFA}{CME} \sim F \left(a-1, ab(r-1), \frac{br}{2\sigma_e^2} \sum_{i=1}^a \alpha_i^2 \right),$$

y

$$\frac{\frac{SCFB(A)}{\cancel{\sigma_e^2(a(b-1))}}}{\frac{SCE}{\cancel{\sigma_e^2 ab(r-1)}}} = \frac{CMFB(A)}{CME} \sim F \left(a(b-1), ab(r-1), \frac{r}{2\sigma_e^2} \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2 \right).$$

Bibliografía

- [1] T. Anderson y D. Darling, "A Test of Goodness of Fit", *Journal of the American Statistical Association*, vol. 49, n°. 268, pp. 765–769, 1954. Disponible en <https://doi.org/10.2307/2281537>
- [2] A. Atkinson, "Two Graphical Displays for Outlying and Influential Observations in Regression", *Biometrika*, vol. 68, n°. 1, pp. 13–20, 1981.
- [3] J. Baena y C. Martínez, *Evaluación experimental de un modelo matemático de una extrusora plastificante monohusillo*, tesis pregrado, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, 2001.
- [4] D. Ballesteros, *Estudio comparativo sobre la obtención de extractos con actividad citotóxica a partir de residuos frutícolas*, tesis maestría, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, 2015. Disponible en <https://repositorio.unal.edu.co/handle/unal/56323>
- [5] V. Barliza y D. Torres, *Evaluación de los procesos de oxidación avanzada: Fenton, UV/H₂O₂ y Foto-Fenton para la degradación de clorpirifós en aguas residuales a nivel laboratorio en la Universidad de Cartagena*, tesis pregrado, Facultad de Ingeniería, Fundación Universidad de América, Bogotá, 2018. Disponible en <https://hdl.handle.net/20.500.11839/7202>

Bibliografía

- [6] A. Barón y J. Giraldo, *Evaluación de la extracción de compuestos orgánicos presentes en matrices acuosas*, tesis pregrado, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, 2002.
- [7] M. Bartlett, “Properties of Sufficiency and Statistical Tests”, *Proceeding of the Royal Society of London, Series A, Mathematical and Physical Sciences*, vol. 160, n°. 901, pp. 268–282, 1937. Disponible en <https://doi.org/10.1098/rspa.1937.0109>
- [8] P. Bickely K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, vol. I, 2.^a ed., San Francisco: CRC Press - Taylor & Francis Group, 2015.
- [9] M. Brown y A. Forsythe, “Robust Tests for the Equality of Variances”, *Journal of the American Statistical Association*, vol. 69, n°. 346, pp. 364–367, 1974.
- [10] E. Bulla, *Diseño del proceso de producción del biodiesel a partir de aceites de fritura*, tesis maestría, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, 2014. Disponible en <https://repositorio.unal.edu.co/handle/unal/51747>
- [11] G. Camacho-Muñoz, H. Felizzola, M. Ríos, J. Parra, y D. Álvarez-Martínez, “Determinación de las condiciones óptimas de operación para un proceso de corte con plasma a través de diseño experimental”, en *Avances en investigación de operaciones y ciencias administrativas, IOCA 2017 Proceedings Book*; A. Pulido-Rojano, R De la Hoz-Reyes y E. Melamed-Varela (eds.), pp. 25–31, 2017.
- [12] G. Canavos, *Probabilidad y estadística. Aplicaciones y Métodos*, México: McGraw-Hill, 1988.
- [13] M. Chao y R. Glaser, “The Exact Distribution of Bartlett’s Test Statistic for Homogeneity of Variances with Unequal Sample Sizes”, *Journal of the American Statistical Association*, vol. 73, n°. 362, pp. 422–426, 1978. Disponible en <https://doi.org/10.2307/2286677>
- [14] W. Cochran y M. Gertrude, *Diseños experimentales*, 3.^a ed., México: Editorial Trillas, 1974.

Bibliografía

- [15] J. Cole y J. Grizzle, “Applications of Multivariate Analysis of Variance to Repeated Measures Experiments”, *Biometrics*, vol. 22, n°. 4, pp. 810–828, 1966. Disponible en <https://doi.org/10.2307/2528076>
- [16] A. Condori, *Determinación del nitrógeno metabólico fecal, endógeno urinario y dérmico en llamas hembras de dos años de edad*, tesis pregrado, Universidad Nacional del Altiplano, Puno, Perú, 2017.
- [17] W. J. Conover, *Practical Nonparametric Statistics*, 2.^a ed., Nueva York: John Wiley & Sons, 1980.
- [18] W. Conover, M. Johnson, y M. Johnson, “A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data”, *Technometrics*, vol. 23, n°. 1, pp. 351–361, 1981. Disponible en <https://doi.org/10.2307/2335801>
- [19] H. Cortés, M. Ballesteros, y A. Zamudio, “Optimización de una metodología para la extracción de la humina de un Oxisol de Colombia. Fraccionamiento y caracterización fisicoquímica”, *Revista Colombiana de Química*, vol. 36, n°. 3, pp. 335–348, 2007.
- [20] F. Cruz y J. Mejía, *Efecto de diferentes densidades de siembra, dosis de nitrógeno y fósforo sobre la producción de semilla de colza (Brassica sp. variedad nacional)*, tesis pregrado, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, 1987.
- [21] A. Dean, D. Voss, y D. Draguljić, *Design and Analysis of Experiments*, 2.^a ed., Nueva York: Springer, 2017.
- [22] G. del Pino, “On Identifiability and Estimability in Response Function Models”, *Brazilian Journal of Probability and Statistics*, vol. 10, n°. 1, pp. 35–57, 1996.
- [23] C. Dunnett, “New Tables For Multiple Comparisons with a Control”, *Biometrics*, vol. 20, n°. 3, pp. 482–491, 1964. Disponible en <https://doi.org/10.2307/2528490>

Bibliografía

- [24] P. Emerson, “Numerical Construction of Orthogonal Polynomials from a General Recurrence Formula”, *Biometrics*, vol. 24, n°. 3, pp. 695–701, 1968. Disponible en <https://doi.org/10.2307/2528328>
- [25] B. Everitt, *The Cambridge Dictionary of Statistics*, Cambridge: Cambridge University Press, 2006.
- [26] M. Fligner y T. Killeen, “Distribution-Free Two-Sample Test for Scale”, *Journal of the American Statistical Association*, vol. 71, n°. 353, pp. 210–213, 1976. Disponible en <https://doi.org/10.2307/2285771>
- [27] J. García, *Desarrollo de un modelo teórico experimental del proceso de corte del bagazo de caña de azúcar orientado al diseño óptimo de un sistema cortador*, tesis maestría, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, 2014. Disponible en <https://repositorio.unal.edu.co/handle/unal/52119>
- [28] J. Guevara y R. Mateus, *Evaluación de la utilización de probióticos en la fase de levante del ciclo de producción de la mojarra roja (Oreochromis sp.)*, tesis pregrado, Facultad de Medicina Veterinaria y de Zootecnia, Universidad Nacional de Colombia, Bogotá, 2001.
- [29] A. Gutiérrez, *Análisis de algunos aspectos tróficos y reproductivos de la comunidad de peces de un caño de aguas negras amazónicas en cercanías de Leticia (Amazonas, Colombia)*, tesis pregrado, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, 2003.
- [30] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, Nueva York: Springer, 2008.
- [31] K. Hinkelmann y O. Kempthorne, *Design and Analysis of Experiments. Vol. I: Introduction to Experimental Design*, Nueva York: John Wiley & Sons, 1994.
- [32] K. Hinkelmann y O. Kempthorne, *Design and Analysis of Experiments. Vol. 2: Advanced Experimental Design*, John Wiley & Sons, 2005.

Bibliografía

- [33] K. Hinkelmann y O. Kempthorne, *Design and Analysis of Experiments*. Vol. I: *Introduction to Experimental Design*, 2.^a ed., Nueva York: John Wiley & Sons, 2008.
- [34] R. Hocking, *Methods and Applications of Linear Models*, Nueva York: John Wiley & Sons, 1996.
- [35] J. Kloeke y J. McKean, *Nonparametric Statistical Methods Using R*, Boca Raton: CRC Press, 2015.
- [36] C. Kramer, “Extension of Multiple Range Test to Groups Means with Unequal Numbers of Replications”, *Biometrics*, vol. 12, n°. 3, pp. 307–310, 1956. Disponible en <https://doi.org/10.2307/3001469>
- [37] R. Lenth, “Quick and Easy Analysis of Unreplicated Factorials”, *Technometrics*, vol. 31, n°. 4, pp. 469–473, 1989. Disponible en <https://doi.org/10.2307/1269997>
- [38] H. Levene, “Robust test for equality of variances”, en *Contributions to Probability and Statistics*. I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow y H. B. Mann (eds.), Palo Alto: Stanford University Press, 1960, ch. 25, pp. 278–292.
- [39] H. Lilliefors, “On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown”, *Journal of the American Statistical Association*, vol. 62, n°. 318, pp. 399–402, 1967. Disponible en <https://doi.org/10.2307/2283970>
- [40] O. Melo, L. López, y S. Melo, *Diseño de Experimento: métodos y aplicaciones*, 1.^a ed., Bogotá: Universidad Nacional de Colombia, 2007.
- [41] D. Montgomery, *Diseño y análisis de experimentos*, 2.^a ed., México: Editorial Limusa, 2002.
- [42] J. Muñoz, D. Camargo, y D. Gallego, “Aplicación de la metodología de superficies de respuesta en un proceso de absorción del CO_2 de un biogás en una solución alcalina”, *Dyna*, vol. 76, n°. 159, pp. 113–144, 2009.

Bibliografía

- [43] R. Myers, D. Montgomery, y C. Anderson-Cook, *Response Surface Methodology*, 3.^a ed., Nueva York: Wiley, 2009.
- [44] E. Ojasild, *Elaboración de néctares de gulupa (Passiflora edulis f. edulis) y curuba (Passiflora mollissima)*, tesis especialización, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, 2009.
- [45] G. Paula, *Modelos de regressão com apoio computacional*, São Paulo: Universidade de São Paulo, 2012. Disponible en https://www.ime.usp.br/~giapaula/texto_2013.pdf
- [46] M. Pereira y A. Amaral, “Estimation of Genetic Components in Popcorn Based on the Nested Design”, *Crop Breeding and Applied Biotechnology*, vol. 1, n°. 1, pp. 3–10, 2001.
- [47] A. Pérez, *Evaluación de las propiedades farmacotécnicas del almidón de ñame y su aplicación como diluyente en la elaboración de formas farmacéuticas sólidas*, tesis pregrado, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, 2002.
- [48] J. Posada, *Efecto de la cantidad de carga en el consumo de combustible en camiones*, tesis doctorado, Facultad de Minas, Universidad Nacional de Colombia, Medellín, 2012. Disponible en <https://repositorio.unal.edu.co/handle/unal/11067>
- [49] A. Quiroga, *Respuesta a las aplicaciones de silicio en el cultivo de pepino (cucumis sativus L) variedad modan, en condiciones de estrés hídrico bajo cubierta en Culiacán, Sinaloa*, tesis pregrado, Facultad de Ingeniería, Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá, 2016. Disponible en <https://repository.udca.edu.co/handle/11158/582>
- [50] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, 2021. Disponible en <https://www.R-project.org/>

Bibliografía

- [51] J. Rafter, M. Abell, y J. Braselton, "Multiple Comparison Methods for Means", *SIAM Review*, vol. 44, n°. 2, pp. 259–278, 2002. Disponible en <https://doi.org/10.1137/S003614450135723>
- [52] D. Robson, "A Simple Method for Constructing Orthogonal Polynomials when the Independent Variable is Unequally", *Biometrics*, vol. 15, n°. 2, pp. 187–191, 1959. Disponible en <https://doi.org/10.2307/2527668>
- [53] C. Sampedro y H. Sanjuan, *Ensayos de rendimientos de variedades promisorias de cebada*, tesis pregrado, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, 1981.
- [54] C. E. Särndal, B. Swensson, y J. Wretman, *Model Assisted Survey Sampling*, Nueva York: Springer-Verlag, 1992.
- [55] H. Scheffe, "A Method for Judging all Contrasts in the Analysis of Variance", *Biometrika*, vol. 40, n°. 1/2, pp. 87–104, 1953. Disponible en <https://doi.org/10.2307/2333100>
- [56] S. Searle, *Linear Models*, Nueva York: John Wiley & Sons, 1971.
- [57] S. Searle, G. Casella, y C. McCulloch, *Variance Components*, Nueva York: John Wiley & Sons, 2006.
- [58] J. Shao, *Mathematical Statistics*, 2.^a ed., Nueva York: Springer, 2003.
- [59] S. Shapiro y M. Wilk, "An analysis for variance test for normality (complete samples)", *Biometrika*, vol. 52, n°. 3/4, pp. 591–611, 1965. Disponible en <https://doi.org/10.2307/2333709>
- [60] W. Shapiro y R. Francia, "An Approximate Analysis of Variance Test for Normality", *Journal of the American Statistical Association*, vol. 67, n°. 337, pp. 215–216, 1972. Disponible en <https://doi.org/10.1080/01621459.1972.10481232>
- [61] D. Wu y M. Hamada, *Experiments. Planning, Analysis, and Parameter Design Optimization*, Nueva York: John Wiley & Sons, 2000.

Bibliografia

- [62] D. Zucker, "An Analysis of Variance Pitfall: the Fixed Effects Analysis in a Nested Design", *Educational and psychological measurement*, vol. 50, n°. 4, pp. 731–738, 1990. Disponible en <https://doi.org/10.1177/0013164490504002>

Índice

- AIC (información de Akaike), 224
- aleatorización, 9
- alias, 313, 315, 316
- ANOVA
 - ortogonal, 37
 - simple, 36
- causas de variación, 6
- combinación lineal, 29
- contraste, 101
- criterio de menor aberración, 308
- cuadrados medios, 392
- diferencia mínima significativa, 110
- diseño balanceado, 50
- diseño completamente aleatorizado, 384
- diseño experimental, 3
- distribución
 - F no central, 42
 - t no central, 42
- ecuaciones normales, 25
- efecto clear o transparente, 338
- efecto principal, 215
- error
 - de tratamiento, 6
 - experimental, 6
 - observacional, 6
- error observacional, 385
- estimación por mínimos cuadrados, 24
- factor, 7, 8
 - anidado, 355
- factores cruzados, 215
- forma cuadrática, 38
- forma lineal, 38
- función lineal, 21
- función lineal estimable, 29, 96
- matriz de covarianza, 32
- matriz diseño, 22
- mejor estimador lineal insesgado (ME-LI), 31
- modelo
 - de efectos aleatorios, 5
 - de efectos fijos, 5

Índice

- de efectos mixtos, 5
- lineal, 21
- nivel de significancia por comparación, 108
- principio
 - de heredabilidad, 215
 - de orden jerárquico, 215
 - de parsimonia, 215
- producto Kronecker, 402
- relación de definición, 313, 315, 316
- replicación, 9
- respuestas potenciales o conceptuales, 47
- suma de cuadrados
 - modelo, 36
 - residual, 36
 - total, 36
- tamaño de muestra
 - DCA con submuestreo, 79
 - diseño completamente aleatorizado (DCA), 69
 - diseño con arreglos factoriales, 228
 - diseño en bloques, 177
 - diseño en bloques incompletos, 186
 - efectos aleatorios, 83
- tasa de error, 108
- tratamiento, 7, 211
- unidad
 - de observación o de muestreo, 6, 70
- experimental, 5, 70
- variable
 - dependiente, 21
 - independiente, 21