

Notas de Clase - Diseño de Experimentos

Partes I, II y III del curso 3007340

Nelfi González Álvarez

Doctora en Ciencias Estadísticas

Profesora Asociada Escuela de Estadística

Universidad Nacional de Colombia Sede Medellín



UNIVERSIDAD NACIONAL DE COLOMBIA

2021

Índice general

Lista de Abreviaturas	XI
------------------------------	-----------

I INTRODUCCIÓN AL DOE - DISEÑOS COMPLETAMENTE ALEATOR- IZADOS CON UN SÓLO FACTOR DE TRATAMIENTOS	1
---	----------

1. Introducción al diseño de experimentos	3
1.1. Clasificación de los problemas experimentales	4
1.1.1. Comparación de tratamientos	4
1.1.2. Depuración o selección de variables	4
1.1.3. Exploración de superficies de respuesta	4
1.1.4. Optimización de sistemas	5
1.1.5. Robustificación de sistemas	5
1.2. Aproximación sistemática a la planeación e implementación de experi- mentos	5
1.2.1. Establecer el objetivo	6
1.2.2. Elegir la respuesta	6
1.2.3. Elegir factores y niveles	8
1.2.4. Elegir el plan experimental	11
1.2.5. Ejecutar el experimento	11
1.2.6. Analizar los datos	12
1.2.7. Extraer conclusiones y hacer recomendaciones	12
1.3. Contenido del plan experimental	12

1.4. Los tres principios fundamentales de la experimentación	15
1.5. Estructura de un diseño experimental	21
1.5.1. Estructura de tratamientos	22
1.5.2. Estructura de diseño	24
1.6. Clasificación de los diseños experimentales	25
1.7. La investigación experimental vs. DOE	27
1.8. Consideraciones prácticas	27
1.9. Qué se hace después del primer experimento	28
1.10. Qué hacer cuando ningún efecto es significativo	31
1.11. Ejemplos de DOE	32
1.11.1. Efectos de cinco dietas sobre la pérdida de peso	32
1.11.2. Resistencia de cuatro pinturas de uso doméstico a condiciones ambientales	35
1.11.3. Protección del acero contra el calor	38
1.11.4. Efectos de los niveles de nitrógeno (N) y de potasio (K) sobre una variedad de maíz	41
1.11.5. Bloques y replicaciones	42
2. Generalidades de los Modelos ANOVA para DOE	45
2.1. Supuestos básicos	45
2.2. Usos del análisis de varianza	46
2.3. Modelo de análisis de varianza o ANOVA	46
2.3.1. Relación entre regresión y análisis de varianza	46
2.3.2. Elección entre regresión y ANOVA	48
2.3.3. Modelo lineal de la variable respuesta relativo a modelos ANOVA .	49
2.3.4. Modelo ANOVA I (efectos fijos) vs. Modelo ANOVA II (efectos aleato- rios)	50
3. Experimentos con un sólo factor de efectos fijos	53
3.1. Introducción	53

3.2. Diseño completamente aleatorizado o DCA	53
3.2.1. Aleatorización	54
3.2.2. Modelo para un diseño completamente aleatorizado, un factor efectos fijos	58
3.2.3. Test ANOVA para un DCA	61
3.2.4. Distribuciones muestrales de los estimadores	63
3.2.5. Validación de supuestos	63
3.2.6. Ejemplo: Mejora de la productividad (tomado de [7])	67
3.2.7. Ejemplo: Efectos de un fertilizante (tomado de [6])	74
4. Inferencias para medias de tratamientos con un factor	83
4.1. Comparaciones de pares de medias de tratamientos - mínima diferencia significativa o LSD	83
4.1.1. Ejemplo. [4]	84
4.2. Prueba de rangos múltiples de Duncan	85
4.2.1. Ejemplo	86
4.3. Método de Tukey de comparaciones de todos los pares de medias, o rango estudentizado de Tukey	87
4.3.1. Ejemplo	88
4.4. Comparaciones de tratamientos con un control - Método de Dunnett . . .	89
4.5. Métodos de Dunnett unilaterales	89
4.6. Ejemplos de aplicación	90
4.6.1. Con datos experimento de las lechugas	90
4.6.2. Con datos experimento sobre la mejora de la productividad	96
4.7. Comparaciones mediante contrastes	100
4.7.1. Pruebas de hipótesis sobre un contraste de medias	101
4.7.2. Ejemplo, [2]	102
4.7.3. Método de Bonferroni para comparaciones preplaneadas	103
4.7.4. Ejemplo	104

4.7.5. Método de Scheffé	105
4.7.6. Ejemplo	105
4.7.7. Relación entre ANOVA y el método de Scheffé.	106
4.7.8. Ejemplo	107
4.7.9. Ejemplo	111
5. Modelo ANOVA un solo factor efectos aleatorios	113
5.1. Modelo de un sólo factor de efectos aleatorios	117
5.2. Distribución de las variables de respuesta	118
5.3. Análisis de varianza	119
5.4. Estimaciones	120
5.4.1. Estimaciones mediante el método ANOVA	120
5.4.2. Estimaciones por máxima verosimilitud (MLE) y máxima verosimilitud restringida o residual (REML)	123
5.5. Chequeo de supuestos	127
5.6. Ejemplo	128
6. Potencia en ANOVA con un Factor en un DCA	135
6.1. Potencia en experimentos con un factor de efectos fijos en un DCA . . .	135
6.1.1. Tamaños de muestra para alcanzar intervalos de confianza de longitudes especificadas	136
6.1.2. Tamaños de muestra usando la potencia de la prueba ANOVA . .	138
6.2. Potencia en ANOVA un factor de efectos aleatorios en un DCA y tamaños de muestra	149
II DISEÑOS EN BLOQUES COMPLETOS CON UN SÓLO FACTOR DE TRATAMIENTOS	155
7. Diseños en bloques completos aleatorizados	157
7.1. Distinción entre factores de bloque, de ruido y covariables	157

7.2. Tamaño de los bloques	159
7.3. El diseño de bloques completos aleatorizados o DBCA	161
7.3.1. Análisis de diseños de bloques completos aleatorizados	161
7.3.2. Comparaciones múltiples	164
7.3.3. Chequeo de supuestos del modelo	165
7.3.4. Determinación del número de observaciones por tratamiento . . .	165
7.4. Diseño general de bloques completos (sección complementaria)	166
7.4.1. Modelo y ANOVA	166
7.5. Efectividad en la construcción de bloques: Eficiencia relativa (ER) y pre- cisión de un experimento	170
7.6. Ejemplo: Comparación de cuatro métodos de ensamble	173
7.7. Problema para análisis	181
7.8. Diseños en cuadrados latinos DCL	187
7.8.1. Ejemplo [4]	191

III DISEÑOS COMPLETAMENTE ALEATORIZADOS CON TRATAMIENTOS FACTORIALES

201

8. Experimentos factoriales

203

8.1. Introducción	203
8.1.1. Experimentos factoriales vs. la estrategia de mover un factor a la vez	204
8.1.2. Ventajas de los diseños factoriales	204
8.2. Diseños factoriales con dos factores	205
8.2.1. Tipos de experimentos factoriales con dos factores	205
8.3. Modelo factorial dos factores efectos fijos con interacción, en un DCA . .	206
8.3.1. Modelo ANOVA	206
8.3.2. Parámetros de interés	209
8.3.3. Significado de la interacción	209

8.3.4. Encubrimiento o enmascaramiento de efectos principales	213
8.3.5. Estimaciones	213
8.3.6. Análisis de varianza	214
8.3.7. Hipótesis a probar	215
8.3.8. Evaluación de los efectos simples	217
8.3.9. Comparaciones múltiples	222
8.3.10. Validación de supuestos	224
8.4. Modelo factorial dos factores de efectos fijos sin interacción, en un DCA	225
8.4.1. Estimaciones	225
8.4.2. ANOVA	225
8.4.3. Hipótesis a probar	226
8.4.4. Comparaciones múltiples y contrastes de medias	227
8.4.5. Validación de supuestos	228
8.5. Ejemplo experimento factorial dos factores de efectos fijos en un DCA .	229
8.5.1. Modelo	230
8.5.2. Análisis descriptivo	230
8.5.3. ANOVA y tests de hipótesis	233
8.5.4. Inferencias sobre medias de tratamientos técnico*marca	235
8.5.5. Validación de supuestos	243
8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA	247
8.6.1. Modelo ANOVA	248
8.6.2. Parámetros de interés	249
8.6.3. ANOVA	249
8.6.4. Pruebas de hipótesis	252
8.6.5. Estimaciones	253
8.6.6. Validación de supuestos	254
8.7. Modelo factorial de dos factores con efectos aleatorios sin interacción en un diseño DCA	254

8.7.1. ANOVA	255
8.7.2. Hipótesis a probar	255
8.7.3. Estimaciones	257
8.7.4. Validación de supuestos	257
8.8. Ejemplo dos factores de efectos aleatorios en un DCA	257
8.8.1. Modelo para el experimento	258
8.8.2. Análisis descriptivo	259
8.8.3. ANOVA y estimación de componentes de varianza	260
8.8.4. Validación de supuestos sobre los errores	264
8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado	268
8.9.1. Modelo Mixto restringido	269
8.9.2. Modelo mixto no restringido	276
8.10. Modelo factorial de dos factores de efectos mixtos sin interacción, en un DCA balanceado	282
8.10.1. Parámetros de interés	283
8.10.2. ANOVA	284
8.10.3. Pruebas de hipótesis	284
8.10.4. Estimaciones	285
8.10.5. Contrastes de medias $\mu_{i\bullet}$	286
8.10.6. Validación de supuestos	287
8.11. Ejemplo dos factores de efectos mixtos en un DCA	287
8.11.1. Modelo para el experimento	288
8.11.2. Análisis descriptivo	289
8.11.3. ANOVA y estimaciones	291
8.11.4. Validación de supuestos sobre los errores	298

A. Complementos 305

A.1. Funciones estimables de los parámetros	305
---	-----

A.2. Test de Bartlett de homogeneidad de varianzas	305
A.3. Test de Levene de Homogeneidad de varianza	306
A.4. Test de Cochran para homogeneidad de varianza	308
A.4.1. Test C de Cochran	308
A.4.2. Test G de Cochran	309
A.5. Transformaciones para homogeneizar varianza	311
A.6. Análisis con varianzas desiguales	314
A.7. Deducción de los intervalos de confianza para los parámetros en modelo	
DCA un factor de efectos aleatorios	315
A.7.1. I.C para σ^2	316
A.7.2. I.C para σ_α^2	316
A.7.3. I.C para σ_α^2/σ^2	317
A.7.4. I.C para $\sigma_\alpha^2/(\sigma^2 + \sigma_\alpha^2)$	317
A.7.5. I.C para μ	318
A.8. Procedimiento Satterthwaite, [7]	319

Lista de Abreviaturas

<i>I.C</i>	Intervalo de Confianza
<i>IID</i>	Independiente(s) e idénticamente distribuido(s)
<i>m.a</i>	Muestra Aleatoria
<i>TLC</i>	Teorema de Límite Central
<i>v.a</i>	Variable Aleatoria

Índice de figuras

1.1. Diagrama de flujo proceso de fabricación de pulpa de papel. Adaptado de [14].	9
1.2. Diagrama causa - efecto, experimento del moldeo por inyección. Adaptado de [14].	10
1.3. Diagrama de causa - efecto revisado, experimento del moldeo por inyección. Adaptado de [14].	20
1.4. Diagrama de Entrada - Proceso - Salida . Adaptado de [4].	20
1.5. Diagrama de Entrada - Proceso - Salida en la fabricación de un envase plástico. Adaptado de [4].	21
1.6. Estructuras de tratamiento y de diseño y su combinación en un DOE. Adaptado de [9].	22
1.7. Procesos en la investigación experimental. Adaptado de [5].	28
2.1. Relación entre regresión y ANOVA. Adaptado de [7]. (a) Punto de vista de la regresión: interesa la función que establece cómo cambia la media de la respuesta con los cambios en los valores del predictor lineal; (b) Punto de vista del modelo ANOVA: solo interesa que en los distintos niveles de valores del predictor lineal, existen diferentes poblaciones estadísticas de la respuesta que se diferencian en sus medias, pero no interesa modelar los cambios en estas medias como función de los valores del predictor lineal.	48
3.1. Boxplots sobre datos en ejemplo 3.2.6.	73

3.2. Residuos internamente estudentizados y su gráfico de probabilidad normal, para modelo ANOVA sobre datos en ejemplo 3.2.6.	73
3.3. Boxplots sobre datos en ejemplo 3.2.7.	80
3.4. Residuos internamente estudentizados y su gráfico de probabilidad normal, para modelo ANOVA sobre datos en ejemplo 3.2.7.	81
4.1. I.C de Tukey para las diferencias de medias de tratamiento en 3.2.7. Es resultado del uso de <code>plot(TukeyHSD(...))</code> ; vea Código R 4.6.1.	93
4.2. I.C de Tukey para las diferencias de medias de tratamiento en 3.2.6. Es resultado del uso de <code>plot(TukeyHSD(...))</code> ; vea Código R 4.6.3.	100
5.1. Representación del modelo ANOVA de un factor de efectos aleatorios. Adaptado de [7]. Note que las $\mu_i = \mu + A_i$, por tanto, $\mu_i \stackrel{iid}{\sim} N(\mu, \sigma_\alpha^2)$	117
5.2. Boxplots para los datos del experimento del contenido de lana limpia . .	131
5.3. Residuos internamente estudentizados para el experimento del contenido de lana limpia. Gráficos de probabilidad normal también sobre residuos internamente estudentizados y efectos estimados (estandarizados). . . .	132
6.1. Comportamiento de la diferencia $q_{0.05}^2(5, 5(n-1)) - 0.9n$ vs. n , en la solución del tamaño de muestra óptimo según comparaciones de Tukey con un $HSD_{ij} \leq 3 \text{ mm}$, a un nivel de significancia de 0.05 y usando $\hat{\sigma}^2 = 10 \text{ mm}^2$ en un DCA de efectos fijos de un sólo factor con $a = 5$ niveles. . .	137
6.2. Comportamiento de la potencia del test ANOVA vs. n , en un DCA con un factor de efectos fijos, cuando se desea detectar diferencias entre medias a un nivel de significancia de 0.05, con las verdaderas medias sean iguales a $\mu_1 = 15$, $\mu_2 = 13$, $\mu_3 = 19.5$, $\mu_4 = 27.5$	144
6.3. Comportamiento de la potencia del test ANOVA vs. n , en un DCA con un factor de efectos fijos con $a = 3$, cuando se desea detectar diferencias $\Delta = \mu_i - \mu_j = 0.25$, entre cualesquiera dos medias, a un nivel de significancia de 0.05.	148

7.1. Boxplots tiempos de ensamble según método y según operador	174
7.2. Medias de tiempos de ensamble según método y según operador	175
7.3. I.C Tukey del 95 % de confianza, para las diferencias de medias de tratamien- tos (Métodos de ensamble)	177
7.4. Gráficos de residuales internamente estudentizados y test de Normali- dad en ejemplo 7.6.	178
7.5. Boxplots tiempos de montaje según operador y según tipo de montaje .	182
7.6. Medias de tiempos de ensamble según operador y según tipo de montaje	183
7.7. Gráficas de los intervalos de confianza de Tukey	184
7.8. Gráficas de residuos internamente estudentizados Para análisis de supuestos	185
7.9. Boxplots datos Ejemplo 7.8.1	193
7.10. Gráfica de los I.C de Tukey para las diferencias entre medias del factor Marca, en ejemplo 7.8.1.	196
7.11. Gráfica de residuos internamente estudentizados y qqplot normal, en ejemplo 7.8.1.	197
8.1. Gráficos de perfiles de medias.	210
8.2. (a) Perfiles de medias de tratamientos según marca. (b) perfiles de medias de tratamientos según técnico.	231
8.3. (a) Distribuciones según técnico. (b) Distribuciones según marca.	232
8.4. (a) Distribuciones de tiempos de técnicos en cada marca. (b) Distribu- ciones de tiempos por marcas en cada técnico.	233

8.5. Gráficos I.C de Tukey del 95 % de confianza, para las diferencias de medias de tratamientos. Observe que no se detectan diferencias pues el cero está contenido en todos los intervalos, sinembargo, esto es debido a una gran reducción de la potencia ante un número grande de comparaciones. Observe además que en muchos casos el cero está muy cerca a uno de los límites, y en esas circunstancias es muy probable que sí existan diferencias entre las medias de tratamientos para las cuales se observa esta situación.	238
8.6. Gráficos de residuos internamente estudentizados y de probabilidad normal en ejemplo 8.5.	244
8.7. Boxplots según factores y gráficos de interacción, en datos del Ejemplo 8.8	260
8.8. Gráficas para evaluación de supuestos sobre los errores, en datos del Ejemplo 8.8	265
8.9. Boxplots según factores y gráficos de interacción, en datos del Ejemplo 8.11	290
8.10. Gráficos de residuales internamente estudentizados para validar supuestos en modelo factorial mixto, en datos del Ejemplo 8.11	299
A.1. Gráfico del log-verosimilitud vs. λ producido por función R <code>boxCox</code> sobre los datos del ejemplo 3.2.6	313

Índice de tablas

3.1. Estructuras de diseño en experimentos con un sólo factor de efectos fijos, y modelos estadísticos correspondientes	54
3.2. Tabla de datos experimentales de un DOE en un DCA	60
3.3. Tabla ANOVA en un DCA	62
3.4. Estimadores, errores estándar e I.C para los parámetros del modelo . .	63
3.5. Supuestos a evaluar y pruebas correspondientes	64
3.6. Algunas transformaciones de la respuestas para normalidad, con base en la distribución original	67
4.1. Tests de hipótesis sobre contrastes de medias de tratamientos	101
5.1. Tabla ANOVA en un DCA con un sólo factor de efectos aleatorios	119
5.2. Estimadores puntuales y por intervalos de confianza de $(1 - \gamma) 100\%$, . .	121
6.1. Evaluación de los tamaños de muestra satisfaciendo $q_{0.05}^2(5, 5(n - 1)) \leq 0.9n$	137
7.1. Tabla de datos experimentales de un DOE en un DBCA	161
7.2. ANOVA en un DBCA	164
7.3. ANOVA en un DBCA con bloques de tamaño $k = sa$, sin interacción factor-bloques	167
7.4. ANOVA en un DBCA con bloques de tamaño $k = sa$, con interacción factor-bloques	168
7.5. Cuadrado latino 4x4 inicial (izq.); sorteo de columnas 3, 2, 4, 1 (centro); sorteo de las filas 4, 2, 1, 3 (der.)	187

7.6. ANOVA en un DCL $r \times r$	190
8.1. Tratamientos en un experimento con dos factores A y B	204
8.2. Tabla de datos experimentales de un DOE en un DCA balanceado, dos factores	208
8.3. Parámetros y sus estimadores de MCO en un diseño factorial de efectos fijos con interacción, en un DCA.	214
8.4. ANOVA en un DCA balanceado, con dos factores de efectos fijos con interacción	216
8.5. Definición de efectos simples de A en cada nivel de B	218
8.6. Definición de efectos simples de B en cada nivel de A	219
8.7. Estimadores de efectos simples	220
8.8. significancia de los efectos simples	223
8.9. Parámetros y sus estimadores de MCO en un diseño factorial de efectos fijos sin interacción, en un DCA.	226
8.10 ANOVA en un DCA balanceado, con dos factores de efectos fijos sin interacción	227
8.11 ANOVA en un DCA balanceado, con dos factores de efectos aleatorios con interacción	251
8.12 Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos aleatorios con interacción	253
8.13 ANOVA en un DCA balanceado, con dos factores de efectos aleatorios sin interacción	256
8.14 Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos aleatorios sin interacción	257
8.15 Estimaciones en el ejemplo	264
8.16 ANOVA en un DCA balanceado, con dos factores de efectos Mixtos (A factor de efectos fijos y B factor de efectos aleatorios) con interacción, modelo restringido	273

8.17	Estimaciones de las componentes y propociones de varianza en modelo factorial, dos factores de efectos mixtos con interacción, modelo restringido	275
8.18	Estimadores de efectos fijos y medias del factor A, factorial dos factores de efectos mixtos con interacción	275
8.19	ANOVA en un DCA balanceado, con dos factores de efectos Mixtos (A factor de efectos fijos y B factor de efectos aleatorios) con interacción, modelo no restringido	279
8.20	Estimaciones de las componentes y propociones de varianza en modelo factorial, dos factores de efectos mixtos con interacción, modelo no restringido	281
8.21	ANOVA en un DCA balanceado, con dos factores de efectos mixtos sin interacción	285
8.22	Estimaciones de las componentes y propociones de varianza en modelo factorial, dos factores de efectos mixtos sin interacción.	286
8.23	Estimadores de efectos fijos y medias del factor A, factorial dos factores de efectos mixtos sin interacción	286
A.1.	Transformaciones de potencia Y^p usuales. Fuente [6].	312

Parte I

INTRODUCCIÓN AL DOE - DISEÑOS COMPLETAMENTE ALEATORIZADOS CON UN SÓLO FACTOR DE TRATAMIENTOS

Capítulo 1

Introducción al diseño de experimentos

Según [14], La experimentación cubre un amplio rango de aplicaciones que permite averiguar qué pasa con un resultado o respuesta cuando los ajustes sobre ciertas variables de entrada en un sistema son cambiados a propósito, y básicamente permite comprender y/o mejorar el sistema estudiado, el cual puede ser un producto o un proceso, bien sea de ingeniería, de las ciencias, de la manufactura, de los servicios, de la administración, etc. Puede usarse tanto el análisis estadístico como el sentido común para estudiar las relaciones entre variables de entrada y de salida, de modo que una mayor comprensión de cómo las primeras afectan a las segundas conlleven a la mejora del sistema particular en el que se haya inmerso el problema. El diseño de experimentos es un cuerpo de conocimientos y técnicas que permiten conducir de una manera efectiva y eficiente los experimentos y el análisis de los datos recolectados, y hacer conexiones entre las conclusiones de los análisis y los objetivos originales de la investigación.

[9] dice que el diseño de experimentos **se ocupa de la planeación de experimentos con el fin de obtener la máxima cantidad de información de los recursos disponibles.** A menudo en el diseño de un experimento es necesario controlar ciertos efectos llamados **tratamientos, poblaciones, o combinaciones de tratamientos.** El experimentador generalmente también controla la elección de las **unidades experimentales** a ser colocadas en **grupos, llamados bloques.** Un experimento típico involucra **t** tratamientos (o combinaciones de tratamientos) que son comparados o cuyos efectos van a ser

estudiados.

1.1. Clasificación de los problemas experimentales

Los problemas experimentales pueden clasificarse en cinco grandes categorías de acuerdo a sus objetivos, [14]:

1.1.1. Comparación de tratamientos

En este tipo de problemas **interesa comparar varios tratamientos** y seleccionar el o los mejores. Por ejemplo, en la comparación de seis variedades de cebada ¿éstas son diferentes en términos de la producción y la resistencia a la sequía? Si de hecho son diferentes, ¿cómo difieren y cuáles son las mejores?

Ejemplos de tratamientos incluyen: variedades o factores en ensayos agrícolas, posiciones de trabajo en estudios de ergonomía, métodos de instrucción, tipos de máquinas, proveedores, etc.

1.1.2. Depuración o selección de variables

Hay un gran número de variables o factores en un sistema pero **sólo un número relativamente pequeño de ellas son importantes**. Un experimento de depuración (*screening experiment*) puede ser conducido para identificar estas variables. Tales experimentos tienden a ser económicos porque trabajan con pocos grados de libertad para estimar la varianza del error y los términos de alto orden como los efectos cuadráticos o las interacciones entre factores. Una vez se identifican las variables importantes, se puede conducir otro experimento para estudiar con más detalles los efectos de tales factores. Esta fase del estudio cae en la categoría que se presenta a continuación.

1.1.3. Exploración de superficies de respuesta

Una vez se ha identificado el pequeño número de variables más importantes, es necesario explorar sus efectos sobre la respuesta. **La relación entre la respuesta y estas variables es algunas veces denominada superficie de respuesta**. Usualmente el experimento está basado en un diseño que permita estimar efectos lineales y cuadráticos de las variables y algunas de las interacciones entre éstas. Estos ex-

1.2. Aproximación sistemática a la planeación e implementación de experimentos 5

perimentos tienden a ser más grandes que los experimentos de depuración. Se han desarrollado tanto modelos paramétricos como no paramétricos, donde los últimos son más intensivos en términos computacionales pero más flexibles en el ajuste del modelo.

1.1.4. Optimización de sistemas

En muchas investigaciones el interés yace en optimizar un sistema. Por ejemplo, el rendimiento de procesamiento de una planta de ensamble o de la producción de un proceso químico va a ser maximizado; la cantidad de desperdicio o el número de piezas reprocesadas en una operación de estampación va a ser minimizado; el tiempo requerido para procesar una reclamación de reembolso va a ser minimizado. Si se ha identificado una superficie de respuesta, ésta puede ser usada para la optimización. Sin embargo, para el propósito de hallar un óptimo no es necesario trazar la superficie total como en la exploración de una superficie de respuesta. Una inteligente estrategia secuencial puede mover rápidamente al experimento a una región que contiene los ajustes óptimos de las variables. Sólo dentro de esta región se justifica una exploración cuidadosa.

1.1.5. Robustificación de sistemas

Además de optimizar la respuesta, es importante en **mejoramiento de la calidad** hacer al sistema robusto frente a la "variación del ruido". Esto a menudo es alcanzado eligiendo **ajustes de los factores controlables** en los cuales el sistema sea menos sensible a este tipo de variación. Aunque tal variación es difícil de controlar en condiciones normales, se requiere variarla sistemáticamente durante la experimentación. La respuesta en el análisis estadístico es a menudo la varianza (o una transformación de ésta) entre las réplicas de ruido para un ajuste dado del factor de control.

1.2. Aproximación sistemática a la planeación e implementación de experimentos

Según [9], antes de llevar a cabo un experimento es necesario dar respuesta a las siguientes preguntas:

1. ¿Cuántos tratamientos van a ser estudiados?
2. ¿Cuántas veces se necesita observar cada tratamiento?
3. ¿Cuáles son las unidades experimentales?
4. ¿Cómo se aplicarán los tratamientos a las unidades experimentales disponibles y luego observar las respuestas?
5. ¿Puede el diseño resultante ser analizado o pueden ser hechas las comparaciones deseadas?

Las respuestas a estas preguntas no son necesariamente directas y no pueden responderse de manera general. Sin embargo, las ideas y conceptos que se presentarán pueden ser útiles para obtener la información necesaria para hallar las respuestas.

Las siguientes subsecciones (incluyendo los ejemplos que se presentan) son tomadas de [14], quien establece que es necesario seguir en la planeación e implementación de experimentos los pasos que se describen a continuación.

1.2.1. Establecer el objetivo

Es necesario establecer con claridad el objetivo del experimento. Todos los interesados deberían participar. Por ejemplo, para un producto manufacturado, los interesados pueden incluir a los ingenieros de diseño del producto, a los ingenieros de diseño de proceso, a los ingenieros de producción, a los trabajadores de la línea, a los administradores, proveedores y consumidores.

1.2.2. Elegir la respuesta

La respuesta es un resultado u observación experimental. En un experimento pueden haber múltiples respuestas. Varias cuestiones resultan al elegir una respuesta. Las respuestas pueden ser discretas o continuas. Las respuestas discretas pueden ser de conteo o categóricas, es decir, binarias (bueno, malo) u **ordinales (fácil, normal, difícil)**. Generalmente se prefieren respuestas continuas. Por ejemplo, es preferible la

1.2. Aproximación sistemática a la planeación e implementación de experimentos 7

medición continua de la fuerza para abrir una puerta que un juicio ordinal (fácil, normal, difícil de abrir); el registro de una característica continua es preferible al registro del porcentaje de que la característica está dentro de sus especificaciones. Sin embargo, no siempre existe el medio para realizar una medición continua y es preferible usar una respuesta ordinal que no hacer el experimento. Para una medición continua debe contarse con un buen sistema de medición. De hecho un experimento llamado estudio G R & R (*Gauge Repeatability and Reproducibility study*) puede llevarse a cabo para asegurar un sistema de medición continua. Cuando hay un solo mecanismo de medición, la variación debida al sistema de medición puede ser dividido en dos tipos: la variación entre los operadores y la variación dentro de los operadores. Idealmente no debería haber variación entre operadores y la variación dentro de operador debería ser pequeña. El estudio G R & R proporciona estimaciones para estas dos componentes de varianza del sistema de medición.

Finalmente, la respuesta puede ser elegida para incrementar la comprensión de los mecanismos y de las leyes físicas involucradas en el problema. Por ejemplo, en un proceso que produce barras de jabón con sobrepeso, el peso de la barra de jabón es la elección obvia para la respuesta en un experimento para mejorar el problema del sobrepeso. Examinando más de cerca al proceso, hay dos subprocesos que tienen una relación directa con el peso de la barra de jabón: El proceso de mezcla que afecta la densidad de la barra de jabón y el proceso de formación que impacta las dimensiones de las barras de jabón. Con el fin de comprender mejor el mecanismo que causa el problema del sobrepeso, la densidad de la barra de jabón y sus dimensiones son elegidas como las respuestas. Aunque el peso de la barra de jabón no es usado como una respuesta, puede ser fácilmente determinada a partir de su densidad y sus dimensiones. Por tanto, no se pierde ninguna información al estudiar la densidad y las dimensiones. Tal estudio puede revelar nueva información sobre los subprocesos de mezcla y formación, lo cual puede conducir a la vez a una mejor comprensión del problema del sobrepeso.

Las respuestas elegidas pueden ser clasificadas de acuerdo al objetivo establecido.

Se pueden considerar tres amplias categorías: La nominal mejor, mientras más grande mejor, mientras más pequeña mejor.

1.2.3. Elegir factores y niveles

^{cantidad de comida}
Un factor es una variable que es considerada en el experimento, con el fin de estudiar su efecto sobre la variable respuesta. Para esto, deben usarse al menos dos valores del factor. Estos valores son referidos como ^{3 tipos de comida} niveles del factor. Un tratamiento es la combinación de niveles de dos o más factores. ^{variables} Cuando hay un solo factor sus niveles son los tratamientos.

Para el éxito del experimento es crucial identificar los factores potencialmente importantes en la etapa de planificación del experimento. Hay dos métodos gráficos para identificar factores potenciales. El primero, un diagrama de flujo del sistema o proceso es útil para ver dónde los factores aparecen en un proceso de múltiples etapas. En la Figura 1.1 se presenta un bosquejo aproximado de un proceso de fabricación de pulpa de papel, el cual incluye las materias primas de los proveedores y un proceso químico para hacer una mezcla la cual es pasada por un proceso mecánico para producir la pulpa. Incluir a todos los involucrados en el proceso es muy importante para capturar una descripción exacta del proceso o sistema.

El segundo diagrama es conocido como diagrama de causa - efecto y puede usarse para listar y organizar los factores potenciales que pueden impactar a la respuesta. En la Figura 1.2 se presenta un diagrama de causa efecto el cual lista los factores que se piensan afectan la calidad de un proceso de moldeo por inyección.

Generalmente los factores son clasificados en: Hombre, Máquina, Medición, Material, Método. Debido a su apariencia, los diagramas de causa - efecto también son llamados diagramas de *espina de pescado*. Las diferentes características de los factores necesitan ser reconocidas porque ellas pueden afectar la elección del diseño experimental. Por ejemplo, un factor tal como la temperatura del horno es *difícil de cambiar*. Es decir, después de cambiar el ajuste de la temperatura, puede tomar una considerable cantidad de tiempo antes que la temperatura se estabilice en el nivel de

ajuste.

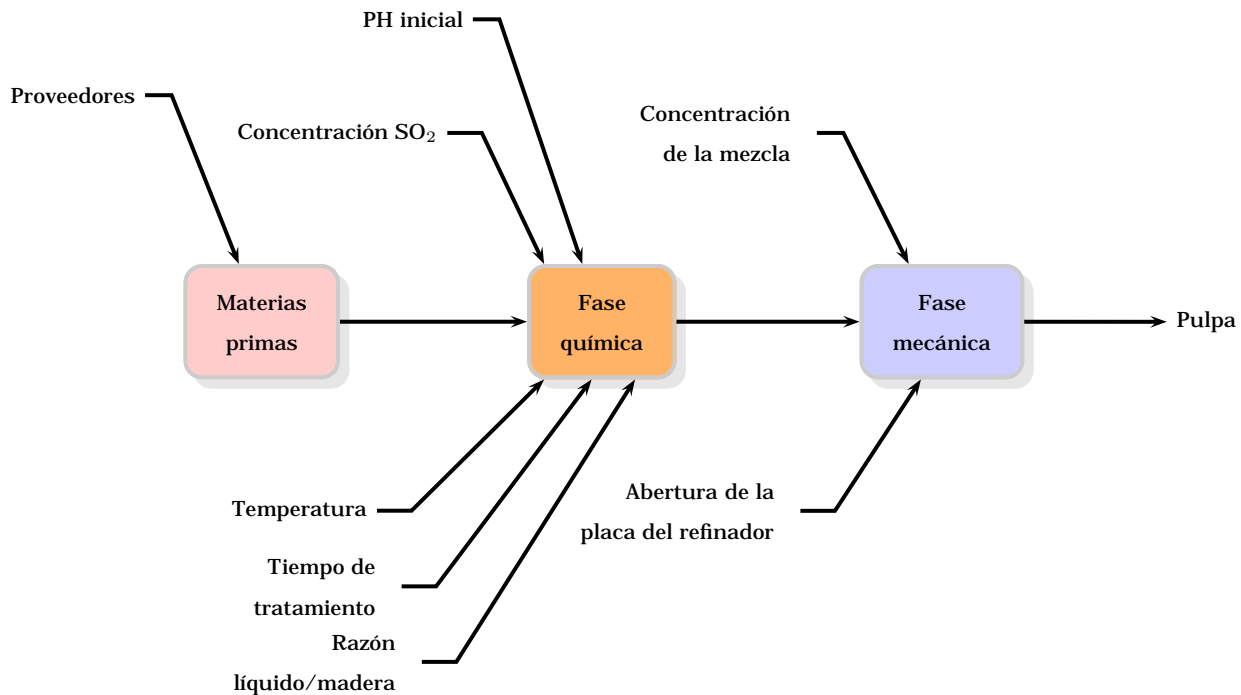


Figura 1.1: Diagrama de flujo proceso de fabricación de pulpa de papel. Adaptado de [14].

Un factor también puede ser *difícil de ajustar* así que el nivel actual usado en el experimento puede ser diferente del nivel previsto. Por ejemplo, la fuerza de impacto actual de un perdigón lanzado contra el parabrisas de un automóvil sólo puede ajustarse dentro de 3 psi de la fuerza de impacto prevista (la fuerza prevista $\pm 3\text{psi}$). Otros factores que pueden ser difíciles o imposibles de controlar son denominados *factores de ruido*. Ejemplos de factores de ruido incluyen las condiciones ambientales y de uso por parte de los consumidores. En general, un factor de ruido es aquél cuyos valores son difíciles de controlar durante el funcionamiento normal del proceso o en condiciones de uso, dentro de esta categoría pueden incluirse:

variación en parámetros de proceso, variación en parámetros de producto, variación ambiental, factores de carga, variación por efectos de pasos previos del proceso (upstream variation), variaciones en condiciones de usos, variación de unidad a unidad o espacial, variación en el tiempo (en escalas cortas de tiempo como mañana a tarde, días en la misma semana, o varias semanas), degradación o deterioro lento de condiciones

o calidad en un periodo largo de tiempo.

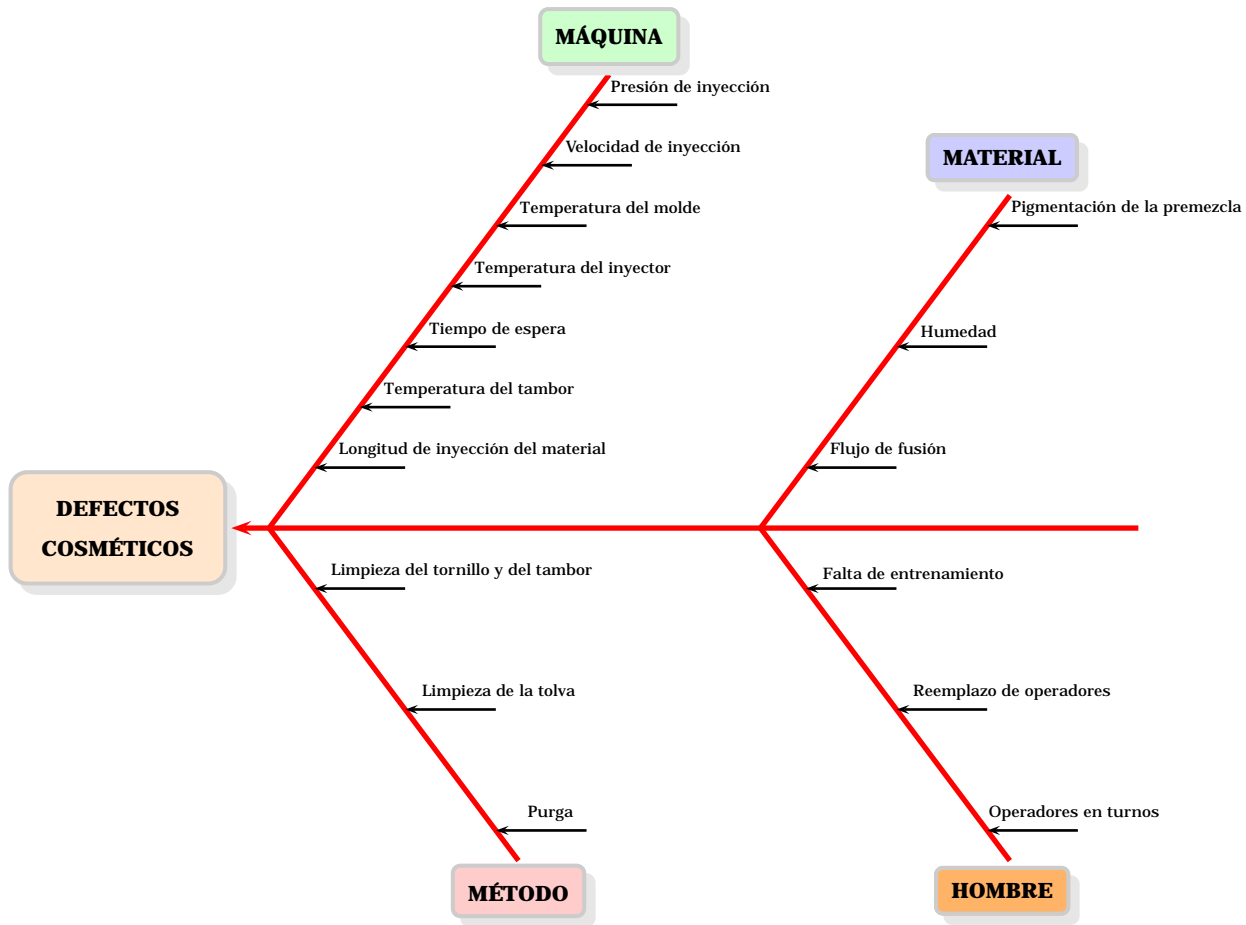


Figura 1.2: Diagrama causa - efecto, experimento del moldeo por inyección. Adaptado de [14].

Por otra parte, los factores pueden ser cuantitativos o cualitativos. Los factores cuantitativos como temperatura, tiempo, y presión toman valores sobre un rango continuo. Los factores cualitativos toman un número discreto de valores, por ejemplo los factores modo de operación, proveedor, posición, línea de producción, etc.

Para los factores cuantitativos es necesario conocer el rango (min, max.) en el cual se puede mover el factor; aparentemente uno podría elegir cualquier número de valores en dicho rango como niveles para el experimento. Por ejemplo, si la temperatura (en °C) está en el rango de 100-200°C, uno podría elegir 130 y 160°C para un experimento con dos niveles, o 125, 150, y 175°C para tres niveles. Si sólo se espera un efecto lineal (la relación entre la respuesta y el factor es lineal), dos niveles resul-

1.2. Aproximación sistemática a la planeación e implementación de experimentos 11

tan suficientes. Pero si se espera una curvatura, entonces se requieren tres o más niveles del factor. En general los niveles de factores cuantitativos deben elegirse lo suficientemente separados de manera que pueda detectarse un efecto, pero no demasiado separados de modo que diferentes mecanismos físicos estén involucrados (lo cual podría dificultar hacer modelación y predicción estadística).

Hay menos flexibilidad al elegir los niveles de factores cualitativos. Suponga que hay tres métodos de ensayo bajo comparación. Todos los tres deben ser incluidos como los tres niveles del factor *método de ensayo*, a menos que el investigador quiera posponer el estudio de un método, así que sólo dos métodos son comparados es un experimento con dos niveles de este factor.

Cuando hay flexibilidad en la elección del número de niveles, la elección puede depender de la disponibilidad de planes experimentales para la combinación dada de niveles de factores. Al elegir factores y niveles, deben considerarse los costos y las restricciones prácticas. Los niveles de los factores deben elegirse cumpliendo las restricciones prácticas. Si una combinación de factores (por ejemplo, alta temperatura y largo periodo de tiempo en una estufa) puede conducir a resultados desastrosos ésta debería evitarse y elegir un plan de experimentación diferente.

1.2.4. Elegir el plan experimental

Para ello debe usarse los tres principios fundamentales del diseño de experimentos (**aleatorización, replicación y bloqueo**) como también otros principios según cada modelo. La elección del plan experimental es crucial. Un diseño pobre puede capturar muy poca información que ningún análisis puede rescatar. De otra parte, si el experimento es bien planeado, los resultados pueden ser tan obvios que no se requieren análisis sofisticados. Ver Sección 1.3.

1.2.5. Ejecutar el experimento

Se recomienda el uso de una **matriz de planeación**. Esta matriz describe el plan experimental en términos de los valores actuales o ajustados de los factores. Por ejemplo, si el factor es presión, una lista de los niveles actuales podría ser 50 o 70 psi. Para

evitar confusión y eliminar problemas potenciales de correr una combinación errónea de niveles de los factores, cada uno de los tratamientos debería especificarse en hojas de instrucción aparte para darlos al personal que ejecutará el experimento. También vale la pena llevar a cabo una corrida de prueba para ver si habrán dificultades al correr el experimento, concretamente, si hay problemas con el ajuste de los factores y la medición de las respuestas. Cualquier desviación del experimento planeado debe ser registrada. Por ejemplo, si hay dificultad para ajustar los factores, entonces los valores reales deberían ser registrados.

1.2.6. Analizar los datos

Los datos deben ser analizados de acuerdo al diseño usado en la recolección de los datos. Este análisis **incluye el ajuste del modelo** y la **validación de sus supuestos a través del análisis de residuales**.

1.2.7. Extraer conclusiones y hacer recomendaciones

Con base en el análisis de los datos, deben presentarse conclusiones en las cuáles se incluyan los factores importantes y el modelo para la respuesta en términos de estos factores. También deben indicarse los ajustes o niveles recomendados para estos factores. Las conclusiones deben retomar los objetivos establecidos para el experimento. También es recomendable realizar un experimento confirmatorio para verificar los ajustes recomendados. Además, pueden presentarse recomendaciones tales como realizar experimentación adicional en un proceso de experimentación secuencial, especialmente en el caso de optimización de un proceso.

1.3. Contenido del plan experimental

De acuerdo a [4], el informe de un plan experimental debe responder a las siguientes siete consideraciones de la planeación experimental.

a) Localizar un problema “importante” susceptible de abordarlo con un diseño experimental:

Esto implica,

- Descripción inicial del problema, su importancia y descripción del proceso en el que se presenta (diagrama de flujo del proceso y localización del problema).
- Establecer cómo se relacionan las variables que intervienen.
- Idea de cuáles variables respuesta son importantes.
- Cuáles factores pueden influir.
- Medir situación actual y especificar qué se ha hecho para atenuar el problema.
- Determinar cómo medir el éxito del proyecto.

b) Determinar los factores de estudio según su supuesta influencia sobre la variable respuesta:

- Diagrama Ishikawa
- Lista de todas las posibles causas (factores)
- Posibles causas (factores) principales
- Ante muchos factores, pueden descartarse algunos usando conocimiento e información previa, y/o diseños factoriales fraccionados saturados.
- Realizar experimentos más informativos en la medida que se reduce el número de factores.

c) Decidir y resumir los niveles de prueba de cada factor: Importa la separación entre niveles:

- Muy juntos no permiten detectar el efecto del factor. Muy separados pueden causar problemas de operación.
- Recurrir a la física del proceso para obtener condiciones de prueba realmente diferentes.
- Si no se pueden fijar en forma exacta, los niveles pueden ser “rangos de variación” disjuntos cuyo centro es el “nivel teórico”.

- Para factores cualitativos seleccionar al menos dos situaciones representativas.
- Ante sospecha fuerte de “curvatura” usar más de dos niveles.
- Si no se sabe cuáles niveles, hacer corridas preliminares con dos niveles (criterio de separación al menos 1.5σ entre medias).

d) Elegir variables respuesta a medir:

- Las que mejor reflejen el problema en cada punto del diseño.
- Las que puedan resultar afectadas por cambios en los factores controlados aunque no sean parte del problema.
- Verificar que se pueden medir de modo confiable: equipos y métodos de medición capaces de repetir y reproducir una medición, con precisión (error) y exactitud (calibración) necesarias. Si es necesario, repetir la medición y reportar el promedio, realizar estudio R & R.

e) Seleccionar el diseño experimental adecuado a factores y objetivos:

- Restricciones en la aleatorización, o en la aplicación de ciertos tratamientos.
- Determinación de replicaciones necesarias de tratamientos, tamaño de prueba en cada condición experimental.
- Proponer primer diseño.
- Investigar por factores de ruido o de bloqueo.
- Proponer diseño final incluyendo aleatorización, replicaciones, restricciones costos, tiempos y objetivos.

f) Planear y organizar el trabajo experimental según diseño:

- Aplicar principios de “aleatorización”, “replicación” y “bloqueo”.
- Definir personas, tareas, responsabilidades, e instrucciones específicas.

- Diseñar hojas de trabajo: Especificar cómo hacer pruebas, orden de corridas, forma de medición y recolección de datos.
- Completar detalles logísticos: Días, horas, materiales a usar, máquinas y puestos donde se harán las pruebas.
- Prever contingencias.
- Si es posible, realizar ensayos para afinar detalles de instrucciones e imprevistos.

g) Bosquejar análisis estadístico para resultados experimentales:

- Software,
- procesamiento datos,
- hipótesis a probar,
- análisis a realizar,
- tipo de conclusiones o decisiones esperadas.

1.4. Los tres principios fundamentales de la experimentación

De acuerdo a [2, 14], hay tres principios que deben considerarse en la experimentación: **la replicación, la aleatorización y el bloqueo.**

Una *unidad experimental* es un término genérico que se refiere a una unidad básica tal como material, animal, persona, máquina o periodo de tiempo, a la cual un tratamiento es aplicado. El término **replicación** significa que cada tratamiento es aplicado a unidades experimentales que son representativas de una población de unidades sobre las cuales las conclusiones del experimento serán aplicadas. Esto facilita la estimación de la magnitud del error experimental (es decir, de la varianza del error) contra el cual se juzgan las diferencias entre los tratamientos. Si hay t tratamientos y t unidades experimentales, puede llevarse a cabo un experimento y la media de cada tratamiento puede estimarse de los datos. Pero una estimación de la

varianza del error no puede ser obtenida a menos que algunos o todos los tratamientos sean *replicados*. Al incrementar el número de replications o réplicas, disminuye la varianza de las estimaciones de los efectos de los tratamientos y proporciona más potencia para detectar diferencias entre tratamientos.

Desde que una *replicación* de un tratamiento es una *observación independiente del tratamiento*, entonces dos replications de un tratamiento deben involucrar a dos unidades experimentales, y por esto es necesario realizar una distinción entre replications y repeticiones. Muy a menudo los investigadores usan *muestras duplicadas o partidas* para generar dos observaciones y las llaman réplicas cuando en realidad son *submuestras o medidas repetidas y no replications*. Por ejemplo, dos medidas independientes de la altura de una persona no proporcionan una medida de la verdadera variación en las alturas de la población; tan sólo son dos submuestras o repeticiones. En general, lecturas o medidas sobre la misma unidad experimental son repeticiones mientras que lecturas cada una sobre unidades experimentales separadas son replications. La varianza del error de las primeras es menor que la varianza del error de las últimas puesto que lecturas repetidas sólo miden la variación debida a los errores en las lecturas mientras que las últimas además miden la variación de unidad a unidad. Una subestimación de la verdadera varianza del error puede resultar en la detección falsa de significancia de efectos [14].

Ejemplo [9]: Considere un experimento para comparar las habilidades de tres conservantes para inhibir el crecimiento de hongos sobre cierto tipo de torta. El panadero hace una torta con cada conservante. Después de nueve días de almacenamiento, se mide el número de esporas de hongos por cm^3 de torta. El panadero quería 10 replications para el análisis, así que dividió cada torta en 10 porciones y obtuvo el conteo de esporas en cada parte. Sin embargo, esas 10 mediciones no resultaron de 10 aplicaciones independientes del conservante. La variación medida por sus submuestras es un índice de la variación dentro de la torta y no de la variación de unidad experimental a unidad experimental. Para tener 10 replications, el panadero necesita batir 10 tortas con cada conservante, cada una mezclada independientemente de las demás.

El principio de la *aleatorización* debe aplicarse a la asignación de las unidades

experimentales a los tratamientos, al orden en el cual los tratamientos son aplicados al desarrollar el experimento y al orden en el cual se miden las respuestas. Lo anterior proporciona protección contra variables desconocidas y que pueden impactar la respuesta. También se reduce la influencia no deseada del juicio subjetivo en la asignación de los tratamientos y más aún, la aleatorización asegura la validez de la estimación del error experimental y proporciona una base para la inferencia al analizar los experimentos [14]. Considere por ejemplo, un experimento que involucra t tratamientos en el cual cada uno de estos es aplicado a r unidades experimentales diferentes. Un modelo matemático que puede usarse para describir la respuesta observada en la j -ésima unidad experimental del i -ésimo tratamiento, es, [9]:

$$y_{ij} = \mu_i + \xi_{ij} \quad i = 1, 2, \dots, t, \quad j = 1, 2, \dots, r. \quad (1.1)$$

Donde μ_i es la verdadera media desconocida de la respuesta del i -ésimo tratamiento y ξ_{ij} es un ruido resultante de la variación natural y de otras posibles fuentes de error aleatorio y no aleatorio.

De acuerdo a [9], para llevar a cabo este experimento, el experimentador debe seleccionar rt unidades experimentales y luego asignar aleatoriamente cada tratamiento a r de las unidades experimentales. *La aleatorización, la cual es muy importante, es la base para el desarrollo de la teoría para un análisis correcto.* Una correcta aleatorización previene la introducción de sesgo sistemático en el experimento. *Si no usamos la aleatorización, no podemos decir si una diferencia observada es debida a las diferencias entre los tratamientos o debida al método sistemático usado para asignar los tratamientos a las unidades experimentales.*

El principio de *bloqueo* tiene que ver con el concepto de bloque. Sea r el número de réplicas necesarias y t el número de tratamientos. En la mayoría de los experimentos es imposible seleccionar rt unidades experimentales idénticas. La no similaridad entre las unidades experimentales contribuye al ruido ξ_{ij} . Así, los experimentos pueden ser mejorados si agrupamos las unidades experimentales dentro de grupos de unidades cercanamente similares, llamadas unidades *homogéneas*. Un bloque es un grupo de

unidades homogéneas, por ejemplo, días de la semana, semanas, mañana vs. tarde, lotes, parcelas, entre otros. Para que el bloqueo sea efectivo, las unidades experimentales deberían ser organizadas de manera que la variación dentro de bloque sea mucho más pequeña que la variación entre bloques. Según [9], como el objetivo de un experimento es comparar las respuestas observadas de los tratamientos sobre las unidades experimentales, mientras más parecidas sean las unidades experimentales, mejor serán las comparaciones entre los tratamientos. Al comparar los tratamientos dentro del mismo bloque, los efectos de bloques son separados en la comparación de los efectos de tratamientos, haciendo de tal modo el experimento más eficiente y permitiendo medir sin sesgos los efectos de los tratamientos. Por ejemplo, puede haber un efecto conocido del día sobre la respuesta de manera que si todos los tratamientos pueden ser aplicados en el mismo día, la variación de día a día es eliminada [14]. Se debe tener en cuenta que cuando se desean hacer inferencias de gran alcance, las condiciones experimentales deberían ser bastante variadas, sin embargo esto puede resultar en un incremento en la variabilidad de la respuesta y por tanto se hace necesaria la aplicación de la técnica de bloqueo [2]. Si el bloqueo es efectivo, debería aplicarse para remover la variación de bloque a bloque. La aleatorización puede ser luego aplicada en la asignación de los tratamientos a las unidades dentro de cada bloque para reducir más la influencia de variables desconocidas [14]. En general, cuando el bloqueo es bien realizado, las comparaciones de dos o más tratamientos resultan más precisas que aquellas comparaciones realizadas a partir de un diseño sin bloqueo [2].

Para entender cómo la inclusión de bloques afecta en el diseño, considere r bloques con t unidades experimentales, cada tratamiento ocurriendo en cada bloque. Un modelo que representa la respuesta observada del i -ésimo tratamiento en el j -ésimo bloque es:

$$y_{ij} = \mu_i + b_j + \xi_{ij}^* \quad i = 1, 2, \dots, t, \quad j = 1, 2, \dots, r. \quad (1.2)$$

Comparando las ecuaciones (1.1) y (1.2) vemos que el ruido en (1.1), ξ_{ij} , ha sido

reemplazado por $b_j + \xi_{ij}^*$, es decir, la variación entre grupos o bloques de unidades experimentales ha sido identificada y aislada de ξ_{ij}^* . Dos tratamientos cualesquiera de los t considerados pueden compararse, libres de los efectos de bloques, tomando las diferencias dentro de los bloques de las respuestas de los dos tratamientos:

$$y_{ij} - y_{kj} = \mu_i - \mu_k + \xi_{ij}^* - \xi_{kj}^* \quad (1.3)$$

que como vemos **no depende del efecto del bloque b_j** .

Recapitulando, un objetivo del diseño experimental es *seleccionar y agrupar el material experimental de modo que el ruido o error experimental en el experimento sea reducido. Así, las unidades experimentales sobre las cuales los tratamientos van a ser comparados deberían ser tan parecidas como sea posible para que la menor diferencia significativa entre dos tratamientos pueda ser detectada.*

Por último, de acuerdo a [14], una adición útil al diagrama de causa - efecto es indicar cómo el diseño experimental propuesto se dirige a cada factor listado. Se sugieren las siguientes designaciones: *E* para un factor experimental, *B* para un factor manipulado por bloqueo, *O* para un factor que es mantenido constante en un valor, y *R* para un factor manipulado por aleatorización. Esta designación indica claramente cómo el diseño propuesto trata con cada uno de los factores potencialmente importantes. La designación *O* sirve para recordarle al experimentador que el factor es mantenido constante durante el experimento actual pero puede ser variado en un experimento futuro. Ver la Figura 1.3.

También pueden usarse otras designaciones para los factores, por ejemplo, los factores experimentales pueden dividirse en dos tipos: factores controlables y factores de ruido, o también como factores “difícil de cambiar” y “fácil de cambiar”.

En la Figura 1.4 se presenta otra manera de esquematizar el proceso y sus variables para el diseño experimental. En la Figura 1.5 se ilustra la aplicación del análisis de los factores y variables en un proceso de fabricación de un envase plástico.

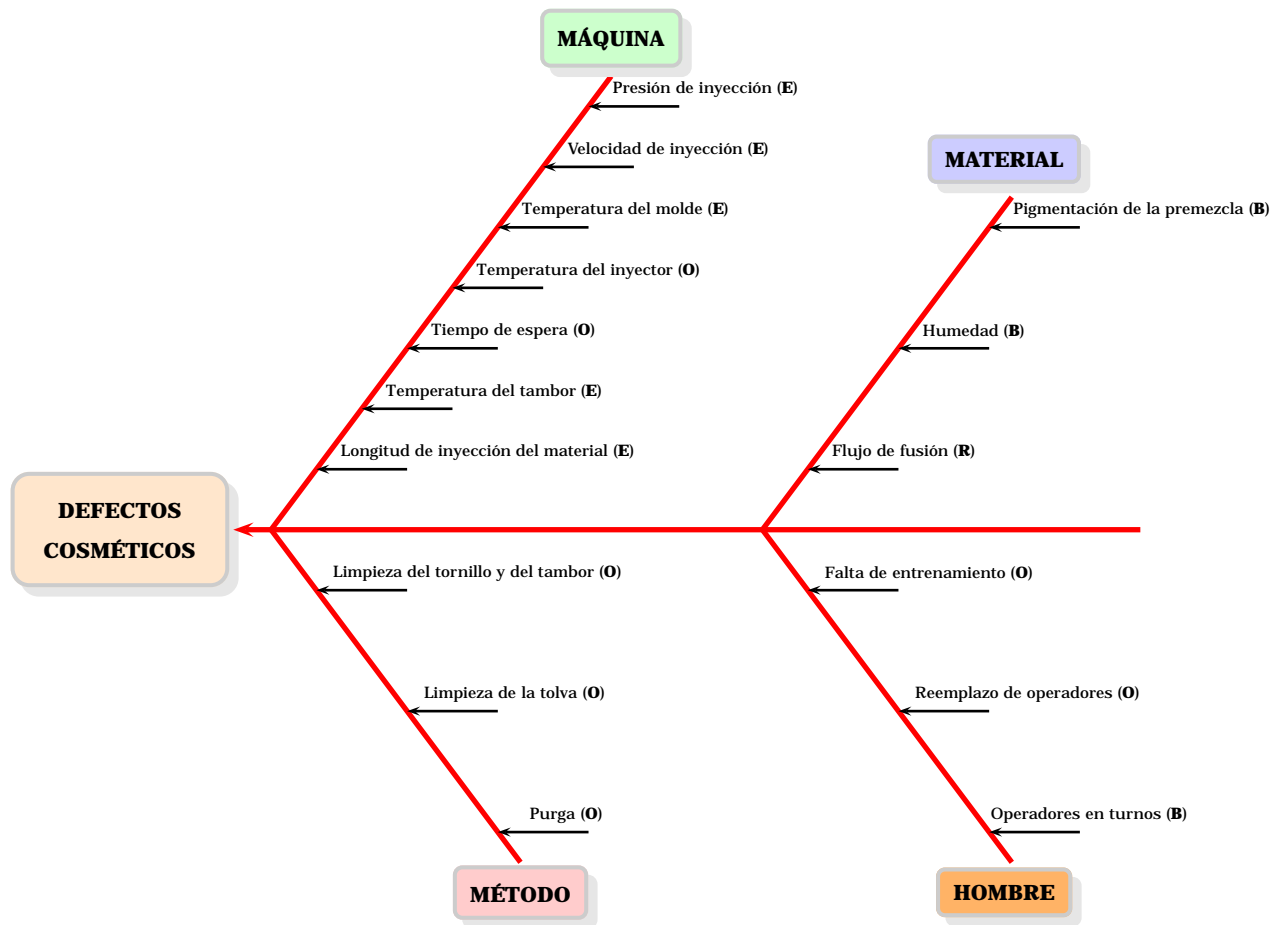


Figura 1.3: Diagrama de causa - efecto revisado, experimento del moldeo por inyección. Adaptado de [14].



Figura 1.4: Diagrama de Entrada - Proceso - Salida . Adaptado de [4].

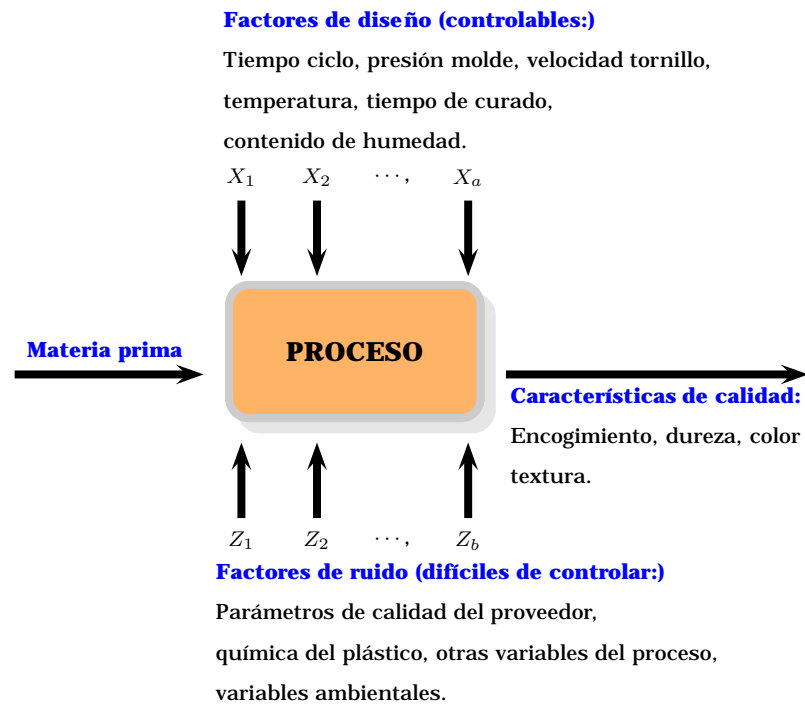


Figura 1.5: Diagrama de Entrada - Proceso - Salida en la fabricación de un envase plástico.
 Adaptado de [4].

1.5. Estructura de un diseño experimental

Esta Sección fue tomada de [9]. Un diseño experimental consiste de dos estructuras básicas, y es muy importante ser capaz de identificarlas y diferenciarlas, a saber:

- La estructura de tratamientos,
- La estructura de diseño.

y el diseño experimental comprende:

- La elección de la estructura de tratamientos,
- La elección de la estructura de diseño y el método de aleatorización.

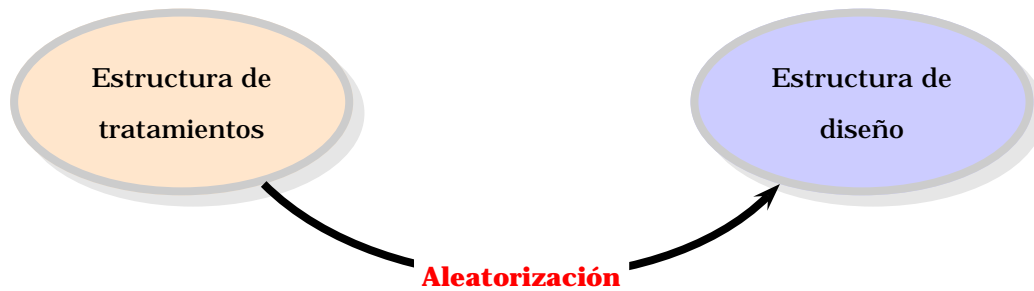


Figura 1.6: Estructuras de tratamiento y de diseño y su combinación en un DOE. Adaptado de [9].

1.5.1. Estructura de tratamientos

Consiste del conjunto de tratamientos, combinaciones de tratamientos, o poblaciones que el experimentador ha seleccionado para estudiar y/o comparar. Se describen a continuación algunas de las estructuras de tratamientos comunes, algunas también de las cuales serán estudiadas en el curso.

- Estructura de tratamiento de un sólo factor: Esta estructura consiste de t tratamientos o poblaciones donde no se asume relación entre tratamientos.
- Estructura de tratamientos de dos factores: Consiste del conjunto de tratamientos construido por combinación de los niveles de dos diferentes tipos de factores. El conjunto resultante de tratamientos, llamado *combinaciones de tratamientos*, es generado combinando cada nivel de un tipo de tratamiento con cada nivel del otro tipo de tratamiento. Si el primer tipo tiene a niveles y el segundo tiene b niveles, estos producen ab combinaciones de tratamientos, como por ejemplo el que se muestra a continuación, donde el factor A tiene $a = 3$ niveles y el factor B tiene $b = 4$ niveles, generando $ab = 12$ combinaciones de tratamientos, A_iB_j , $i = 1, 2, 3$, $j = 1, 2, 3, 4$:
- Estructura de tratamiento de arreglos factoriales, estructura de tratamientos 2^k y estructura de tratamientos factoriales fraccionados: Una estructura de tratamiento de arreglos factoriales consiste del conjunto de combinaciones de tratamien-

	Factor B			
Factor A	B_1	B_2	B_3	B_4
A_1	A_1B_1	A_1B_2	A_1B_3	A_1B_4
A_2	A_2B_1	A_2B_2	A_2B_3	A_2B_4
A_3	A_3B_1	A_3B_2	A_3B_3	A_3B_4

tos contruidos por combinación de los niveles de dos o más tipos diferentes de factores. La estructura de tratamiento de dos factores es también un arreglo factorial. En general, una estructura de tratamiento k -factorial es generada combinando k diferentes tipos de factores, cada uno con a_1, a_2, \dots, a_k niveles, respectivamente, dando $a_1 \times a_2 \times \dots \times a_k$ combinaciones de tratamientos. Cuando hay k factores y cada uno observado en dos niveles, la estructura de tratamientos se denomina 2^k . Una estructura de tratamiento de arreglos factoriales fraccionados Consiste de sólo una parte o fracción de las posibles combinaciones de tratamientos en una estructura de tratamiento de arreglo factorial. Hay muchas técnicas sistemáticas para seleccionar una fracción apropiada, la mayoría de las cuales dependen de los supuestos que el experimentador hace acerca de las interacciones entre los varios tipos de tratamientos en la estructura de tratamiento. Una estructura de tratamiento de arreglo de cuadrado latino involucra un arreglo factorial de tres factores con n tratamientos fila, n tratamientos columna y n tratamientos celda. El arreglo de cuadrado latino consiste de n^2 de las n^3 combinaciones de tratamientos posibles, o de una fracción n^2/n^3 -ésima de las n^3 combinaciones de tratamientos posibles. Un uso posible de este tipo de arreglo es cuando puede asumirse que no hay interacciones de dos factores ni de tres factores entre los tres tipos de tratamientos.

- Arreglos factoriales con **uno o más controles**: Las estructuras de tratamientos pueden ser generadas de muchas formas diferentes y a menudo incluyen la combinación de algunas de las estructuras anteriores. Por ejemplo, una estructura de tratamiento podría consistir de una estructura de un sólo factor con t

controles como niveles combinada con una estructura de tratamientos de un arreglo factorial de dos factores.

1.5.2. Estructura de diseño

Consiste del agrupamiento de las unidades experimentales dentro de grupos homogéneos o bloques de manera que las condiciones bajo las cuales son observados los tratamientos, son tan uniformes como resulte posible.

- Diseño completamente **aleatorizado**: Si todas las unidades experimentales son muy homogéneas, entonces sólo se necesita un único grupo o bloque de observaciones, y las unidades experimentales pueden ser asignadas a los tratamientos completamente al azar. Generalmente, los tratamientos son asignados a un número igual de unidades experimentales, aunque esto no es requerido. Esta estructura de diseño puede también ser usada cuando las unidades experimentales no son homogéneas y el experimentador no puede agruparlas en grupos más homogéneos.
- Diseño **bloqueado**: Si se requiere más de un grupo de unidades experimentales de manera que las unidades dentro de cada grupo son más homogéneas que las unidades experimentales entre grupos.
 - De bloques completos aleatorizados: Si hay t tratamientos, entonces la estructura de diseño de bloques completos aleatorizados consiste de bloques de t unidades experimentales cada uno. **Cada tratamiento es asignado aleatoriamente a una unidad experimental en cada bloque**. Si cada bloque consiste de $c \times t$ unidades experimentales, donde c es un entero, entonces **cada tratamiento puede ser asignado a c unidades** experimentales dentro de cada bloque. Esta es también una estructura de diseño de bloques completos aleatorizados. Una estructura de diseño de bloques completos aleatorizados es pues cualquier esquema de bloqueo en el cual el número de unidades experimentales dentro de un bloque es un múltiplo del número de tratamien-

tos, y así un conjunto completo de tratamientos puede ser asignado completamente al azar a c unidades experimentales en cada bloque.

- Cuadrados latinos: Esta estructura de diseño consiste del bloqueo en dos direcciones. Para un experimento que involucra t tratamientos, t^2 unidades experimentales son organizadas en un cuadrado de $t \times t$ donde las filas son llamadas *bloques fila* y las columnas son llamadas *bloques columna*. Así el arreglo de $t \times t$ unidades experimentales es bloqueado en dos direcciones. Para construir un diseño experimental cuadrado latino, los tratamientos son asignados aleatoriamente a las unidades experimentales en el cuadrado de manera que cada tratamiento ocurra sólo una vez en cada bloque fila y en cada bloque columna.
- De bloques incompletos: Estos diseños ocurren cuando el número de tratamientos excede al número de unidades experimentales en un bloque, de modo que un conjunto completo de tratamientos no puede ocurrir dentro de cada bloque.
- Varias combinaciones y generalizaciones: Hay varias maneras de agrupar a las unidades experimentales. Algunas veces un agrupamiento no satisface las definiciones anteriores pero todavía proporciona una estructura de diseño válida. Un ejemplo es cuando los tamaños de bloque varía de bloque a bloque o donde algunos bloques son incompletos mientras otros son completos. En cualquier caso, estos otros esquemas de bloqueo pueden proporcionar un experimento con estructuras de diseño muy viables.

1.6. Clasificación de los diseños experimentales

De acuerdo a [4], los diseños experimentales pueden clasificarse de la siguiente manera,

Para comparar dos o más tratamientos:

- Completamente aleatorizados

- En bloques completos e incompletos
- Cuadrados latinos y grecolatinos

Para estudiar el efecto de varios factores sobre una o más variables respuestas:

- Factoriales 2^k y 3^k
- Factoriales fraccionados 2^{k-p}

Diseños para optimizar procesos:

- Diseños para modelos de primer orden:
 - Diseños Factoriales 2^k y 2^{k-p}
 - Diseños de Plakett-Burman
 - Diseños simplex
- Diseños para modelos de segundo orden:
 - Diseño central compuesto
 - Diseño de Box-Behnken
 - Diseños factoriales 3^k y 3^{k-p}

Diseños robustos:

- Arreglos ortogonales (diseños factoriales)
- Diseño con arreglo interno y externo

Diseño de mezclas (mezclas de ingredientes o formulación):

- Diseño lattice-simplex
- Diseño simplex con centroide
- Diseño con restricciones
- Diseño axial

Diseños con unidades experimentales de distinto tamaño:

- Diseños experimentales de parcelas divididas (split - plot)
- Diseños de medidas repetidas

Estos diseños están contruidos de estructuras de diseño de bloques incompletos y estructuras de tratamiento de arreglos factoriales que incluyen dos o más factores o tipos de tratamientos.

1.7. La investigación experimental vs. DOE

Según [5], *la investigación experimental* es básicamente secuencial y en ésta un experimento conduce a otro, en la medida en que se va obteniendo cierto conocimiento sobre un proceso y que nuevas preguntas son formuladas. Vea Figura 1.7.

[5] dice que la mayoría de los experimentos son exploratorios y en este tipo de investigación el investigador está interesado en encontrar los factores que tienen influencia sobre los resultados de cierto proceso. Por ejemplo, uno puede estar interesado en si el grado de concentración de un compuesto químico, el tiempo de cocción, la temperatura de horneado, el grado de enfriamiento, y la cantidad de presión, tienen un efecto ya sea individual y/o conjunto sobre la ruptura de un tipo de plato de cerámica. El procedimiento obvio aquí, es variar los niveles de esos factores y comparar la ejecución de las diferentes combinaciones de niveles. Luego de un experimento exploratorio puede seguir un experimento confirmatorio, en el cual por ejemplo, se puede comparar el mejor procedimiento hallado en el experimento exploratorio, con el procedimiento actual y establecer cuál es mejor.

1.8. Consideraciones prácticas

De acuerdo a [4], las metodologías estadísticas por sí solas no garantizan investigaciones exitosas:

- El conocimiento no estadístico es vital: Se requiere buen conocimiento técnico y práctico del proceso para seleccionar mejor los factores y el diseño, y lograr el mejor análisis estadístico, interpretación y conclusión.

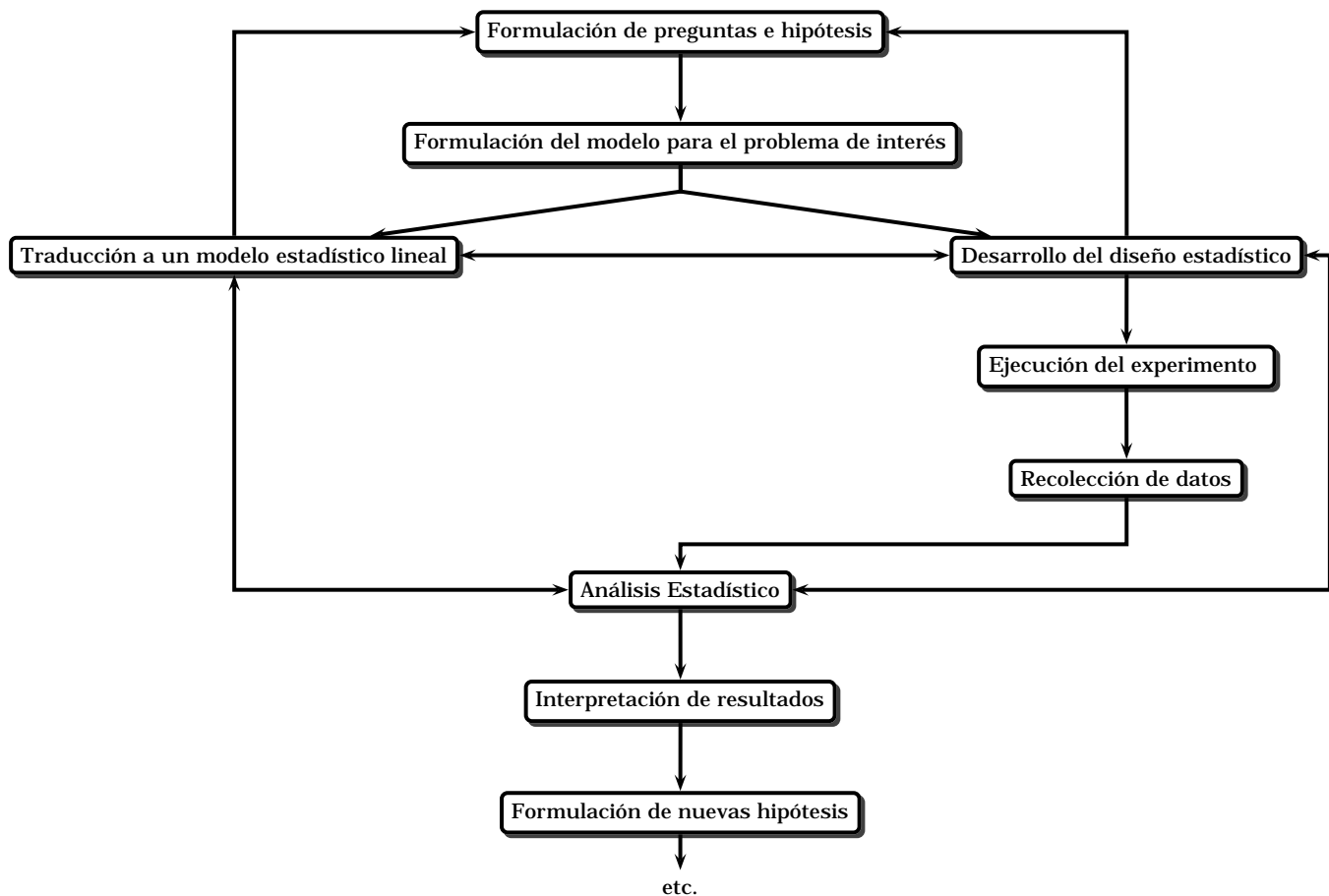


Figura 1.7: Procesos en la investigación experimental. Adaptado de [5].

- Reconocer entre significancia estadística e importancia práctica: Diferencias estadísticas no siempre corresponden a diferencias prácticas de importancia.
- Preferir la experimentación secuencial que un experimento único y definitivo: La experimentación secuencial alcanza paulatinamente mayor precisión en los conocimientos y soluciones. Seguir una estrategia bien definida y pensada. No confundir el término “secuencial” con “mover un factor a la vez”.

1.9. Qué se hace después del primer experimento

De acuerdo a [4], aunque se haga una buena planeación y ejecución de un experimento, a veces el estudio no es definitivo ni concluyente, o bien, se cometieron errores, o se quiere llegar a más y mejores decisiones. Qué hacer después del primer experimento depende de los resultados obtenidos. [4] establecen que existen al menos

los siguientes tres tipos de posibles acciones:

- a) El estudio es planeado para realizarse por etapas, ya sea por la gran cantidad de factores o porque así lo establece la metodología: Cuando se tienen mucho factores de estudio rara vez el primer experimento es definitivo, por lo general al inicio se corre un diseño **altamente fraccionado o saturado**, que permite estudiar sólo los **efectos principales** y no las **interacciones** entre factores, para detectar a los pocos factores influyentes y plantear posteriormente con estos un estudio más completo.
- b) El estudio se planea para desarrollarse por etapas, aunque no sean muchos los factores a considerar. Por cuestión de economía y eficiencia no se gastan todos los recursos experimentales de una vez. Se recomienda inicialmente correr un **diseño factorial fraccionado de resolución III ó IV**, y posteriormente si es necesario aclarar confusión de efectos, correr una fracción adicional. A veces también aunque son pocos los factores de estudio, en el análisis se encuentran dificultades en la determinación de la significancia de algunos efectos, por lo que luego se corre una réplica adicional para obtener más potencia en el experimento.
- c) Se comete algún error de planeación. Un error común es la poca planeación del experimento en el cual por lo general se incurre en seleccionar inadecuadamente **el ancho de los niveles** de los factores, lo cual luego sólo se corrige corriendo parte o todo el experimento de nuevo. Otro error es *no incluir un factor potencialmente importante*, y sobre todo cuando su inclusión no aumenta el número de corridas, ni el costo experimental.

Algunas de las acciones específicas después de un primer experimento:

- a) Agregar otra **fracción** en un **diseño factorial fraccionado** para eliminar dudas o ambigüedades en la interpretación de efectos en **experimentos de resolución III ó IV**.
- b) Reescalar, cuando no se tomó un espaciamiento adecuado entre niveles de uno o varios factores, lo cual puede implicar volver a correr todo el experimento de

nuevo. Un mal espaciamiento entre niveles impide detectar el efecto de un factor aún cuando éste si sea influyente.

- c) Quitar factores (colapsar) sobre los que se ha comprobado no afectan de ningún modo a la respuesta, o agregar otro, que no fue considerado en principio pero que ahora interesa estudiar. En el primer experimento no deberían dejarse por fuera ningún factor controlable que pueda afectar sólo o interactuando con otros factores.
- d) Obtener más replicaciones del experimento (todo o parcialmente) cuando no se ha podido ser concluyente sobre algún (o algunos) efecto (s). En el ANOVA tales efectos son reconocidos porque sus valores P están entre 0.05 y 0.1, y puede ser que sea debido a baja potencia de las pruebas estadísticas.
- e) Moverse de lugar. Es parte de la **metodología de superficie de respuesta**, para buscar condiciones óptimas de operación del proceso, donde más que hallar el mejor tratamiento en el primer experimento, se buscan las mejores maneras de operar el proceso. Del primer experimento mediante un modelo de regresión ajustado, se sacan conclusiones acerca de hacia qué dirección mover los niveles de los factores (se trabaja con pocos factores), corriendo puntos necesarios para ajustar un **modelo de primer orden** (sólo efectos principales), incluyendo en el diseño un punto central para detectar curvatura.
- f) Aumentar el experimento cuando se detecte **curvatura significativa**, y consiste en seleccionar en forma adecuada puntos adicionales y correrlos, para estudiar tanto los **efectos principales e interacciones** como **efectos cuadráticos puros**. Es necesario al menos tres niveles en cada factor para este tipo de estudios. Ej. **diseño central compuesto** que a partir de un **diseño factorial** que detectó curvatura agrega **puntos axiales**.

1.10. Qué hacer cuando ningún efecto es significativo

[4] establecen que a veces al correr un experimento puede suceder que ninguno de los factores estudiados resultan significativos, lo cual se refleja en un R^2 bajo. Algunas de las razones y posibles soluciones, son:

- a) Variación excesivamente alta en el proceso conduce a un MSE muy grande, haciendo que sea difícil detectar factores activos. En ese caso hay que evaluar qué otras fuentes de variación que quedaron en el término de error estuvieron activas durante el experimento. En caso de hallarlas, considerar un nuevo experimento donde se controlen o bloqueen tales fuentes de variación.
- b) Niveles de factores poco espaciados, de forma que resulta imperceptible captar variación entre respuestas medias. Se requiere reescalar y volver a correr el experimento. Si se eligen bien los niveles y un factor es realmente significativo, la dispersión en la variable respuesta durante el experimento debe ser mayor que la dispersión durante la operación normal del proceso. Una manera de verificar si los niveles son suficientemente amplios es comparar la variación observada durante el experimento contra la variación normal del proceso, y si la primera no es más grande de modo evidente, entonces puede que los niveles no fueron elegidos adecuadamente.
- c) Los factores elegidos no son los verdaderos responsables de la variación en la variable respuesta, por tanto hay otros factores no considerados que deben ser influyentes.
- d) Poca variación en la variable respuesta, bien sea por cualquiera de las dos situaciones anteriores, o porque el tamaño de la prueba fue insuficiente para detectar cambios importantes. Esto ocurre en particular cuando la variable respuesta es de conteo, E.j. la proporción de artículos defectuosos en cada corrida, que resulta normalmente pequeño, por lo que se requieren muchas piezas o unidades para observar algún producto defectuoso. Por tanto, si el tamaño de prueba no es sufi-

cientemente habrá que volver a correr el experimento para agregar mayor información.

- e) Si no se está satisfecho con los resultados de un experimento, antes de desecharlos hay que analizar qué pasó, investigar cuáles de las cuatro situaciones anteriores pudo presentarse. La experimentación debe ser un proceso de generación de información y de aprendizaje.

1.11. Ejemplos de DOE

Los siguientes ejemplos tomados del texto de Milliken, G. A. y Johnson, D. E. “*Analysis of Messy Data. Vol. 1: Designed Experiments*”. Van Nostrand Reinhold. New York, pp. 47-66, se ofrecen aquí con el fin de aclarar los conceptos sobre el diseño de experimentos y el análisis de varianza. No se dan fórmulas para el cálculo de los estadísticos en las tablas de análisis de varianza, pero posteriormente, una vez se avance en dichos tópicos en el curso, el estudiante estará en capacidad de establecer dichos cálculos.

1.11.1. Efectos de cinco dietas sobre la pérdida de peso

Un nutricionista quiere estudiar el efecto de cinco dietas sobre la pérdida de peso. La estructura de tratamientos de este experimento es una clasificación en una vía o de un sólo factor que involucra cinco tratamientos (las dietas). Muchas estructuras de diseño diferentes pueden ser seleccionadas. Si hay 20 personas homogéneas, entonces una estructura de diseño completamente aleatorizada puede ser usada donde cada dieta es asignada aleatoriamente a cuatro personas. Un modelo para la estructura de tratamiento de una sólo vía en una estructura de diseño completamente aleatorizada es:

$$y_{ij} = \mu_i + \xi_{ij}, \quad i = 1, 2, \dots, t, \quad j = 1, 2, \dots, n \quad (1.4)$$

Donde μ_i denota la media del i -ésimo tratamiento (dieta) y ξ_{ij} denota el error aleatorio. La tabla de análisis de varianza para el anterior modelo lista las fuentes de variación y los grados de libertad:

Fuentes de variación	gl
Dieta	4
Error	15

Si hay 10 hombres y 10 mujeres en vez de 20 personas homogéneas, el sexo de la persona podría ser usada como un factor de bloqueo; una estructura de diseño de bloques aleatorizados podría ser usada, donde cada dieta es asignada a dos hombres y dos mujeres. El modelo para una estructura de tratamientos de una vía en una estructura de diseño de bloques completos aleatorizados es:

$$y_{ij} = \mu_i + b_j + \xi_{ij}, \quad i = 1, 2, \dots, t, \quad j = 1, 2, \dots, n \quad (1.5)$$

Donde b_j denota el efecto del bloque j . La tabla de análisis de varianza para este modelo está dada por

Fuentes de variación	gl
Bloque (sexo)	1
Dieta	4
Error	14

En algunos casos el sexo de la persona no es una buena elección para un factor de bloqueo, dado que éste puede también ser un tipo de tratamiento. En ese caso la estructura de tratamiento es un arreglo factorial de dos vías o una estructura de tratamientos de dos vías consistiendo de las 10 combinaciones de tratamientos generadas por la combinación de los dos niveles de sexo con los cinco niveles de dieta. La estructura de diseño sería un diseño completamente aleatorizado con cada combinación de tratamiento observada dos veces.

Un modelo para una estructura de tratamiento de dos vías en una estructura de diseño completamente aleatorizada es:

$$y_{ijk} = \mu_{ij} + \xi_{ijk}, \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, n \quad (1.6)$$

Donde μ_{ij} es la media de la combinación del sexo i y de la dieta j . Algunas veces la

media μ_{ij} es expresada como

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad (1.7)$$

donde μ es la media total, α_i es el efecto del i -ésimo sexo, β_j es el efecto de la j -ésima dieta y γ_{ij} es el efecto interacción. Las tablas de análisis de varianza para el modelo (1.6) para ambas expresiones de μ_{ij} son dadas a continuación:

Fuentes de variación	gl
Modelo μ_{ij}	
Sexo*Dieta	9
Error	10
Modelo $\mu + \alpha_i + \beta_j + \gamma_{ij}$	
Sexo	1
Dieta	4
Sexo*Dieta	4
Error	10

Suponga ahora que las dietas tienen una estructura que consiste de una dieta de control y cuatro dietas hechas de las combinaciones de dos niveles de proteína y dos niveles de carbohidrato. La estructura de tratamientos de dietas es un arreglo factorial de dos vías con un control que, cuando es cruzada con el sexo de la persona, genera una estructura de tratamientos de tres vías con dos controles (uno para hombres y otro para mujeres). La estructura de diseño es completamente aleatorizado. El modelo es:

$$y_{ijk} = \mu_{ij} + \xi_{ijk}, \quad i = 1, 2, \quad j = 0, 1, 2, 3, 4 \quad k = 1, 2, \quad (1.8)$$

donde μ_{10} y μ_{20} denotan los controles y los μ_{ij} , $i = 1, 2$ y $j = 0, 1, 2, 3, 4$ denota el sexo de la persona por combinaciones de tratamiento de dieta. La tabla de análisis de varianza para el modelo (1.8) está dada a continuación:

Fuentes de variación	gl
Sexo	1
Dieta	4
Control vs. 2 ²	1
Proteína	1
Carbohidrato	1
Proteína*Carbohidrato	1
Sexo*Dieta	4
Sexo*Control vs. 2 ²	1
Sexo*Proteína	1
Sexo*Carbohidrato	1
Sexo*Carbohidrato*Proteína	1
Error	10

Donde 2² denota una comparación entre la dieta control y el promedio de las cuatro combinaciones de tratamientos proteína×carbohidrato.

1.11.2. Resistencia de cuatro pinturas de uso doméstico a condiciones ambientales

Una compañía de pinturas quiere comparar las habilidades de cuatro pinturas blancas para casas para resistir las condiciones ambientales. Cuatro casas cuadradas, cada una con una cara mirando exactamente hacia el norte, estuvieron disponibles para el experimento.

Cada lado de una casa está posiblemente expuesto a diferentes tipos de condiciones climáticas, así los lados (indicado aquí por las direcciones: norte, sur, este y oeste) de las casas pueden ser usadas también como factor de bloqueo. Dado que el número de tratamientos (las cuatro pinturas) eran iguales al número de bloques, se usó una estructura de diseño de cuadrados latinos, donde cada pintura puede ocurrir una y sólo una vez en cada casa y una y sólo una vez en cada dirección. Uno de tales arreglos de pinturas para casa y direcciones se muestra en la tabla siguiente (el esquema de aleatorización es asignar aleatoriamente las casas a los bloques fila y las

direcciones a los bloques columna).

Dirección				
Casa	N	S	E	W
1	A	B	C	D
2	D	A	B	C
3	C	D	A	B
4	B	C	D	A

Donde A, B, C y D denotan las cuatro pinturas.

El diseño experimental es una estructura de tratamientos de una vía en una estructura de diseño de cuadrado latino. Un modelo correspondiente es:

$$\mu_i + h_j + d_k + \xi_{ijk}, \quad i = 1, 2, 3, 4 \quad j = 1, 2, 3, 4 \quad k = 1, 2, 3, 4 \quad (1.9)$$

Donde μ_i denota la media para la pintura i , h_j denota el efecto de la j -ésima casa, d_k denota el efecto de la k -ésima dirección, y ξ_{ijk} denota el error. La tabla de análisis de varianza para el modelo (1.9) es:

Fuente de variación	gl
Casa	3
Dirección	3
Pintura	3
Error	6

Ahora suponga que las pinturas tienen una estructura dada por (a) pintura base, (b) base más aditivo I, (c) base más aditivo II, (d) base más aditivo I y más aditivo II. Esta es una estructura de tratamientos de dos vías donde un tipo de tratamiento es aditivo I con dos niveles, cero y algo, y el segundo tipo de tratamiento es aditivo II con dos niveles, cero y algo. Las cuatro combinaciones de tratamientos se muestran a continuación:

Aditivo II		
Aditivo I	Nada	algo
Nada	Base	Base + II
Algo	Base + I	Base + I + II

Un modelo para la estructura de tratamientos de dos vías en una estructura de diseño de cuadrado latino es:

$$y_{ijkm} = \mu + \gamma_i + \beta_j + (\gamma\beta)_{ij} + h_k + d_m + \xi_{ijkm} \quad (1.10)$$

Donde γ_i denota el efecto del aditivo I, β_j denota el efecto del aditivo II y $(\gamma\beta)_{ij}$ denota la interacción entre los dos aditivos. La tabla de análisis de varianza para el modelo (1.10) está dada por:

Fuente de variación	gl
Casa	3
Dirección	3
Pintura	3
I	1
II	1
I*II	1
Error	6

La única diferencia entre el análisis de los modelos (1.9) y (1.10) es que en (1.10) las pinturas tienen una estructura que es usada para particionar el efecto de la pintura entre los efectos debidos al aditivo I, el aditivo II, y la interacción de los aditivos I y II. La parte del análisis correspondiente a la estructura de diseño permanece igual, aunque el análisis de la estructura de tratamiento fue cambiada.

Finalmente, suponga que ocho casas estaban disponibles de modo que el experimento podía ser conducido usando dos estructuras de diseño de cuadrado latino repetidas. La siguiente tabla muestra una posible asignación de las pinturas a las combinaciones casa-dirección:

CASA								
C1					C2			
Dirección	1	2	3	4	5	6	7	8
N	C	A	B	D	D	C	A	B
S	D	B	C	A	C	B	D	A
E	A	C	D	B	A	D	B	C
W	B	D	A	C	B	A	C	D

Si las pinturas tienen una estructura de tratamientos de dos vías como la descrita en el caso de los aditivos, entonces el modelo está dado por

$$y_{ijkmn} = \mu + \gamma_i + \beta_j + (\gamma\beta)_{ij} + s_k + h_{km} + d_n + \xi_{ijkmn} \quad (1.11)$$

donde s_k denota el cuadrado k y h_{km} denota la casa m en el cuadrado k . La tabla de análisis de varianza está dada por:

Fuente de variación	gl
Casa	7
Cuadrado	1
Casa (cuadrado)	6
Dirección	3
Pintura	3
I	1
II	1
I*II	1
Error	18

1.11.3. Protección del acero contra el calor

Una estructura de diseño de cuadrado latino es muy útil cuando hay una necesidad de bloquear en dos direcciones. Pero un arreglo de cuadrado latino usado por los experimentadores no es una estructura de diseño de cuadrado latino. Este ejemplo muestra las consecuencias de usar una estructura de tratamiento de arreglo cuadrado latino.

Dos tipos de aditivos para pintura van a combinarse y placas de acero son pintados. El objetivo del experimento es estudiar la habilidad de las combinaciones de pintura para proteger al acero del calor. Hay cinco niveles de cada aditivo y cinco temperaturas en las cuales se chequeará la capacidad de protección.

Este experimento es acomodado en un arreglo cuadrado latino donde los niveles del aditivo I son las filas, los niveles del aditivo II son las columnas, y los niveles de temperatura son asignados a las celdas dentro del cuadrado. Este arreglo genera 25 combinaciones de tratamientos. Las unidades experimentales son 25 placas de acero de 0.2 cm de espesor y 1 m² de área. El proceso de aleatorización usado debería asignar aleatoriamente una de las 25 combinaciones de tratamiento a cada una de las placas de acero.

En este caso, la estructura de tratamiento es una fracción 1/5 de un arreglo factorial 5³ (como consiste de 25 de las 125 combinaciones de tratamientos posibles), llamada un arreglo cuadrado latino. La estructura de diseño es una asignación completamente aleatorizada, en la medida en que las combinaciones son asignadas completamente al azar a las placas de acero. Dado que este es un factorial fraccionado, cada uno de los efectos principales está parcialmente aliado con la interacción doble de los otros dos factores y con la interacción de los tres factores.

Con el fin de analizar apropiadamente este diseño experimental, deben hacerse algunos supuestos sobre los parámetros en el modelo. Los supuestos usuales son que no hay interacciones de dos o de tres factores. Sin embargo, deberíamos ser muy cuidadosos de no hacer tal supuesto sin tener alguna información previa (la cual puede provenir de otros experimentos, de la literatura existente, etc.) que muestre que tales interacciones, de hecho, son insignificantes. Uno de tales arreglos cuadrado latino se da en la siguiente tabla:

Nivel de aditivo I	Nivel de aditivo II				
	1	2	3	4	5
1	T_1	T_2	T_3	T_4	T_5
2	T_5	T_1	T_2	T_3	T_4
3	T_4	T_5	T_1	T_2	T_3
4	T_3	T_4	T_5	T_1	T_2
5	T_2	T_3	T_4	T_5	T_1

Un modelo para una estructura de tratamientos de cuadrado latino es:

$$y_{ijk} = \mu + A1_i + A2_j + T_k + \xi_{ijk} \quad (1.12)$$

Donde $A1_i$ denota el efecto del i -ésimo nivel del aditivo I, $A2_j$ denota el efecto del j -ésimo nivel del aditivo II, y T_k denota el efecto del k -ésimo nivel de temperatura. La tabla de análisis de varianza para este modelo es:

Fuente de variación	gl
Aditivo I	4
Aditivo II	4
Temperatura	4
Residual	12

El término *Residual* es usado en vez de *Error*, dado que la correspondiente suma de cuadrados involucra el error más cualesquiera efectos de interacción que puedan ser no nulos. Si no es correcto el supuesto de interacciones nulas, entonces el cuadrado medio residual será demasiado grande y el estadístico de la prueba F resultante será demasiado pequeño. Consecuentemente, si hay interacción en el experimento, esta no puede ser descubierta y cualesquiera efectos de tratamientos detectables pueden ser enmascarados (ocultos).

1.11.4. Efectos de los niveles de nitrógeno (N) y de potasio (K) sobre una variedad de maíz

Un modelo y el análisis resultante consiste de tres componentes básicos, denominados la estructura de tratamientos del modelo, la estructura de diseño y la estructura de error. Este ejemplo demuestra cómo pueden usarse estas tres componentes básicas.

Un cultivador quiere estudiar el efecto de tres niveles de nitrógeno y cuatro niveles de potasio sobre su variedad de maíz. Su estructura de tratamientos es un arreglo factorial de dos vías con 12 (3 niveles de N \times 4 niveles de K) combinaciones de tratamientos. Él tiene tres parcelas llamadas bloques, cada una de las cuales es dividida en 12 partes. Cada combinación de tratamientos es asignada aleatoriamente a una parte en cada parcela. Así, la estructura de diseño es un bloque completo aleatorizado, pues cada combinación de tratamiento ocurre una vez en cada bloque. El diseño experimental es una estructura de tratamiento de dos vías (factores) en una estructura de diseño de bloque completo aleatorizado. (Los bloques en un diseño de bloques completos aleatorizados son llamados replicaciones por algunos autores, sin embargo, aquí se prefiere llamarlos bloques o replicaciones bloqueadas con el fin de distinguirlos de las replicaciones en el diseño completamente aleatorizado).

El modelo para este ejemplo es:

$$y_{ijk} = \mu_{ij} + b_k + \xi_{ijk} \quad (1.13)$$

Donde μ_{ij} es la media del i -ésimo nivel de N con el j -ésimo nivel de K, b_k es el efecto del k -ésimo bloque, y ξ_{ijk} denota el error aleatorio. En general, cuando un modelo es construido, involucra la suma de las tres partes:

$$y = \text{Estructura de Tratamiento} + \text{Estructura de Diseño} + \text{Estructura de Error(es)} \quad (1.14)$$

De la misma forma, la tabla de análisis de varianza tiene tres partes. La tabla de análisis de varianza general para el modelo en (1.14) es:

Fuente de variación	gl
Estructura de Diseño	(gl)
Estructura de Tratamiento	(gl)
Estructura de Error	(gl)

La tabla de análisis de varianza para el modelo en (1.13) es:

Fuente de variación	gl
Diseño	
Bloques	2
Tratamientos	11
N	2
K	3
N*K	6
Error	22

En general una concesión debe ser hecha para posibilitar más de un término de error. Por ejemplo, los modelos split-plot y de medidas repetidas tendrán más de un término de error.

1.11.5. Bloques y replicaciones

En la mayoría de los textos sobre diseño de experimentos, o no se hace distinción entre bloques y replicaciones, o hay confusión sobre tal distinción. Este ejemplo se incluye para demostrar la diferencia entre los dos conceptos.

Suponga que el investigador quiere estudiar cuatro tratamientos en una estructura de tratamiento de una vía (un sólo factor) y sólo tiene dos unidades experimentales homogéneas por bloque. En este caso la estructura de diseño que debe ser usada es un diseño de bloques incompletos. Si hay suficientes bloques de modo que cada par de tratamientos puedan ocurrir juntos en un bloque el mismo número de veces, entonces es posible usar un diseño de bloques incompletos balanceados. Por ejemplo, los cuatro tratamientos podrían ser asignados a los bloques como se muestra en la

siguiente tabla:

Bloques	Tratamientos
1	1 y 2
2	1 y 3
3	1 y 4
4	2 y 3
5	2 y 4
6	3 y 4

En este caso, hay seis bloques en el diseño y cada tratamiento es replicado tres veces. Este punto señala que los bloques y las replicaciones no son el mismo concepto. Ellos son equivalentes únicamente para la estructura de diseño de bloques completos aleatorizados, donde cada tratamiento es observado una y sólo una vez en cada bloque. En este ejemplo, la estructura de diseño está asociada con los seis bloques, y no con las tres replicaciones que sólo pueden ocurrir. El modelo para el arreglo en la anterior tabla es:

$$y_{ij} = \mu_i + b_j + \xi_{ij} \quad (1.15)$$

donde el par (i, j) puede tomar sólo aquellos valores de combinaciones de tratamiento*bloque que son observados. La tabla de análisis de varianza corresponde a:

Fuente de variación	gl
Bloque	5
Tratamiento	3
Error	3

Los anteriores fueron sólo algunos ejemplos de experimentos comunes, sin embargo, hay muchas otras formas de construir diseños experimentales combinando varias estructuras de diseño y estructuras de tratamientos.

Capítulo 2

Generalidades de los Modelos ANOVA para DOE

Debe tenerse en cuenta que el modelo estadístico a usar para realizar un correcto análisis de los resultados debe ser elegido según el diseño experimental.

2.1. Supuestos básicos

Según [9], en la construcción del modelo estadístico en un DOE se hacen dos supuestos básicos:

1. Las componentes de la **estructura de diseño son efectos aleatorios**, es decir, los bloques usados son una muestra aleatoria de la población de bloques o de unidades experimentales.
2. Se asume que **no hay interacción entre** las componentes de la estructura de **diseño** con las componentes de la estructura de **tratamientos**, en otras palabras, se asume que las relaciones que existen entre los tratamientos serán consistentes de bloque a bloque (excepto por la variación aleatoria), o dicho de otra forma, que **los bloques no influenciarán las relaciones entre los tratamientos**. Sin embargo, **si pudiera ocurrir tal interacción, los factores de ruido o bloqueo deberían incluirse** en la estructura de tratamientos o los valores de estos factores de ruido podrían ser considerados como covariables con la posibilidad de coeficientes de regresión diferentes.

Por lo anterior, [9] afirma que la estructura de diseño debe ser seleccionada usando todo el conocimiento disponible sobre las unidades experimentales y de manera inde-

pendiente de la estructura de tratamientos, es decir, no permitir que la estructura de tratamientos determine el uso de una pobre estructura de diseño. Así mismo el experimentador debería seleccionar la estructura de tratamientos sin ningún conocimiento sobre la estructura de diseño, es decir, no dejar que esta última estorbe en la selección del conjunto necesario de tratamientos. Aunque es posible que después de que han sido seleccionadas la estructura de diseño apropiada y la estructura de tratamientos deseada, resulten necesarias algunas modificaciones en una u otra estructura para hacerlas compatibles y para permitir llevar a cabo un experimento efectivo.

2.2. Usos del análisis de varianza

Los modelos de análisis de varianza son usados básicamente para analizar los efectos de las variables explicatorias bajo estudio sobre la variable respuesta. Específicamente, los estudios de un sólo factor son utilizados para comparar los efectos de diferentes niveles del factor, para determinar “el mejor nivel”, y similarmente, en los estudios multifactoriales, los modelos de análisis de varianza se usan para determinar si los diferentes factores interactúan, cuáles factores son claves, cuáles combinaciones de los factores son “las mejores”, etc.

2.3. Modelo de análisis de varianza o ANOVA

Según [7], los modelos de análisis de varianza son herramientas estadísticas que permiten estudiar la relación entre una variable respuesta y una o más variables explicatorias o variables predictoras. Pero estos modelos no requieren ningún supuesto acerca de la relación estadística entre la variable respuesta y las variables predictoras, ni requieren que las variables predictoras sean cuantitativas.

2.3.1. Relación entre regresión y análisis de varianza

Siguiendo a [7], como es bien sabido, el análisis de regresión estudia la relación estadística entre una o más variables predictoras y una variable respuesta. Tanto las variables predictoras y la respuesta son cuantitativas en los modelos de regresión ordinarios (excepto en modelos con variables indicadoras). **La función de regresión**

establece la naturaleza de la relación entre la respuesta media y los niveles de las variables predictoras. En el análisis de regresión se hace el análisis de varianza para probar la significancia de la regresión, pero el análisis de varianza es una técnica más general. Los modelos de análisis de varianza son un tipo básico de modelo estadístico, que al igual que la regresión, están interesados en la relación entre una o más variables predictoras y una variable respuesta, también estos modelos son apropiados tanto para datos observacionales como datos experimentales y la variable respuesta es de tipo cuantitativo. Pero los modelos de análisis de varianza se diferencian de los modelos de regresión en lo siguiente:

1. Las variables predictoras o explicatorias pueden ser cualitativas (género, ubicación geográfica, tamaño de la planta, etc)
2. Si las variables predictoras son cuantitativas, no se hacen supuestos acerca de la naturaleza de la relación estadística entre ellas y la variable respuesta. Por tanto la necesidad de especificar la naturaleza de la función de regresión encontrada en el análisis de regresión ordinaria no se presenta en los modelos de análisis de varianza.

La Figura 2.1 ilustra las diferencias esenciales entre regresión y los modelos de análisis de varianza. Vemos que en el análisis de regresión se busca ajustar la curva que relaciona el volumen promedio de ventas (Y) a tres niveles dados del gasto en publicidad (X), en cambio, en el análisis de varianza los tres niveles de gastos en publicidad son tratados como poblaciones separadas, cada una conduciendo a una distribución de probabilidad de volumen de ventas. Las diferencias cuantitativas entre los tres niveles de gastos en publicidad y sus relaciones estadísticas con los valores esperados de volumen de ventas no son consideradas por el modelo de análisis de varianza.

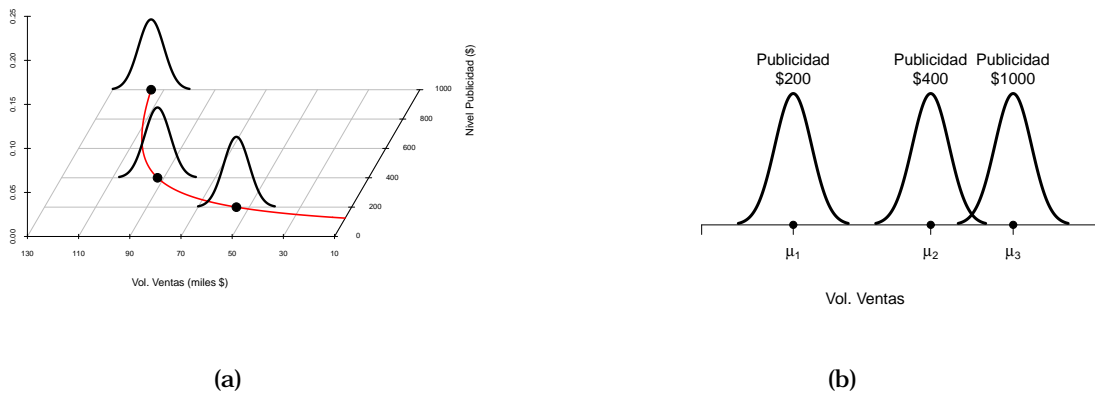


Figura 2.1: Relación entre regresión y ANOVA. Adaptado de [7]. (a) Punto de vista de la regresión: interesa la función que establece cómo cambia la media de la respuesta con los cambios en los valores del predictor lineal; (b) Punto de vista del modelo ANOVA: solo interesa que en los distintos niveles de valores del predictor lineal, existen diferentes poblaciones estadísticas de la respuesta que se diferencian en sus medias, pero no interesa modelar los cambios en estas medias como función de los valores del predictor lineal.

2.3.2. Elección entre regresión y ANOVA

Tal como establece [7], si las variables explicatorias son cualitativas, podemos trabajar con variables indicadoras en el análisis de regresión o recurrir a los modelos de análisis de varianza. Ambos métodos llevan a los mismos resultados, sin embargo, el análisis de varianza permite ciertas simplificaciones computacionales, pero de cualquier modo bajo estas circunstancias no es fundamental elegir entre regresión y un modelo de análisis de varianza. De otra parte, hay que elegir entre los dos tipos de modelación cuando las variables predictoras son cuantitativas. Una posibilidad es reconocer explícitamente la naturaleza cuantitativa de las variables predictoras y esto sólo puede hacerse con un modelo de regresión. La otra posibilidad es establecer clases o rangos para cada variable cuantitativa y emplear bien sea variables indicado-

ras en un modelo de regresión o un modelo de análisis de varianza. La estrategia de establecer clases para las variables predictoras cuantitativas es algunas veces usada en estudios a gran escala como un medio para obtener un ajuste de regresión no paramétrico cuando hay duda sustancial acerca de la naturaleza de la relación estadística. En este caso el análisis de regresión y los modelos de análisis de varianza no conducen al mismo resultado.

En general, los modelos de análisis de varianza son aplicados cuando es de interés estudiar los efectos de una o más variables predictoras sobre una variable respuesta.

Las variables predictoras son denominadas factores o tratamientos.

2.3.3. Modelo lineal de la variable respuesta relativo a modelos ANOVA

En los modelos ANOVA se asume que cada variable respuesta es medible y está representada matemáticamente en un modelo lineal. La idea general del modelo es expresar las observaciones generalmente denotadas por Y , en términos de “efectos” que contribuyen a Y , los cuales se pueden clasificar en tres categorías:

- Efectos de Tratamiento
- Efectos de Diseño
- Efectos de Error

Los efectos de tratamientos son un reflejo de los efectos de los tratamientos simples o de las combinaciones de factores. Los efectos de diseño son determinados por el diseño para el control del error, en particular, los efectos debidos a las varias clases de bloqueo. Los efectos de error representan diferentes clases de variación aleatoria. Según [4], el error aleatorio es la variabilidad observada que no se puede explicar por los factores estudiados, sino que es debida a “causas comunes o aleatorias” que generan la “variabilidad inherente del proceso”. Incluye efectos pequeños de factores no estudiados, variabilidad de las mediciones hechas bajo mismas condiciones, y el error experimental. El error experimental representa los errores del experimentador en la planeación y ejecución del experimento. Si el factor o factores estudiados son

influyentes sobre la variable respuesta, se espera que la variabilidad observada en ésta durante la experimentación sea explicada significativamente por tales factores y en menor medida por el error aleatorio, y que éste último sea realmente aleatorio.

El modelo ANOVA es lineal desde que la variable respuesta es una función lineal de los parámetros del modelo.

2.3.4. Modelo ANOVA I (efectos fijos) vs. Modelo ANOVA II (efectos aleatorios)

Siguiendo a [7], para establecer con simplicidad qué son y en qué se diferencian estos modelos, considere el caso de un sólo factor. Podemos distinguir dos modelos ANOVA: Modelo ANOVA I y modelo ANOVA II. En el primero sólo estamos interesados en considerar y concluir para un conjunto fijo de niveles del factor, en cambio en el segundo, los niveles considerados constituyen una muestra aleatoria tomada de la población de niveles del factor, con el fin de generalizar los resultados a toda la población de niveles.

Supuestos básicos

En el modelo ANOVA I con un sólo factor se tienen las siguientes consideraciones:

1. Para cada nivel del factor hay una distribución de probabilidad asociada para la variable respuesta,
2. Todas las distribuciones de probabilidad son normales,
3. Todas las distribuciones de probabilidad tienen la misma varianza,
4. Las respuestas en cada nivel del factor son selecciones aleatoriamente de la correspondiente distribución de probabilidad y son independientes de las respuestas en cualquier otro nivel del factor.

Bajo estas condiciones las distribuciones de las respuestas en cada nivel del factor sólo difieren con respecto a sus medias. Las diferencias en las medias reflejan por tanto los efectos esenciales de los niveles del factor. Este modelo es conocido como modelo de efectos fijos.

Por el contrario, en el modelo ANOVA II, aunque continúa operando los supuestos de normalidad, con un sólo factor las medias de sus niveles son en sí variables aleatorias y la variabilidad entre estas medias es un parámetro de interés. La varianza total de la respuesta es descompuesta en dos componentes una de las cuales es debida a la variabilidad en las medias de niveles del factor. Las respuestas observadas en niveles diferentes son independientes, sin embargo, las respuestas en un mismo nivel no son independientes. Este modelo es también denominado de componentes de varianza o modelo de efectos aleatorios.

Capítulo 3

Experimentos con un sólo factor de efectos fijos

3.1. Introducción

Los experimentos con un sólo factor comprenden experimentos que se usan cuando el objetivo es comparar más de dos tratamientos pero que corresponden a niveles de un mismo factor, con el fin de tomar una decisión en la solución de un problema real:

- Comparar dos o tres máquinas
- Comparar varios procesos para la obtención de un producto o resultado dado
- Comparar varios materiales,
- Comparar varias dietas
- Etc.

Por lo general las comparaciones se hacen en términos de las medias poblacionales, aunque también es importante la comparación de varianzas y capacidad actual para cumplir ciertas especificaciones. Las estructuras de diseños pueden ser como se indica en la Tabla 3.1.

3.2. Diseño completamente aleatorizado o DCA

Es el más simple de todos los diseños pues contempla sólo dos fuentes de variabilidad: los tratamientos y el error aleatorio. Se denominan completamente aleatorizados

Tabla 3.1: Estructuras de diseño en experimentos con un sólo factor de efectos fijos, y modelos estadísticos correspondientes

Diseño	Factores de bloque	Método estadístico	Modelo estadístico
Completamente aleatorizado	0	ANOVA con un solo criterio de clasificación	$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
En bloques completos aleatorizados	1	ANOVA con dos criterios de clasificación	$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$
En cuadrados latinos	2	ANOVA con tres criterios de clasificación	$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$
En cuadrados grecolatinos	3	ANOVA con cuatro criterios de clasificación	$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$
Notación: Y es la variable respuesta; μ es la media global; α_i es el efecto del i -ésimo tratamiento; β_j , γ_k y δ_l , son los efectos de bloques y ε es el error aleatorio			

porque las corridas experimentales se realizan en orden aleatorio completo, dado que no hay restricciones impuestas por factores de bloqueo.

3.2.1. Aleatorización

Para lograr que el diseño sea completamente aleatorizado es necesario un mecanismo para tal fin. A continuación se explica un procedimiento sugerido por [2] que puede ser realizado a mano o con software que incluya una función de generación de números aleatorios de una distribución $U(0,1)$ y que permita ordenar por los valores de una columna: Suponga a niveles o tratamientos del factor considerado. Sea n_i el número de replicaciones a ser tomadas en el i -ésimo tratamiento, y $N = \sum_{i=1}^a n_i$ el total de observaciones (el número total de unidades experimentales). Codifique los tratamientos de 1 hasta a y etiquete las unidades experimentales de 1 hasta N .

Paso 1: En una columna ingrese n_1 veces el número 1, n_2 veces el número 2, ..., n_a veces el número a , teniendo así un total de N entradas en dicha columna; estos valores representan a las etiquetas de los tratamientos.

Paso 2: Ingrese en otra columna N números aleatorios de una $U(0,1)$ con suficientes dígitos para evitar empates (números iguales).

Paso 3: Reordene ambas columnas usando como criterio el orden ascendente de los números aleatorios, así quedan ordenadas en forma aleatoria las etiquetas de los tratamientos en la primera columna.

Paso 4: Asigne la unidad experimental l al tratamiento cuya etiqueta quedó en la fila l -ésima, $l = 1, 2, \dots, N$.

Tenga en cuenta que si el número total de unidades experimentales N es un entero con k dígitos, entonces los números aleatorios en la columna 2 deberían ser una lista de números aleatorios de k dígitos.

En R podemos realizar de manera sencilla lo anterior. Suponga que se va a ejecutar un DOE bajo un DCA con un factor de efectos fijos en $a = 3$ niveles: 35, 40 y 45, cada uno replicado $n_i = 4$ veces, $i = 1, 2, 3$, por lo que se usarán $N = 12$ unidades experimentales y por tanto serán 12 corridas. Un programa R para construir el plan de aleatorización es como sigue.

Código R 3.2.1. *Plan de aleatorización en un DCA balanceado unifactorial.*

```
ni=4 #tamaño muestras balanceadas
a=3 #niveles del factor
tratamientos=factor(rep(c(35,40,45),each=ni)) #etiquetas de los tratamientos replicados ni=4
tratamientos
N=a*ni #total de unidades experimentales
set.seed(1234)
fac=sample(tratamientos,N) #m.a.s sin reemplazo sobre los tratamientos replicados
UE=1:N #etiquetas de las unidades experimentales
plan=data.frame(Unidad=UE,Tratamiento=fac)
plan
```

Esta programación produce los siguientes resultados

Salida R 3.2.1.

```
> tratamientos
[1] 35 35 35 35 40 40 40 40 45 45 45 45
Levels: 35 40 45
```

```
> plan
      Unidad Tratamiento
1         1           35
2         2           40
3         3           45
4         4           40
5         5           45
6         6           40
7         7           35
8         8           45
9         9           35
10        10           40
11        11           35
12        12           45
```

Siguiendo los pasos 1 a de aleatorización descritos, el programa también puede diseñarse de la siguiente forma:

Código R 3.2.2. *Plan de aleatorización en un DCA balanceado unifactorial.*

```
ni=4 #tamaño muestras balanceadas
a=3 #niveles del factor
tratamientos=factor(rep(c(35,40,45),each=ni)) #etiquetas de los tratamientos replicados ni=4
tratamientos
N=a*ni #total de unidades experimentales
set.seed(1234)
numaleat=runif(N) #Extrayendo N números aleatorios U(0,1)
numaleat
plani=data.frame(tratamientos,numaleat)
planf=data.frame(plani[order(numaleat),],UE=1:N,row.names =1:N)
planf
```

En anterior programa produce los siguientes resultados:

Salida R 3.2.2.

```
> tratamientos
[1] 35 35 35 35 40 40 40 40 45 45 45 45
Levels: 35 40 45
```

```
> numaleat
[1] 0.113703411 0.622299405 0.609274733 0.623379442 0.860915384 0.640310605
[7] 0.009495756 0.232550506 0.666083758 0.514251141 0.693591292 0.544974836
> planf
      tratamientos      numaleat UE
1           40 0.009495756 1
2           35 0.113703411 2
3           40 0.232550506 3
4           45 0.514251141 4
5           45 0.544974836 5
6           35 0.609274733 6
7           35 0.622299405 7
8           35 0.623379442 8
9           40 0.640310605 9
10          45 0.666083758 10
11          45 0.693591292 11
12          40 0.860915384 12
```

Si el diseño fuera desbalanceado, es decir, si el número de réplicas o tamaño de muestra en cada nivel no fueran iguales, podría ejecutarse la aleatorización en R como se ejemplifica a continuación. Suponga de nuevo un factor de tres niveles: 35, 40 y 45 pero con $n_1 = 5$, $n_2 = 4$ y $n_3 = 3$, respectivamente

Código R 3.2.3. *Plan de aleatorización en un DCA desbalanceado unifactorial. Note el uso del argumento `times` en la función R `rep()` en lugar del argumento `each` usado en los dos códigos R previos.*

```
ni=c(5,4,3) #tamaño muestras balanceadas
tratamientos=factor(rep(c(35,40,45),times=ni)) #etiquetas tratamientos replicados ni, i=1, 2, 3
tratamientos
N=sum(ni) #total de unidades experimentales
set.seed(1234)
fac=sample(tratamientos,N) #m.a.s sin reemplazo sobre los tratamientos replicados
UE=1:N #etiquetas de las unidades experimentales
plan=data.frame(Unidad=UE,Tratamiento=fac)
plan
```

El resultado de este código es como sigue:

Salida R 3.2.3.

```
> tratamientos
[1] 35 35 35 35 35 40 40 40 40 45 45 45
Levels: 35 40 45

> plan
  Unidad Tratamiento
1      1           35
2      2           40
3      3           45
4      4           40
5      5           45
6      6           35
7      7           35
8      8           45
9      9           35
10     10          40
11     11          35
12     12          40
```

Si la sucesión de corridas es aleatoria, cada sucesión posible tiene la misma oportunidad de ocurrir, luego las corridas y las asignaciones se dicen son completamente aleatorias. Sin embargo, aún con mecanismos de aleatorización pueden aparecer tratamientos en secuencia, y por tanto el experimentador podría no desear ciertos órdenes en la experimentación, pero al mirar la secuencia resultante no se debería decidir cambiarla subjetivamente porque no parece aleatoria.

3.2.2. Modelo para un diseño completamente aleatorizado, un factor efectos fijos

Existen dos posibles modelos en este caso. El primero es el modelo de medias de tratamientos,

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad (3.1)$$

y el segundo es el modelo de efectos de tratamientos,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \text{ sujeto a } \sum_{i=1}^a n_i \alpha_i = 0. \quad (3.2)$$

donde

- Y_{ij} denota la variable aleatoria que corresponde a la respuesta de la j -ésima unidad experimental asignada al i -ésimo tratamiento, $i = 1, 2, \dots, a$ y $j = 1, 2, \dots, n_i$,
- μ_i es la respuesta esperada al i -ésimo tratamiento si ésta es observada bajo condiciones experimentales idénticas y medida sin error,
- ε_{ij} es el error aleatorio.

Observe que en (3.2) μ_i es reemplazado por $\mu + \alpha_i$, donde,

- μ denota la respuesta global promedio y
- α_i es la desviación o el efecto del i -ésimo tratamiento sobre la respuesta media global.

Por tanto, el análisis de las diferencias entre las medias de tratamiento μ_i es equivalente a examinar las diferencias entre los efectos α_i , bajo la restricción lineal $\sum_{i=1}^a n_i \alpha_i = 0$, para una solución única del sistema de ecuaciones de mínimos cuadrados. En cualquiera de los dos modelos se establecen los siguientes supuestos: Los errores son variables aleatorias que se distribuyen normales, independientes, de media cero y varianza constante σ^2 , es decir, $\varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2)$, $\forall i = 1, 2, \dots, a$, $j = 1, 2, \dots, n_i$. Los parámetros a estimar son las constantes μ , μ_i , α_i y σ^2 .

Note que si $n_i = n$, $\forall i = 1, 2, \dots, a$, entonces el modelo (3.2) queda de la siguiente manera,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \text{ sujeto a } \sum_{i=1}^a \alpha_i = 0. \quad (3.3)$$

En la Tabla 3.2 se ilustra la tabulación de los datos experimentales.

Para ejemplificar la aproximación por regresión de los dos modelos DCA suponga un diseño balanceado para un factor de tres niveles, con $n_i = n = 3$, $\forall i = 1, 2, 3$. Para el modelo en (3.1) definimos las variables indicadoras X_k , $k = 1, 2, 3$ así:

$$X_k = \begin{cases} 1 & \text{si el caso proviene del nivel } k \text{ del factor} \\ 0 & \text{c.c} \end{cases}$$

Tabla 3.2: Tabla de datos experimentales de un DOE en un DCA

	Niveles factor A			
	A_1	A_2	\dots	A_a
Datos respuesta	Y_{11}	Y_{21}	\dots	Y_{a1}
	Y_{12}	Y_{22}	\dots	Y_{a2}
	\vdots	\vdots	\vdots	\vdots
	Y_{1n_1}	Y_{2n_2}	\dots	Y_{an_a}
Total	$Y_{1\bullet}$	$Y_{2\bullet}$	\dots	$Y_{a\bullet}$

Entonces el modelo de regresión equivalente al modelo ANOVA (3.1) es

$$Y_{ij} = \mu_1 X_{1,ij} + \mu_2 X_{2,ij} + \mu_3 X_{3,ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2), \quad \forall i = 1, 2, 3 \quad j = 1, 2, 3,$$

el cual corresponde a un MRLM sin intercepto y donde los coeficientes de regresión son las medias μ_i . En forma matricial el sistema de ecuaciones resultante al evaluar el modelo en el conjunto de datos experimentales es de la forma $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde,

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \end{bmatrix}.$$

Para definir el modelo de regresión equivalente al modelo ANOVA en (3.3), de nuevo con $a = 3$ niveles y $n_i = 3$, tenemos que considerar la restricción $\sum_{i=1}^3 \alpha_i = 0$, por tanto podemos, escribir $\alpha_3 = -\alpha_1 - \alpha_2$ y definimos $a - 1 = 2$ variables explicatorias X_k , para $k = 1, 2$, así:

$$X_k = \begin{cases} 1 & \text{si el caso es del nivel } k \text{ del factor} \\ -1 & \text{si el caso es del nivel } 3 \text{ del factor,} \\ 0 & \text{c.c} \end{cases}$$

y la ecuación del modelo de regresión es,

$$Y_{ij} = \mu + \alpha_1 X_{1,ij} + \alpha_2 X_{2,ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{IID}{\sim} N(0, \sigma^2), \quad \forall i = 1, 2, 3 \quad j = 1, 2, 3,$$

que como puede verse, es un MRLM con intercepto. El sistema de ecuaciones en su forma matricial $Y = X\beta + \varepsilon$ tiene ahora como matriz de diseño X y vector de parámetros β a

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

3.2.3. Test ANOVA para un DCA

Considere la estructura de datos representada en la Tabla 3.2, donde

- $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$ es la suma de las observaciones en el nivel o tratamiento i ,
- $Y_{\bullet\bullet} = \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}$ es la Suma o total de todas las N observaciones en el experimento.

Con estas cantidades se obtienen los siguientes estimadores de mínimos cuadrados en el modelo de efectos de tratamientos (ver Anexo A.1: Funciones estimables de los parámetros):

$$\hat{\mu} = \bar{Y}_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^a Y_{i\bullet} = \frac{Y_{\bullet\bullet}}{N}, \quad \hat{\mu}_i = \bar{Y}_{i\bullet} = \frac{Y_{i\bullet}}{n_i}, \quad \hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet} \quad (3.4)$$

Es de anotar que la suma de cuadrados a ser minimizada, es

$$SSE = \sum_{i=1}^a \sum_{j=1}^{n_i} \varepsilon_{ij}^2 = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \mu - \alpha_i)^2 \quad \text{sujeto a} \quad \sum_{i=1}^a n_i \alpha_i = 0. \quad (3.5)$$

Es fácil ver que las respuestas ajustadas para el tratamiento i son $\hat{Y}_{ij} = \bar{Y}_{i\bullet}$. Por tanto los residuales del modelo corresponden a $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_{i\bullet}$.

Las sumas de cuadrados del modelo ANOVA son

- Suma de cuadrados totales o SST , que es la variabilidad total observada en la respuesta, con $N - 1$ grados de libertad,

$$SST = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}^2 - N\bar{Y}_{\bullet\bullet}^2, \quad (3.6)$$

y que se descompone como $SST = SSA + SSE$;

- Suma de cuadrados debido al factor o SSA , que es la variabilidad en la respuesta explicada por el factor A, con $a - 1$ grados de libertad,

$$SSA = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^a n_i \bar{Y}_{i\bullet}^2 - N\bar{Y}_{\bullet\bullet}^2 \quad (3.7)$$

- Suma de cuadrados del error o SSE , que es la variabilidad en la respuesta explicada por la aleatoriedad, con $N - a$ grados de libertad,

$$SSE = \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2 = SST - SSA \quad (3.8)$$

Por tanto obtenemos la siguiente tabla ANOVA:

Tabla 3.3: Tabla ANOVA en un DCA

Fuente de variación	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F_0	Valor P
Factor	$a - 1$	SSA	$MSA = \frac{SSA}{a-1}$	$\sigma^2 + \frac{\sum_{i=1}^a n_i \alpha_i^2}{a-1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, N-a} > F_0)$
Error	$N - a$	SSE	$MSE = \frac{SSE}{N-a}$	σ^2		
Total	$N - 1$	SST				

Observe de la Tabla anterior que el MSE es un estimador insesgado de la varianza σ^2 .

Bajo los supuestos sobre los errores, la prueba de hipótesis fundamental asociada al ANOVA, es la siguiente:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a \text{ vs. } H_1 : \text{algún par } \mu_i \neq \mu_j \quad (3.9)$$

o de forma equivalente,

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \text{ vs. } H_1 : \text{algún } \alpha_i \neq 0. \quad (3.10)$$

Si F_0 es grande, se rechaza la hipótesis nula, en tanto que si es pequeño, no se tiene suficiente evidencia en contra de H_0 .

3.2.4. Distribuciones muestrales de los estimadores

Puede demostrarse que bajo los supuestos sobre los errores del modelo ANOVA,

$$\bar{Y}_{..} \sim N\left(\mu, \frac{\sigma^2}{N}\right), \quad \bar{Y}_{i\bullet} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right), \quad \hat{\alpha}_i \sim N\left(\alpha_i, \sigma^2 \left[\frac{1}{n_i} - \frac{1}{N}\right]\right) \quad (3.11)$$

Por tanto, los correspondientes errores estándar e intervalos de confianza están dados por:

Tabla 3.4: Estimadores, errores estándar e I.C para los parámetros del modelo

Estimador	Error estándar	Intervalo de confianza de $(1 - \gamma) \%$
$\bar{Y}_{..}$	$S_{\bar{Y}_{..}} = \sqrt{\frac{MSE}{N}}$	$\mu \in \bar{Y}_{..} \pm t_{\gamma/2, N-a} \times S_{\bar{Y}_{..}}$
$\bar{Y}_{i\bullet}$	$S_{\bar{Y}_{i\bullet}} = \sqrt{\frac{MSE}{n_i}}$	$\mu_i \in \bar{Y}_{i\bullet} \pm t_{\gamma/2, N-a} \times S_{\bar{Y}_{i\bullet}}$
$\hat{\alpha}_i$	$S_{\hat{\alpha}_i} = \sqrt{MSE \left(\frac{1}{n_i} - \frac{1}{N}\right)}$	$\alpha_i \in \hat{\alpha}_i \pm t_{\gamma/2, N-a} \times S_{\hat{\alpha}_i}$

3.2.5. Validación de supuestos

Para cada supuesto sobre los errores del modelo existen pruebas analíticas y gráficas, y de acuerdo a [4] éstas últimas aunque no son exactas, funcionan relativamente bien con pocos datos, y se requiere que una fuerte evidencia visual en contra de un supuesto esté soportada por más de dos puntos para concluir que no se cumple. Cuando sólo un par de puntos se alejan del comportamiento esperado en una gráfica de diagnóstico, esto puede tratarse como un problema de datos atípicos. Un resumen de los tests a realizar se muestra en la Tabla 3.5.

Tabla 3.5: Supuestos a evaluar y pruebas correspondientes

Supuesto	Hipótesis	Test estadístico	Gráfico
Normalidad	$H_0 : \varepsilon_{ij} \sim \text{Normal vs.}$ $H_1 : \varepsilon_{ij} \text{ no son Normales}$	Shapiro Wilk	Gráfico de probabilidad normal sobre residuales, o residuales estandarizados o estudentizados
Varianza Constante	$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2 \text{ vs.}$ $H_1 : \text{algún par } \sigma_i^2 \neq \sigma_j^2$	Bartlett, Cochran, Levene	Gráfico de residuales vs. valores ajustados, residuales vs. niveles del factor
Independencia	$H_0 : \text{corr}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0 \forall (i, j) \neq (i', j')$ $H_1 : \text{corr}(\varepsilon_{ij}, \varepsilon_{i'j'}) \neq 0 \text{ para algún } (i, j) \neq (i', j')$	Ljung-Box, ACF, Durbin - Watson de orden 1	Gráfico de residuales (comunes) vs. orden de corrida
Nota: Para test de independencia se necesita conocer el orden o secuencia temporal entre las observaciones y evaluar con los residuos organizados en ese orden de tiempo. Ver en Apéndices A.2, A.3 y A.4, los tests de Bartlett, Levene y Cochram, respectivamente. Para los tests Ljung-Box y Durbin Watson remitirse a textos de series de tiempo.			

Adicionalmente, se chequean, [2]:

- La forma del modelo (ajuste del modelo), es decir, si realmente las respuestas medias de los tratamientos son adecuadamente descritos por el modelo lineal postulado, $\mu_i = \mu + \alpha_i$, mediante el gráfico de residuales estandarizados o estudentizados vs. niveles del factor. Patrones no aleatorios alrededor de la línea cero de este gráfico son indicios de carencia de ajuste.
- Outliers, mediante gráfico de residuales estandarizados o estudentizados vs. niveles del factor; si todas las presunciones del modelo se cumplen incluyendo la normalidad, aproximadamente el 68 % de los residuales estandarizados o estudentizados deberían estar entre -1 y $+1$, aproximadamente el 95 % entre -2 y $+2$, y aproximadamente 99.7 % entre -3 y $+3$. Si hay muchos outliers, los niveles de confianza son más bajos de los esperados.

De acuerdo a [2],

- El supuesto de independencia debe verificarse antes de los otros supuestos. La no independencia en los errores puede ser causada por la similitud de las unidades experimentales cercanamente juntas en el tiempo o espacio. La no independencia causa que los verdaderos niveles de significancia de las pruebas de hipótesis sean mucho más altas que las establecidas y que los niveles de confianza y las potencias de las pruebas sean más bajas que lo esperado.
- Las pruebas formales sobre igualdad de varianzas tienden a ser poco **potentes** con pocas réplicas por nivel y muy sensibles a no normalidad. Una regla de dedo que usualmente se aplica es que el estadístico F de la ANOVA y los métodos de comparaciones múltiples de medias son apropiadas, si se ha probado que la razón de la estimación de la varianza más grande a la varianza más pequeña de los tratamientos no es mayor de tres (3); sin embargo hay que ser cuidadosos con esta regla porque aún cuando las presunciones del modelo son correctas tal razón de varianzas puede dar mayor que 3.
- La violación del supuesto de varianza constante es preocupante en particular cuando los diseños son desbalanceados (número de réplicas no son iguales). Si el nivel con mayor varianza es el menos replicado, entonces los procedimientos estadísticos son más liberales y se tiene mayor probabilidad de cometer error tipo I, y menores niveles de confianza que los esperados. Por el contrario, si el tratamiento o nivel de mayor varianza es el más replicado, entonces los procedimientos estadísticos son más conservadores, es decir, niveles de significancia son más bajos y niveles de confianza más altos de los esperados. En general se recomienda trabajar con diseños balanceados.
- Pequeñas desviaciones de la normalidad no afectan fuertemente a los niveles de significancia, niveles de confianza o la potencia. Con diseños balanceados el caso de mayor preocupación es cuando la distribución tiene colas más pesadas

que la normal, y en ese caso deberían usarse métodos de análisis de varianza no paramétricos (ej. Kruskall - Wallis).

Soluciones al problema de varianza no constante

- Transformaciones de la variable respuesta para estabilizar varianza cuando la varianza desigual está asociada a las medias de tratamientos (ver Apéndice A.5).
- Análisis con varianzas desiguales (ver Apéndice A.6).
- Análisis por mínimos cuadrados ponderados. Cuando la varianza no constante no está relacionada con las medias de tratamientos no se recomienda aplicar transformaciones. en este caso, una solución posible es usar mínimos cuadrados ponderados, en le cual el vector de parámetros estimados $\hat{\beta}$ es la solución de las ecuaciones normales $X^t W X \beta = X^t W Y$, es decir, es tal que minimiza la suma de cuadrados ponderadas del error, $S(\beta) = (Y - X\beta)^t W (Y - X\beta)$, donde W es una matriz diagonal con los recíprocos de las varianzas muestrales en cada nivel del factor. En R la función `lm()` con la cual se ajustan los modelos de regresión lineal, es posible llevar a cabo mínimos cuadrados ponderados mediante el uso de su argumento `weights=...` Para más detalles teóricos sobre el ajuste de modelos de regresión ponderados ver Sección 10.1 de [7].

Soluciones al problema de no normalidad

- Métodos no paramétricos
- Transformar la variable respuesta, tal como la transformación logarítmica. Aunque a veces transformar puede originar problemas de varianza no constante que no existía con los datos originales. Otra posibilidad que se escapa de los objetivos de este curso, es trabajar con modelos lineales generalizados.
- Si la varianza no es constante y si hay suficientes datos por nivel del factor, analizar la normalidad por separado en cada nivel.

Tabla 3.6: Algunas transformaciones de la respuestas para normalidad, con base en la distribución original

Distribución de y	Varianza como función de $E[y] = \mu$	Transformación $f(y)$
Binomial	$\frac{\mu(1-\mu)}{n}$	$\sin^{-1} \sqrt{y/n}$ (radianes)
poisson	μ	\sqrt{y} ó $\sqrt{y+0.5}$
Lognormal	$c\mu^2$	$\log(y)$

Soluciones al problema de no independencia

De acuerdo a [2], la no independencia es difícil de corregir; si hay una clara tendencia en el gráfico de residuales tal como una línea recta, es posible incorporar al modelo términos que representen el efecto temporal o espacial. Por ejemplo, con una tendencia lineal en el tiempo, el modelo adecuado sería $Y_{ij} = \mu + \alpha_i + \lambda t_{ij} + \varepsilon_{ij}$, $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, donde t_{ij} es el tiempo en el cual la observación j del tratamiento i fue tomada. Este modelo es uno denominado de *análisis de covarianza* o ANCOVA.

Veamos los siguientes dos ejemplos.

3.2.6. Ejemplo: Mejora de la productividad (tomado de [7])

Un economista compiló datos sobre mejoramiento de la productividad en los últimos tres años para una muestra de firmas productoras de equipos de computación. Las firmas fueron clasificadas de acuerdo a nivel de sus gastos en investigación y desarrollo en los pasados tres años, en bajo, moderado y alto. Los resultados del estudio sobre la mejora de la productividad (mprod, medida en una escala de 0 a 100) se presentan en la tabla a continuación.

ID	Observaciones											
	1	2	3	4	5	6	7	8	9	10	11	12
alto	8.5	9.7	10.1	7.8	9.6	9.5						
bajo	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
medio	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4

Asumiendo que un modelo de efectos fijos es apropiado:

a) Plantee un modelo estadístico apropiado para el ANOVA. Identifique sus térmi-

nos y los supuestos necesarios.

- b) Haga un análisis descriptivo de los datos. ¿Qué puede concluir de estos en cuanto al promedio de mejoramiento de la productividad? ¿La variabilidad dentro de cada nivel del gasto en investigación y desarrollo puede considerarse o no aproximadamente constante?
- c) Calcule la tabla ANOVA, la tabla de parámetros estimados para los datos y la tabla de medias. Analice los resultados y pruebe a un nivel de significancia de 0.05, si la media de mejoramiento de la productividad difiere según el nivel de gastos en investigación y desarrollo. Tenga en cuenta en sus cálculos que este problema es un caso de diseño desbalanceado.
- d) Dé una expresión general para el cálculo de los residuales ordinarios. Analice los gráficos de residuales internamente estudentizados $z_{ij} = \frac{\hat{\epsilon}_{ij}}{\sqrt{\left(\frac{n_i-1}{n_i}\right)MSE}}$ y úselos para validar los supuestos del modelo. Realice el teste de normalidad con base en estos residuales estudentizados.
- e) Calcule intervalos de confianza del 95 % para la mejora media de productividad para cada nivel de gasto en inversión y desarrollo.
- f) Estime los efectos de cada nivel del factor y sus respectivos errores estándar, y pruebe la significancia de cada uno usando para cada prueba un nivel de significancia de 0.05.

Formule los tests de hipótesis y su procedimiento de prueba, donde sea necesario. Un programa R para obtener los resultados pedidos en b) a f) es el siguiente (vea Apéndices A.2, A.3 y A.4 aplicación de tests de homogeneidad de varianza sobre estos datos):

Código R 3.2.4. Programa R para solución ejemplo 3.2.6

```

library(gmodels)
library(multcomp)

#CREANDO FUNCIÓN USUARIO mismediastratamientos()
mismediastratamientos=function(modeloanova,nivel=95){
MSE=anova(modeloanova)["Mean Sq"][2,]
df=anova(modeloanova)["Df"][2,]
ni=unlist(model.tables(modeloanova,type = "means")["n"])
alfa=1-nivel/100
alfa.med=(1-(nivel/100))/2
t=qt(alfa.med,df=df,lower.tail=F)
medias.tratam=unlist(model.tables(modeloanova,type = "means")["tables"])[-1]
interval=cbind(ni=ni,Medias=medias.tratam,LIC=medias.tratam-t*sqrt(MSE/ni),
               LSC=medias.tratam+t*sqrt(MSE/ni))
cat("Tabla de medias de tratamientos y sus I.C de",nivel,"%","\n")
cat("alfa","          ",alfa,"\n")
cat("grados de libertad    ",df,"\n")
cat("error cuadrático medio",MSE,"\n")
cat("valor crítico t        ",t,"\n","\n")
interval
}

ni=c(6,9,12) #tamaños de muestra en orden niveles alto, bajo y medio
#ENTRADA DE DATOS: En el orden de niveles alto, bajo y medio
diseño=data.frame(ID=factor(rep(c("alto","bajo","medio"),times=ni)),mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)

#CALCULANDO MEDIAS DE TRATAMIENTO PARA LUEGO USAR EN GRÁFICA DE BOXPLOTS
mediasy=sapply(split(mprod,ID),mean)
#BOXPLOTS COMPARATIVOS
boxplot(mprod~ID,boxwex = 0.5)
lines(1:3,mediasy,col=2,lty=2,type="b",pch=19)

#AJUSTE MODELO ANOVA UN FACTOR
modelo=aov(mprod~ID)

```

```

#OBTENIENDO TABLA ANOVA
anova(modelo)

#o bien
summary(modelo)

#OBTENIENDO MEDIAS ESTIMADAS POR NIVEL DEL FACTOR
model.tables(modelo,type = "means",se=TRUE)

#OBTENCIÓN DE MEDIAS DE TRATAMIENTO CON SUS I.C DEL 95%
mismediastratamientos(modelo)

#ESTIMACIÓN DE LOS EFECTOS PRINCIPALES:
model.tables(modelo,type = "effects",se=TRUE)

#CÁLCULO DE LOS EFECTOS DE TRATAMIENTOS, SUS TESTES T Y I.C DEL 95%:
#La función fit.contrast requiere la librería gmodels
efect.alto=fit.contrast(modelo,"ID",rbind(":efecto alto"=c(2/3,-1/3,-1/3)),conf=0.95)
efect.bajo=fit.contrast(modelo,"ID",rbind(":efecto bajo"=c(-1/3,2/3,-1/3)),conf=0.95)
efect.medio=fit.contrast(modelo,"ID",rbind(":efecto medio"=c(-1/3,-1/3,2/3)),conf=0.95)
rbind(efect.alto,efect.bajo,efect.medio)

#OTRA FORMA PARA CÁLCULO INDIVIDUAL DE EFECTOS CON SUS INTERVALOS DE CONFIANZA DEL 95%:
#La función glht requiere la librería multcomp
contr.alto= rbind("efecto id alto"= c(2/3,-1/3,-1/3))
contr.bajo= rbind("efecto id bajo"= c(-1/3,2/3,-1/3))
contr.medio= rbind("efecto id medio"= c(-1/3,-1/3,2/3))

rbind(confint(glht(modelo,linfct=mcp(ID=contr.alto)))$confint,
confint(glht(modelo,linfct=mcp(ID=contr.bajo)))$confint,
confint(glht(modelo,linfct=mcp(ID=contr.medio)))$confint)

#GRÁFICOS PARA ANÁLISIS DE RESIDUALES INTERNAMENTE ESTUDENTIZADOS
layout(rbind(c(1,1,2,2),c(0,3,3,0)))
stripchart(rstandard(modelo)~ID,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1) #residuos internamente
                                                                    #estudentizados vs.
                                                                    #niveles factor

abline(h=c(-2,0,2),lty=2)
plot(fitted(modelo),rstandard(modelo),ylim=c(-2.5,2.5)) #residuales internamente estudentizados vs.
                                                                    #valores ajustados

abline(h=c(-2,0,2),lty=2)
qqnorm(rstandard(modelo)) #gráfico de prob. normal de residuos internamente estudentizados
qqline(rstandard(modelo),lty=2)

```

```
detach(datos)
```

En las siguientes salidas R se observan los resultados numéricos obtenidos. En las Figuras 3.1 y 3.2 se muestran las gráficas generadas.

Salida R 3.2.4. Tabla ANOVA

```
> anova(modelo)
Analysis of Variance Table

Response: mprod
          Df Sum Sq Mean Sq F value    Pr(>F)
ID          2  20.125  10.0626    15.72 4.331e-05 ***
Residuals 24  15.362   0.6401
```

Salida R 3.2.5. Medias y efectos de tratamientos estimados con función `R model.tables()`.

Observe mensaje al comienzo de cada resultado anunciando que no se calculó errores estándar debido a que el diseño es desbalanceado.

```
> model.tables(modelo, type = "means", se=TRUE)
Design is unbalanced - use se.contrast() for se's
Tables of means
Grand mean
7.951852
ID
      alto bajo medio
      9.2 6.878 8.133
rep 6.0 9.000 12.000

> model.tables(modelo, type = "effects", se=TRUE)
Design is unbalanced - use se.contrast() for se's
Tables of effects
ID
      alto bajo medio
      1.248 -1.074 0.1815
rep 6.000 9.000 12.0000
```

Salida R 3.2.6. Medias de tratamientos con sus errores estándar e I.C del 95 %, obtenidos con función usuario `mismediastratamientos()`

```
> mismediastratamientos(modelo)
Tabla de medias de tratamientos y sus I.C de 95 %
alfa                0.05
grados de libertad    24
error cuadrático medio 0.6400926
valor crítico t        2.063899
```

	ni	Medias	LIC	LSC
n.ID.alto	6	9.200000	8.525885	9.874115
n.ID.bajo	9	6.877778	6.327365	7.428191
n.ID.medio	12	8.133333	7.656662	8.610005

Salida R 3.2.7. Estimación de los efectos de tratamientos con sus errores estándar y sus I.C del 95 % mediante la función `fit.contrast()` de la librería `gmodels`. Revise en el código R 3.2.4 la creación de los objetos `efect.alto`, `efect.bajo`, `efect.medio` mediante la función `fit.contrast()`.

```
> rbind(efect.alto,efect.bajo,efect.medio)
```

	Estimate	Std. Error	t value	Pr(> t)	lower CI	upper CI
ID:efecto alto	1.12962963	0.2474741	4.5646379	1.256483e-04	0.6188682	1.6403911
ID:efecto bajo	-1.19259259	0.2222383	-5.3662785	1.648749e-05	-1.6512699	-0.7339153
ID:efecto medio	0.06296296	0.2084780	0.3020125	7.652442e-01	-0.3673145	0.4932404

Salida R 3.2.8. Estimación de los efectos de tratamientos con sus errores estándar y sus I.C del 95 % mediante la combinación de las funciones R `confint()` y `glht()`, ésta última de la librería `multcomp`. Revise en el código R 3.2.4 la creación de los objetos `contr.alto`, `contr.bajo`, `contr.medio`, antes del uso de `confint(glht(...))`.

```
> rbind(confint(glht(modelo,linfct=mcp(ID=contr.alto)))$confint,
+ confint(glht(modelo,linfct=mcp(ID=contr.bajo)))$confint,
+ confint(glht(modelo,linfct=mcp(ID=contr.medio)))$confint)
```

	Estimate	lwr	upr
efecto id alto	1.12962963	0.6188682	1.6403911
efecto id bajo	-1.19259259	-1.6512699	-0.7339153
efecto id medio	0.06296296	-0.3673145	0.4932404

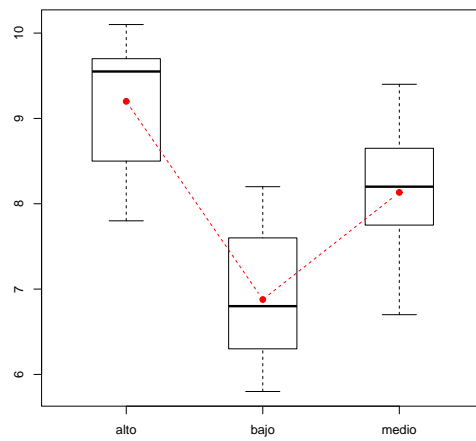


Figura 3.1: Boxplots sobre datos en ejemplo 3.2.6.

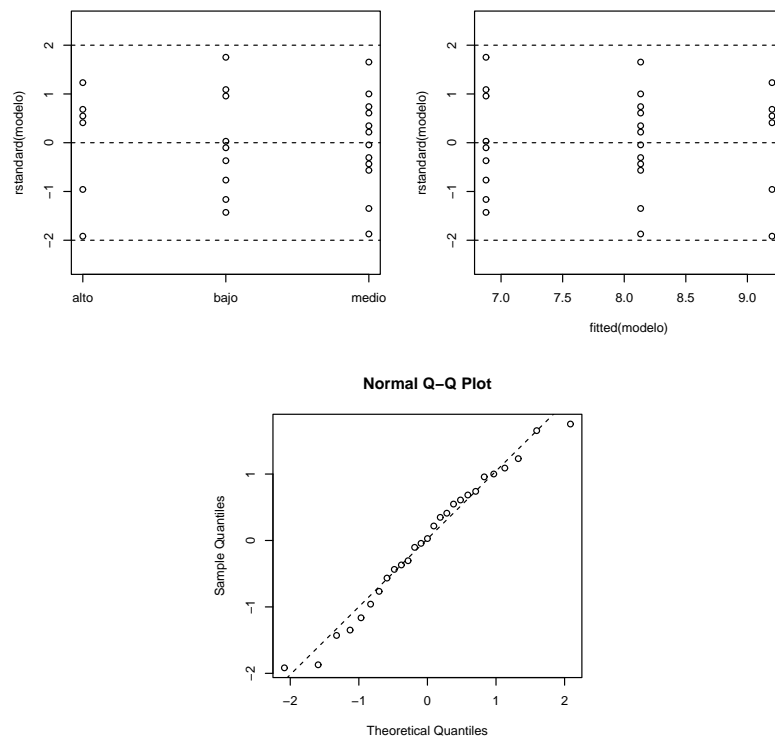


Figura 3.2: Residuos internamente estudentizados y su gráfico de probabilidad normal, para modelo ANOVA sobre datos en ejemplo 3.2.6.

3.2.7. Ejemplo: Efectos de un fertilizante (tomado de [6])

Se llevó a cabo un experimento para probar los efectos de un fertilizante nitrogenado en la producción de lechuga. Se aplicaron cinco dosis diferentes de nitrato de amonio a cuatro parcelas (las réplicas) en un diseño totalmente aleatorizado. Los datos son el número de lechugas cosechadas de las parcelas.

Tratamiento (lb N/acre)	lechuga/parcela			
0	104	114	90	140
50	134	130	144	174
100	146	142	152	156
150	147	160	160	163
200	131	148	154	163

- Especifique la estructura de tratamiento y de diseño empleado en este experimento. Escriba el modelo ANOVA apropiado con sus supuestos y restricciones.
- Realice un análisis descriptivo de los datos experimentales establezca conclusiones preliminares sobre los posibles efectos en la producción promedio de los diferentes niveles de aplicación de nitrato.
- Realice el test de hipótesis asociado al modelo ANOVA, interprete a la luz del problema.
- Estime medias y efectos de tratamientos con sus intervalos de confianza del 95
- Valide los supuestos del modelo ANOVA de este experimento. Analice con residuos internamente estudentizados.

Nota 3.1. En los casos donde aplique, exprese apropiadamente la prueba de hipótesis, el estadístico de prueba, el criterio de rechazo. Un programa R para obtener los resultados necesario es como sigue,

Código R 3.2.5. Programa R para solución ejemplo 3.2.7

```

#PROGRAMA R EJEMPLO LECHUGAS

library(gmodels)
library(multcomp)
library(daewr)
library(car)
library(outliers)

#CREANDO FUNCIÓN USUARIO mismediastratamientos()
mismediastratamientos=function(modeloanova,nivel=95){
MSE=anova(modeloanova)["Mean Sq"][2,]
df=anova(modeloanova)["Df"][2,]
ni=unlist(model.tables(modeloanova,type = "means")["n"])
alfa=1-nivel/100
alfa.med=(1-(nivel/100))/2
t=qt(alfa.med,df=df,lower.tail=F)
medias.tratam=unlist(model.tables(modeloanova,type = "means")["tables"])[-1]
interval=cbind(ni=ni,Medias=medias.tratam,LIC=medias.tratam-t*sqrt(MSE/ni),LSC=medias.tratam+t*sqrt(MSE/ni))
cat("Tabla de medias de tratamientos y sus I.C de",nivel,"%","\n")
cat("alfa","          ",alfa,"\n")
cat("grados de libertad   ",df,"\n")
cat("error cuadrático medio",MSE,"\n")
cat("valor crítico t       ",t,"\n","\n")
interval
}

#ENTRADA DE DATOS
ni=rep(4,5) #No. réplicas por nivel
#obseve que los renglones de datos corresponden a cada nivel desde 0 a 200, respectivamente
diseño2=data.frame(nitrato=factor(rep(c(0,50,100,150,200),times=ni)),nlechug=scan())
104 114 90 140
134 130 144 174
146 142 152 156
147 160 160 163
131 148 154 163

attach(diseño2)

#CALCULANDO MEDIAS DE TRATAMIENTO PARA LUEGO USAR EN GRÁFICA
mediasy2=sapply(split(nlechug,nitrato),mean)

```

```

#BOXPLOTS COMPARATIVOS
boxplot(nlechug~nitrato,boxwex = 0.5)
lines(1:5,mediasy2,col=2,lty=2,type="b",pch=19)

#AJUSTE MODELO DE ANÁLISIS DE VARIANZA
modelo2=aov(nlechug~nitrato)
#OBTENIENDO TABLA ANOVA
anova(modelo2)

#OBTENIENDO MEDIAS ESTIMADAS POR NIVEL DEL FACTOR
model.tables(modelo2,type = "means",se=TRUE)
#OBTENCIÓN DE MEDIAS DE TRATAMIENTO CON SUS I.C DEL 95%
mismediastratamientos(modelo2)

#ESTIMACIÓN DE LOS EFECTOS PRINCIPALES
model.tables(modelo2,type = "effects",se=TRUE)
#CÁLCULO INDIVIDUAL DE EFECTOS, SUS TESTES T Y I.C DEL 95%
efect.0=fit.contrast(modelo2,"nitrato",
rbind(":efecto 0"=c(4/5,-1/5,-1/5,-1/5,-1/5)),conf=0.95)
efect.50=fit.contrast(modelo2,"nitrato",
rbind(":efecto 50"=c(-1/5,4/5,-1/5,-1/5,-1/5)),conf=0.95)
efect.100=fit.contrast(modelo2,"nitrato",
rbind(":efecto 100"=c(-1/5,-1/5,4/5,-1/5,-1/5)),conf=0.95)
efect.150=fit.contrast(modelo2,"nitrato",
rbind(":efecto 150"=c(-1/5,-1/5,-1/5,4/5,-1/5)),conf=0.95)
efect.200=fit.contrast(modelo2,"nitrato",
rbind(":efecto 200"=c(-1/5,-1/5,-1/5,-1/5,4/5)),conf=0.95)
rbind(efect.0,efect.50,efect.100,efect.150,efect.200)

#OTRA FORMA PARA CÁLCULO INDIVIDUAL DE EFECTOS CON SUS INTERVALOS DE CONFIANZA DEL 95%:
contr0=rbind("efecto nitrato 0"= c(4/5,-1/5,-1/5,-1/5,-1/5))
contr50=rbind("efecto nitrato 50"= c(-1/5,4/5,-1/5,-1/5,-1/5))
contr100=rbind("efecto nitrato 100"= c(-1/5,-1/5,4/5,-1/5,-1/5))
contr150=rbind("efecto nitrato 150"= c(-1/5,-1/5,-1/5,4/5,-1/5))
contr200=rbind("efecto nitrato 200"= c(-1/5,-1/5,-1/5,-1/5,4/5))

rbind(confint(glht(modelo2,linfct=mcp(nitrato=contr0)))$confint,
confint(glht(modelo2,linfct=mcp(nitrato=contr50)))$confint,
confint(glht(modelo2,linfct=mcp(nitrato=contr100)))$confint,
confint(glht(modelo2,linfct=mcp(nitrato=contr150)))$confint,
confint(glht(modelo2,linfct=mcp(nitrato=contr200)))$confint)

```



```
#OBTENIENDO GRÁFICOS DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS,
layout(rbind(c(1,1,2,2),c(0,3,3,0)))
stripchart(rstandard(modelo2)~nitrato,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,xlab="Lb N/acre")
abline(h=c(-2,0,2),lty=2)
plot(fitted(modelo2),rstandard(modelo2),ylim=c(-2.5,2.5))
abline(h=c(-2,0,2),lty=2)
qqnorm(rstandard(modelo2))
qqline(rstandard(modelo2),lty=2)

#TEST DE NORMALIDAD CON RESIDUOS INTERNAMENTE ESTUDENTIZADOS
shapiro.test(rstandard(modelo2))
#TESTES PARA HOMOGENEIDAD DE VARIANZA
bartlett.test(nlechug~nitrato)
leveneTest(nlechug~nitrato)
cochran.test(nlechug~nitrato,data=diseño2)
detach(diseño2)
```

El anterior código produce los siguientes resultados

Salida R 3.2.9. Tabla ANOVA

```
> anova(modelo2)
Analysis of Variance Table
Response: nlechug
          Df Sum Sq Mean Sq F value    Pr(>F)
nitrato    4 4994.8 1248.70   5.6113 0.005757 **
Residuals 15 3338.0   222.53
```

Salida R 3.2.10. Medias y efectos de tratamientos estimados con función `R model.tables()`.

```
> model.tables(modelo2,type = "means",se=TRUE)
Tables of means
Grand mean
142.6
      nitrato
nitrato
      0      50     100     150     200
112.0 145.5 149.0 157.5 149.0

Standard errors for differences of means
```

```

      nitrato
      10.55
replic.      4

> model.tables(modelo2,type = "effects",se=TRUE)
Tables of effects
  nitrato
nitrato
  0    50   100   150   200
-30.6  2.9   6.4  14.9   6.4

Standard errors of effects
      nitrato
      7.459
replic.      4

```

Salida R 3.2.11. Medias de tratamientos con sus errores estándar e I.C del 95 %, obtenidos con función usuario `mismediastratamientos()`

```

> mismediastratamientos(modelo2)
Tabla de medias de tratamientos y sus I.C de 95 %
alfa                0.05
grados de libertad    15
error cuadrático medio 222.5333
valor crítico t        2.13145

      ni Medias    LIC    LSC
tables.nitrato.0    4  112.0  96.102 127.898
tables.nitrato.50    4  145.5 129.602 161.398
tables.nitrato.100    4  149.0 133.102 164.898
tables.nitrato.150    4  157.5 141.602 173.398
tables.nitrato.200    4  149.0 133.102 164.898

```

Salida R 3.2.12. Estimación de los efectos de tratamientos con sus errores estándar y sus I.c del 95 % mediante la función `fit.contrast()` de la librería `gmodels`. Revise en el código R 3.2.5 la creación de los objetos `efect.0`, `efect.50`, `efect.100`, `efect.150`, `efect.200`, mediante la función `fit.contrast()`.

```
> rbind(efect.0,efect.50,efect.100,efect.150,efect.200)
```

	Estimate	Std. Error	t value	Pr(> t)	lower CI	upper CI
nitrato:efecto 0	-30.6	6.671332	-4.5867904	0.0003561942	-44.8196069	-16.38039
nitrato:efecto 50	2.9	6.671332	0.4346958	0.6699719245	-11.3196069	17.11961
nitrato:efecto 100	6.4	6.671332	0.9593287	0.3526022798	-7.8196069	20.61961
nitrato:efecto 150	14.9	6.671332	2.2334371	0.0411761195	0.6803931	29.11961
nitrato:efecto 200	6.4	6.671332	0.9593287	0.3526022798	-7.8196069	20.61961

Salida R 3.2.13. Estimación de los efectos de tratamientos con sus errores estándar y sus I.c del 95 % mediante la combinación de las funciones R `confint()` y `glht()`, ésta última de la librería `multcomp`. Revise en el código R 3.2.5 la creación de los objetos `contr0`, `contr50`, `contr100`, `contr150`, `contr200`, antes del uso de `confint(glht(...))`.

```
> rbind(confint(glht(modelo2,linfct=mcp(nitrato=contr0)))$confint,
+ confint(glht(modelo2,linfct=mcp(nitrato=contr50)))$confint,
+ confint(glht(modelo2,linfct=mcp(nitrato=contr100)))$confint,
+ confint(glht(modelo2,linfct=mcp(nitrato=contr150)))$confint,
+ confint(glht(modelo2,linfct=mcp(nitrato=contr200)))$confint)
```

	Estimate	lwr	upr
efecto nitrato 0	-30.6	-44.8196069	-16.38039
efecto nitrato 50	2.9	-11.3196069	17.11961
efecto nitrato 100	6.4	-7.8196069	20.61961
efecto nitrato 150	14.9	0.6803931	29.11961
efecto nitrato 200	6.4	-7.8196069	20.61961

Salida R 3.2.14. Test de normalidad Shapiro Wilk sobre residuos internamente estudentizados, tests de homogeneidad de varianzas de Bartlett, Levene y C-Cochran.

```
> shapiro.test(rstandard(modelo2))
Shapiro-Wilk normality test
data: rstandard(modelo2)
W = 0.94466, p-value = 0.2932
> bartlett.test(nlechug~nitrato)
Bartlett test of homogeneity of variances
data: nlechug by nitrato
Bartlett's K-squared = 5.7049, df = 4, p-value = 0.2223
```

```

> leveneTest(nlechug~nitrato)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.0143 0.4312
      15

> cochrans.test(nlechug~nitrato,data=diseño2)
Cochran test for outlying variance
data:  nlechug ~ nitrato
C = 0.40024, df = 4, k = 5, p-value = 0.4745
alternative hypothesis: Group 0 has outlying variance
sample estimates:
      0      50     100     150     200
445.33333 395.66667 38.66667 51.00000 182.00000

```

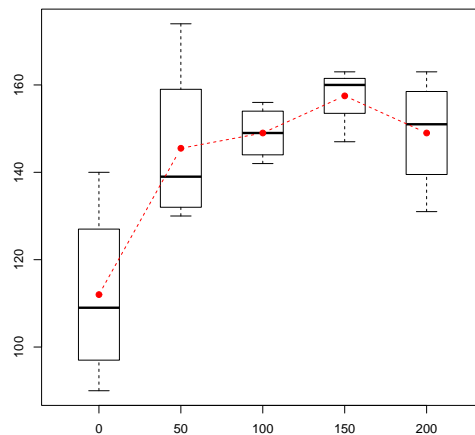


Figura 3.3: Boxplots sobre datos en ejemplo 3.2.7.

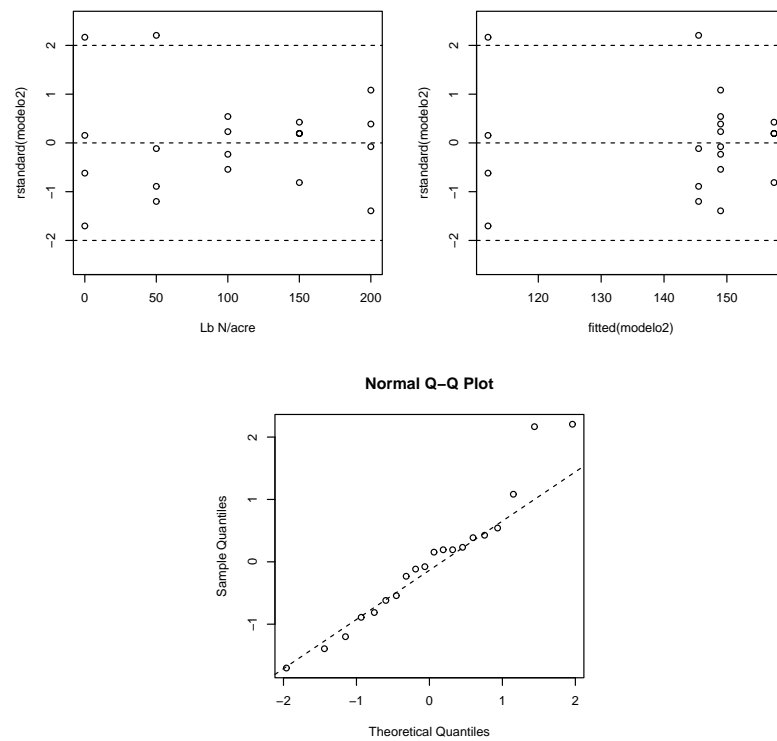


Figura 3.4: Residuos internamente estudentizados y su gráfico de probabilidad normal, para modelo ANOVA sobre datos en ejemplo 3.2.7.

Capítulo 4

Inferencias para medias de tratamientos con un factor

Luego de un ANOVA donde se rechazó la hipótesis nula asociada, es necesario determinar cuáles tratamientos son estadísticamente distintos. Para ello, se recurre a diferentes técnicas de comparación de medias de tratamiento:

- Comparaciones entre pares de medias
- Comparaciones múltiples o pruebas de rango múltiple.
- Comparaciones por contrastes
- Comparaciones de tratamientos con un control

4.1. Comparaciones de pares de medias de tratamientos - mínima diferencia significativa o LSD

Para un par dado de medias de tratamientos, las comparaciones pueden realizarse probando la igualdad de tales medias o construyendo el I.C para la diferencia. En particular, se podrían realizar todas las $a(a-1)/2$ comparaciones entre pares de medias. Un método sencillo es el método LSD (mínima diferencia significativa), en la cual se tiene (para $i \neq j$):

$$H_0 : \mu_i = \mu_j \text{ vs. } H_1 : \mu_i \neq \mu_j \quad (4.1)$$

El estadístico de prueba, bajo los supuestos del modelo lineal ANOVA un factor efectos fijos es: $D_{ij} = |\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}|$, el cual se compara con la LSD_{ij} (mínima diferencia

significativa) a un nivel de significancia γ , dada por,

$$LSD_{ij} = t_{\gamma/2, N-a} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (4.2)$$

Se rechaza H_0 si $D_{ij} > LSD_{ij}$. También podemos construir un I.C de $(1 - \gamma)100\%$ para la diferencia de medias de tratamientos:

$$(\mu_i - \mu_j) \in (\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) \pm LSD_{ij}, \quad (4.3)$$

de modo que se rechaza la hipótesis nula si $0 \notin (\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) \pm LSD_{ij}$

4.1.1. Ejemplo. [4]

Suponga que en un DCA se compararon los tiempos de cuatro métodos de ensamble A, B, C, y D, usando cuatro réplicas con cada método. El ANOVA arrojó los siguientes resultados, en donde visiblemente la hipótesis de igualdad de todos los tiempos promedios de ensamble de las cuatro métodos es rechazada al 5 % de significancia:

FUENTE	SC	g.l	CM	F_0	$P(f_{3,12} > F_0)$
Métodos	69.5	3	23.17	9.42	0.0018
Error	29.5	12	2.46		
Total	99.0	15			

Las medias muestrales por tratamiento, fueron respectivamente 7.25, 8.50, 12.75, y 10.50 y la media global de 9.75. todas las $n_i = 4$. A continuación se presentan las comparaciones por pares de medias, a un nivel de significancia de 0.05, donde $LSD_{ij} = 2.42$ en todas las comparaciones:

Diferencia Poblacional	Diferencia muestral absoluta	Decisión
$\mu_A - \mu_B$	$1.25 < 2.42$	Iguales
$\mu_A - \mu_C$	$5.50 > 2.42$	Diferentes
$\mu_A - \mu_D$	$3.25 > 2.42$	Diferentes
$\mu_B - \mu_C$	$4.25 > 2.42$	Diferentes
$\mu_B - \mu_D$	$2.00 < 2.42$	Iguales
$\mu_C - \mu_D$	$2.25 < 2.42$	Iguales

Con este procedimiento se concluye que las medias de tratamientos estadísticamente iguales pueden agruparse así:

- GRUPO 1: $\mu_A = \mu_B$
- GRUPO 2: $\mu_B = \mu_D$
- GRUPO 3: $\mu_C = \mu_D$

Como puede observarse hay traslape entre estos grupos.

Sin embargo el anterior método no es potente porque no controla la **tasa de error experimental** de todas las comparaciones. Dicha tasa es definida como la probabilidad de cometer error tipo I en al menos una de las $a(a-1)/2$ comparaciones entre pares.

Existen otros métodos de comparación múltiple más potentes para detectar diferencias no nulas entre pares de medias, pero que no controlan la tasa de error experimental, entre ellas la Prueba de Rango Múltiple de Duncan.

4.2. Prueba de rangos múltiples de Duncan

Si las a muestras son de igual tamaño n , sus respectivos promedios se ordenan en forma descendente y el error estándar de cada promedio se estima por $S_{\bar{Y}_{i\bullet}} = \sqrt{\frac{MSE}{n}}$. Si existen diferencias en los tamaños muestrales entonces el error estándar de las medias es calculado usando en vez de n , la media armónica de las n_i , dada por, [4]:

$$n_h = \frac{a}{\sum_{i=1}^a n_i^{-1}} \quad (4.4)$$

En este procedimiento las diferencias absolutas de los pares de medias ordenadas, separadas p lugares (incluyendo a las medias a comparar), son comparadas a un nivel de significancia de γ , con los rangos de mínima significancia correspondientes dados por

$$R_p = r_\gamma(p, \mathbf{g}, \mathbf{l}) S_{\bar{Y}_{i\bullet}}, \quad p = 2, 3, \dots, a, \quad (4.5)$$

donde $r_\gamma(p, g.l)$ son los valores críticos de rangos significativos de Duncan, los cuales se hayan tabulados en distintos textos de diseño de experimentos y de probabilidad y estadística, con g.l los grados de libertad del MSE.

Procedimiento: Tal como describe [4], se compara la diferencia entre la media más grande y la más pequeña con el rango R_a . Luego, se Compara la diferencia entre la media más grande y la segunda media más pequeña con el rango R_{a-1} . Estas comparaciones continúan hasta que se haya comparado la mayor media con todas las demás. A continuación se compara la diferencia entre la segunda media más grande y la menor media con el rango R_{a-1} . Después, se compara la diferencia entre la segunda media más grande y la segunda más pequeña con el rango R_{a-2} , y así sucesivamente hasta que se comparan las $a(a-1)/2$ pares de medias posibles con el rango que les corresponda.

OJO: En las comparaciones donde la diferencia observada es mayor que el rango respectivo, se concluye que las medias correspondientes son estadísticamente distintas. Si una, dos o más medias caen entre otras dos que son estadísticamente iguales, entonces tales medias se consideran un grupo de medias estadísticamente iguales (incluyendo a las dos medias en los extremos de la comparación).

4.2.1. Ejemplo

Considere de nuevo el experimento de comparación de los cuatro métodos de ensamble. Para una significancia de $\gamma = 0.05$, $n = 4$, g.l = 12 y $MSE = 2.46$ se tiene $S_{\bar{Y}_{i\bullet}} = 0.78$. Los rangos mínimos significantes a usar son:

$$R_2 = r_{0.05}(2, 12)S_{\bar{Y}_{i\bullet}} = 3.08 \times 0.78 = 2.40$$

$$R_3 = r_{0.05}(3, 12)S_{\bar{Y}_{i\bullet}} = 3.23 \times 0.78 = 2.52$$

$$R_4 = r_{0.05}(4, 12)S_{\bar{Y}_{i\bullet}} = 3.33 \times 0.78 = 2.60$$

Las medias muestrales, de menor a mayor, son $\bar{Y}_A = 7.25$, $\bar{Y}_B = 8.50$, $\bar{Y}_D = 10.50$, $\bar{Y}_C = 12.75$. Entonces las comparaciones a realizar son como sigue:

4.3. Método de Tukey de comparaciones de todos los pares de medias, o rango estudentizado de Tukey

Diferencia Poblacional	Diferencia muestral	Decisión
$\mu_C - \mu_A$	$5.50 > 2.60 = R_4$	Diferentes
$\mu_C - \mu_B$	$4.25 > 2.52 = R_3$	Diferentes
$\mu_C - \mu_D$	$2.25 < 2.40 = R_2$	Iguales
$\mu_D - \mu_A$	$3.25 > 2.52 = R_3$	Diferentes
$\mu_D - \mu_B$	$2.00 < 2.40 = R_2$	Iguales
$\mu_B - \mu_A$	$1.25 < 2.40 = R_2$	Iguales

En este procedimiento se concluye que las medias de tratamientos estadísticamente iguales pueden agruparse así:

- GRUPO 1: $\mu_A = \mu_B$
- GRUPO 2: $\mu_B = \mu_D$
- GRUPO 3: $\mu_C = \mu_D$

Como puede observarse, se obtuvieron en este caso los mismos grupos del método LSD.

4.3. Método de Tukey de comparaciones de todos los pares de medias, o rango estudentizado de Tukey

Es un método de comparación de pares de medias más conservador que los dos anteriores (menos potente para detectar pequeñas diferencias), pero con la característica de que controla la tasa de error experimental. Sin embargo, proporciona intervalos de confianza para las diferencias por pares de medias más cortos que los arrojados por los métodos de Bonferroni y Scheffé. Puede usarse en diseños de bloques completos aleatorizados y diseños de bloques incompletos balanceados. Para un diseño DCA de un solo factor efectos fijos, las diferencias absolutas de pares de medias se comparan a un nivel de significancia con el valor crítico HSD_{ij} llamado la diferencia honestamente significativa, que corresponde a:

$$HSD_{ij} = \frac{1}{\sqrt{2}} q_{\gamma}(a, g.l) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (4.6)$$

donde $q_{\gamma}(a, g.l)$ es el valor crítico del rango estudentizado, cuyos valores se hayan tabulados, y $g.l$ son los grados de libertad del MSE. Los pares de medias de tratamien-

tos μ_i y μ_j difieren significativamente cuando: $D_{ij} = |\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}| > HSD_{ij}$. También puede construirse los I.C de Tukey para las diferencias de medias. Si el cero no pertenece a tal intervalo entonces las dos medias correspondientes son estadísticamente distintas:

$$(\mu_i - \mu_j) \in \left[(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) \pm \frac{1}{\sqrt{2}} q_\gamma(a, g, l) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right] \quad (4.7)$$

Nota 4.1. El rango estudentizado en términos de las medias de a muestras independientes, es definido como sigue: Suponga que se toma una muestra de tamaño n de cada una de a poblaciones con la misma distribución $N(0, \sigma^2)$ y suponga que \bar{Y}_{\min} y \bar{Y}_{\max} corresponden, respectivamente, a la media más pequeña y más grande de estas medias muestrales. Sea $S_p^2 = \frac{\sum_{i=1}^a (n_i - 1) S_i^2}{\sum_{i=1}^a (n_i - 1)}$, la estimación pooled de la varianza a partir de las a muestras. Entonces la siguiente variable aleatoria es el rango estudentizado,

$$q(a, g, l) = \frac{\bar{Y}_{\max} - \bar{Y}_{\min}}{S_p / \sqrt{n}} \quad (4.8)$$

donde g, l son los grados de libertad de S_p^2 , y $n_i = n$.

4.3.1. Ejemplo

De nuevo considere el caso de comparación de los cuatro métodos de ensamble: La diferencia honestamente significativa es $HSD_{ij} = \frac{1}{\sqrt{2}} q_{0.05}(4, 12) \sqrt{MSE \left(\frac{2}{4} \right)} = 4.20 \times 0.78 = 3.28$, para todas las comparaciones.

Diferencia Poblacional	Diferencia muestral absoluta	Decisión
$\mu_A - \mu_B$	$1.25 < 3.28$	Igual
$\mu_A - \mu_C$	$5.50 > 3.28$	Diferente
$\mu_A - \mu_D$	$3.25 < 3.28$	Igual
$\mu_B - \mu_C$	$4.25 > 3.28$	Diferente
$\mu_B - \mu_D$	$2.00 < 3.28$	Igual
$\mu_C - \mu_D$	$2.25 < 3.28$	Igual

Con este procedimiento se concluye que las medias de tratamientos estadísticamente iguales pueden agruparse así:

- GRUPO 1: $\mu_A = \mu_B = \mu_D$

- GRUPO 2: $\mu_C = \mu_D$

Como puede observarse hay un traslapo entre estos dos grupos.

4.4. Comparaciones de tratamientos con un control - Método de Dunnet

Algunas veces el experimentador puede estar interesado en comparar los efectos de un tratamiento especial llamado el control, con los efectos de cada uno de los otros tratamientos, bien sea en un DCA, en diseños de bloques completos aleatorizados y en diseños de bloques incompletos balanceados. Ej. Un experimento farmacéutico involucra una o más drogas experimentales juntas a comparar con una droga estándar que ha estado en el mercado por varios años. Frecuentemente el objetivo es comparar el efecto de cada droga experimental con el de la droga estándar pero no necesariamente con los efectos de las otras drogas experimentales. La droga estándar es el control. En muchos otros casos el tratamiento de control se refiere a la ausencia de tratamiento o la aplicación de un placebo.

Si se denota al tratamiento de control como el tratamiento 1, entonces comparar respecto al control implica probar: $H_0 : \mu_i = \mu_1$ vs. $H_1 : \mu_i \neq \mu_1$, con $i = 2, 3, 4, \dots, a$. La hipótesis nula se rechaza con nivel de significancia γ , si:

$$|\bar{Y}_{i\bullet} - \bar{Y}_{1\bullet}| > D_\gamma(a-1, g.l) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_1} \right)}, \quad (4.9)$$

donde $D_\gamma(a-1, g.l)$ es el valor crítico para la prueba Dunnet bilateral al nivel de significancia γ , y g.l es los grados de libertad del MSE.

4.5. Métodos de Dunnett unilaterales

Si el propósito de un experimento es determinar cuál de los tratamientos experimentales da una respuesta promedio significativamente más alta que el tratamiento de control, deben usarse un intervalos de confianza unilaterales. Para un DCA de un solo factor, los intervalos de confianza unilaterales simultáneos de Dunnett del $(1 - \gamma)100\%$ para las diferencias $\mu_i - \mu_1$ (donde el nivel 1 es el tratamiento de control,

$i = 2, 3, \dots, a$), corresponden a:

$$\mu_i - \mu_1 \geq (\bar{Y}_{i\bullet} - \bar{Y}_{1\bullet}) - D_{1\gamma}(a-1, \mathbf{g}, \mathbf{l}) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_1} \right)}. \quad (4.10)$$

Por otro lado, si el objetivo es determinar cuál de los tratamientos experimentales da una respuesta promedio significativamente menor a la del control, entonces los intervalos de confianza simultáneos están dados por:

$$\mu_i - \mu_1 \leq (\bar{Y}_{i\bullet} - \bar{Y}_{1\bullet}) + D_{1\gamma}(a-1, \mathbf{g}, \mathbf{l}) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_1} \right)} \quad (4.11)$$

4.6. Ejemplos de aplicación

4.6.1. Con datos experimento de las lechugas

Considere de nuevo el experimento de las lechugas presentado en la Sección 3.2.7. Suponga al tratamiento “0” libras de nitrato como el nivel de control. A continuación se proporciona un programa R para el ajuste del modelo ANOVA, la obtención de los tests de Tukey, Duncan, LSD, y las comparaciones tipo bilateral de Dunnett. En este programa, las funciones `HSD.test()`, `LSD.test()` y `duncan.test()` dependen de la librería `agricolae`, en tanto que la función `glht()` depende de la librería `multcomp`. La función `TukeyHSD`, es de la librería `stats` por tanto siempre está disponible. Interprete los resultados y agrupe las medias de tratamiento.

Código R 4.6.1. Programa R para ajuste modelo Anova y comparación medias de tratamientos, ejemplo 3.2.7.

```
library(multcomp)
library(agricolae)
#ENTRADA DE DATOS
ni=rep(4,5) #No. réplicas por nivel
#obseve que los renglones de datos corresponden a cada nivel desde 0 a 200, respectivamente
diseño2=data.frame(nitrato=factor(rep(c(0,50,100,150,200),times=ni)),nlechug=scan())
104 114 90 140
134 130 144 174
146 142 152 156
147 160 160 163
```

```

131 148 154 163

attach(diseño2)

#AJUSTE MODELO DE ANÁLISIS DE VARIANZA
modelo2=aov(nlechug~nitrato)

#INTERVALOS DE TUKEY
TukeyHSD(modelo2, conf.level = 0.95) #La función es cargada por defecto por librería stats
plot(TukeyHSD(modelo2, conf.level = 0.95),cex.lab=0.8,las=1) #Grafica I.C de Tukey para
#las diferencias de medias

HSD.test(modelo2,"nitrato", group=TRUE,console=TRUE) #Comparaciones de Tukey
duncan.test(modelo2,"nitrato",alpha=0.05,group=TRUE,console=TRUE) #Rango múltiple de Duncan
LSD.test(modelo2,"nitrato",group=TRUE,console=TRUE) #Método LSD
#Comparaciones de Dunnett; la función glht asume que primer nivel es el control
summary(glht(modelo2, linfct=mcp(nitrato="Dunnett"),alternative="two.sided"))
confint(glht(modelo2, linfct=mcp(nitrato="Dunnett"),alternative="two.sided"))

#I.C Dunnett unilateral inferior del 95%
confint(glht(modelo2, linfct=mcp(nitrato="Dunnett"),alternative="greater"))
detach(diseño2)

```

Salida R 4.6.1. Resultados para comparaciones según LSD

```

> LSD.test(modelo2,"nitrato",group=TRUE,console=TRUE) #Método LSD
Study: modelo2 ~ "nitrato"
LSD t Test for nlechug
Mean Square Error: 222.5333
nitrato, means and individual ( 95 %) CI

```

	nlechug	std r	LCL	UCL	Min	Max
0	112.0	21.102922 4	96.102	127.898	90	140
100	149.0	6.218253 4	133.102	164.898	142	156
150	157.5	7.141428 4	141.602	173.398	147	163
200	149.0	13.490738 4	133.102	164.898	131	163
50	145.5	19.891372 4	129.602	161.398	130	174

```

Alpha: 0.05 ; DF Error: 15
Critical Value of t: 2.13145
Least Significant Difference 22.48317
Means with the same letter are not significantly different

```

Groups, Treatments and means

a	150	157.5
a	100	149
a	200	149
a	50	145.5
b	0	112

Salida R 4.6.2. Resultados según método de Tukey. Vea además la Figura 4.1.

```
> TukeyHSD(modelo2, conf.level = 0.95) #I.C de Tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = nlechug ~ nitrato)
$nittrato
      diff      lwr      upr    p adj
50-0      3.350000e+01  0.9276412 66.07236 0.0424154
100-0     3.700000e+01  4.4276412 69.57236 0.0225650
150-0     4.550000e+01 12.9276412 78.07236 0.0047391
200-0     3.700000e+01  4.4276412 69.57236 0.0225650
100-50    3.500000e+00 -29.0723588 36.07236 0.9970821
150-50    1.200000e+01 -20.5723588 44.57236 0.7847018
200-50    3.500000e+00 -29.0723588 36.07236 0.9970821
150-100   8.500000e+00 -24.0723588 41.07236 0.9248001
200-100   2.842171e-14 -32.5723588 32.57236 1.0000000
200-150  -8.500000e+00 -41.0723588 24.07236 0.9248001

> HSD.test(modelo2,"nitrato", group=TRUE,console=TRUE) #Comparaciones de Tukey
Study: modelo2 ~ "nitrato"
HSD Test for nlechug
Mean Square Error: 222.5333
nitrato, means
      nlechug      std r Min Max
0      112.0 21.102922 4  90 140
100    149.0  6.218253 4 142 156
150    157.5  7.141428 4 147 163
200    149.0 13.490738 4 131 163
50     145.5 19.891372 4 130 174

Alpha: 0.05 ; DF Error: 15
Critical Value of Studentized Range: 4.366985
Honestly Significant Difference: 32.57236
Means with the same letter are not significantly different.
```


Groups, Treatments and means

a	150	157.5
a	100	149
a	200	149
a	50	145.5
b	0	112

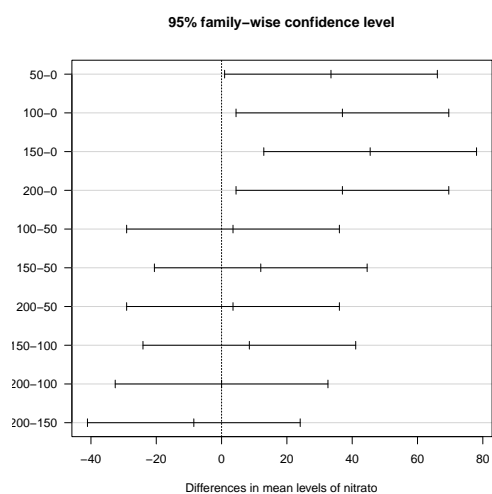


Figura 4.1: I.C de Tukey para las diferencias de medias de tratamiento en 3.2.7. Es resultado del uso de `plot(TukeyHSD(...))`; vea Código R 4.6.1.

En la Figura 4.1 puede observar la representación de los límites de los intervalos de confianza tipo Tukey para la diferencia por pares de medias de tratamientos. Observe la línea de referencia vertical en cero. Si algún intervalo es cortado por tal línea entonces el par de medias correspondientes no son estadísticamente distintas, de lo contrario, si el intervalo no es cortado por tal línea, entonces el par de medias correspondientes son estadísticamente diferentes.

Salida R 4.6.3. Resultados para comparaciones según Duncan.

```
> duncan.test(modelo2, "nitrato", alpha=0.05, group=TRUE, console=TRUE) #Rango múltiple de Duncan
Study: modelo2 ~ "nitrato"
Duncan's new multiple range test
for nlechug
Mean Square Error: 222.5333
```

```

nitrato, means
      nlechug      std r Min Max
0      112.0 21.102922 4  90 140
100     149.0  6.218253 4 142 156
150     157.5  7.141428 4 147 163
200     149.0 13.490738 4 131 163
50      145.5 19.891372 4 130 174
Alpha: 0.05 ; DF Error: 15
Critical Range
      2      3      4      5
22.48317 23.56843 24.24287 24.70233
Means with the same letter are not significantly different.
Groups, Treatments and means
a      150      157.5
a      100      149
a      200      149
a      50      145.5
b      0      112

```

Salida R 4.6.4. Resultados comparaciones de Dunnett usando como control el primer nivel por defecto: “0” Libras.

```

> summary(glht(modelo2, linfct=mcp(nitrato="Dunnett"),alternative="two.sided"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = nlechug ~ nitrato)
Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
50 - 0 == 0      33.50      10.55   3.176  0.0208 *
100 - 0 == 0      37.00      10.55   3.508  0.0109 *
150 - 0 == 0      45.50      10.55   4.313  0.0022 **
200 - 0 == 0      37.00      10.55   3.508  0.0107 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> confint(glht(modelo2, linfct=mcp(nitrato="Dunnett"),alternative="two.sided"))
      Simultaneous Confidence Intervals
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = nlechug ~ nitrato)
Quantile = 2.7286
95% family-wise confidence level

```

```

Linear Hypotheses:
              Estimate   lwr   upr
50 - 0 == 0  33.5000    4.7178 62.2822
100 - 0 == 0  37.0000    8.2178 65.7822
150 - 0 == 0  45.5000   16.7178 74.2822
200 - 0 == 0  37.0000    8.2178 65.7822

> confint(glht(modelo2, linfct=mcp(nitrato="Dunnett"), alternative="greater"))
              Simultaneous Confidence Intervals
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = nlechug ~ nitrato)
Quantile = -2.3565
95% family-wise confidence level
Linear Hypotheses:
              Estimate   lwr   upr
50 - 0 <= 0  33.5000    8.6431  Inf
100 - 0 <= 0  37.0000   12.1431  Inf
150 - 0 <= 0  45.5000   20.6431  Inf
200 - 0 <= 0  37.0000   12.1431  Inf

```

Nota 4.2. El nivel de referencia o de control para el método de Dunnett puede especificarse de modo que no sea usado por defecto el primer nivel del factor. Como ilustración, aunque no tiene sentido práctico, tomemos en el ejemplo de las lechugas el nivel de 50 libras de nitrato como nivel de control. Para ello, usamos la función `relevel()` y ajustamos de nuevo el modelo anova con los niveles del factor reordenados. veamos.

Código R 4.6.2. Programa R para ajuste modelo Anova y comparación medias de tratamientos por Dunnet pero usando como nivel de referencia el de 50 Lb de nitrato en el ejemplo 3.2.7.

```

#Para hacer Dunnet cambiando nivel de control a 50 y el resultado lo guardamos en nuevo objeto R
#aquí denominado nitrato2,
nitrato2=relevel(nitrato,ref="50")
modelo3=aov(nlechug~nitrato2)
summary(glht(modelo3, linfct=mcp(nitrato2="Dunnett"), alternative="two.sided"))
confint(glht(modelo3, linfct=mcp(nitrato2="Dunnett"), alternative="two.sided"))

```

Salida R 4.6.5. Resultados comparaciones de Dunnett tomando como nivel de control "50" Libras

```
> summary(glht(modelo3, linfct=mcp(nitrato2="Dunnett"), alternative="two.sided"))

      Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = nlechug ~ nitrato2)

Linear Hypotheses:

              Estimate Std. Error t value Pr(>|t|)
0 - 50 == 0      -33.50      10.55  -3.176  0.0208 *
100 - 50 == 0       3.50      10.55   0.332  0.9914
150 - 50 == 0      12.00      10.55   1.138  0.6252
200 - 50 == 0       3.50      10.55   0.332  0.9914
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
(Adjusted p values reported -- single-step method)

> confint(glht(modelo3, linfct=mcp(nitrato2="Dunnett"), alternative="two.sided"))

      Simultaneous Confidence Intervals

Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = nlechug ~ nitrato2)

Quantile = 2.7261

95% family-wise confidence level

Linear Hypotheses:

              Estimate   lwr      upr
0 - 50 == 0    -33.5000 -62.2561  -4.7439
100 - 50 == 0     3.5000 -25.2561  32.2561
150 - 50 == 0    12.0000 -16.7561  40.7561
200 - 50 == 0     3.5000 -25.2561  32.2561
```

4.6.2. Con datos experimento sobre la mejora de la productividad

Considere de nuevo el ejemplo 3.2.6 sobre mejora de la productividad. Considere los diferentes métodos vistos para la comparación de medias de tratamiento y agrúpelas de acuerdo a cada método. Para Dunnett considere como nivel de control el nivel bajo.

Código R 4.6.3. Programa R para ajuste modelo Anova y comparación medias de tratamientos por Dunnett pero usando como nivel de referencia el nivel bajo del factor ID en el Ejemplo 3.2.6.

```

library(multcomp)
library(agricolae)
#ENTRADA DE DATOS
ni=c(6,9,12)
diseño=data.frame(ID=factor(rep(c("alto","bajo","medio"),times=ni)),mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
#AJUSTE MODELO ANOVA UN FACTOR
modelo=aov(mprod~ID)

#INTERVALOS DE TUKEY
TukeyHSD(modelo, conf.level = 0.95) #Función es cargada por defecto por librería stats
plot(TukeyHSD(modelo, conf.level = 0.95),cex.lab=0.8,las=1)
HSD.test(modelo,"ID", group=TRUE,console=TRUE) #Comparaciones de Tukey
duncan.test(modelo,"ID",alpha=0.05,group=TRUE,console=TRUE) #Rango múltiple de Duncan
LSD.test(modelo,"ID",group=TRUE,console=TRUE) #Método LSD

#Como Dunnett asume que primer nivel es el control
#Para cambiar el nivel de control usamos relelevel
ID2=relevel(ID,ref="bajo")
modelo2=aov(mprod~ID2)
summary(glht(modelo2, linfct=mcp(ID="Dunnett"),alternative="two.sided"))
confint(glht(modelo2, linfct=mcp(ID="Dunnett"),alternative="two.sided"))
detach(diseño)

```

Salida R 4.6.6. Resultados para comparaciones de medias de tratamientos en ejemplo

3.2.6. Ver además Figura 4.2.

```

> LSD.test(modelo,"ID",group=TRUE,console=TRUE) #Método LSD
Study: modelo ~ "ID"
LSD t Test for mprod
Mean Square Error: 0.6400926
ID, means and individual ( 95 %) CI

```

	mprod	std	r	LCL	UCL	Min	Max
alto	9.200000	0.8671793	6	8.525885	9.874115	7.8	10.1
bajo	6.877778	0.8135997	9	6.327365	7.428191	5.8	8.2
medio	8.133333	0.7571878	12	7.656662	8.610005	6.7	9.4

```

Alpha: 0.05 ; DF Error: 24
Critical Value of t: 2.063899
Minimum difference changes for each comparison
Means with the same letter are not significantly different
Groups, Treatments and means
a      alto      9.2
b      medio     8.133333
c      bajo      6.877778
> TukeyHSD(modelo, conf.level = 0.95) #Función es cargada por defecto por librería stats
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = mprod ~ ID)
$ID              diff          lwr          upr          p adj
bajo-alto -2.322222 -3.3752471 -1.26919735 0.0000335
medio-alto -1.066667 -2.0656538 -0.06767956 0.0347870
medio-bajo  1.255556  0.3745317  2.13657937 0.0043755

> HSD.test(modelo,"ID", group=TRUE,console=TRUE) #Comparaciones de Tukey
Study: modelo ~ "ID"
HSD Test for mprod
Mean Square Error: 0.6400926
ID, means
      mprod      std  r Min  Max
alto  9.200000 0.8671793  6 7.8 10.1
bajo  6.877778 0.8135997  9 5.8  8.2
medio 8.133333 0.7571878 12 6.7  9.4
Alpha: 0.05 ; DF Error: 24
Critical Value of Studentized Range: 3.531697
Harmonic Mean of Cell Sizes 8.307692
Honestly Significant Difference: 0.9803128
Means with the same letter are not significantly different.
Groups, Treatments and means
a      alto      9.2
b      medio     8.133
c      bajo      6.878

> duncan.test(modelo,"ID",alpha=0.05,group=TRUE,console=TRUE) #Rango múltiple de Duncan
Study: modelo ~ "ID"
Duncan's new multiple range test
for mprod
Mean Square Error: 0.6400926

```

```

ID, means
      mprod      std  r Min  Max
alto  9.200000 0.8671793  6 7.8 10.1
bajo  6.877778 0.8135997  9 5.8  8.2
medio 8.133333 0.7571878 12 6.7  9.4
Alpha: 0.05 ; DF Error: 24
Critical Range
      2      3
0.8101856 0.8509385
Harmonic Mean of Cell Sizes  8.307692
Different value for each comparison
Means with the same letter are not significantly different.
Groups, Treatments and means
a      alto      9.2
b      medio     8.133
c      bajo      6.878
#Dunnet con nivel bajo definido como el control
> summary(glht(modelo2, linfct=mcp(ID2="Dunnett"),alternative="two.sided"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = mprod ~ ID2)
Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
alto - bajo == 0    2.3222     0.4217   5.507 2.28e-05 ***
medio - bajo == 0    1.2556     0.3528   3.559 0.00306 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
(Adjusted p values reported -- single-step method)
> confint(glht(modelo2, linfct=mcp(ID2="Dunnett"),alternative="two.sided"))
      Simultaneous Confidence Intervals
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = mprod ~ ID2)
Quantile = 2.3525
95% family-wise confidence level
Linear Hypotheses:
      Estimate lwr      upr
alto - bajo == 0  2.3222   1.3303 3.3142
medio - bajo == 0  1.2556   0.4256 2.0855

```

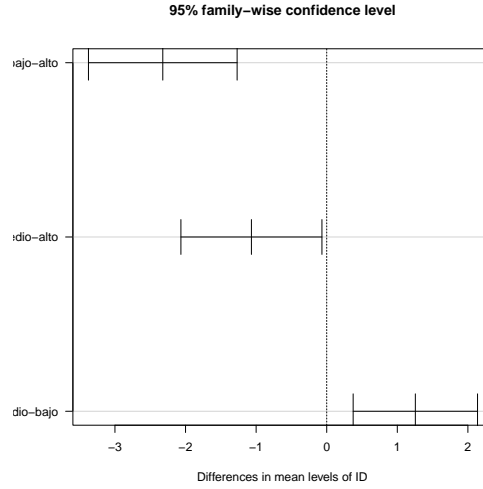


Figura 4.2: I.C de Tukey para las diferencias de medias de tratamiento en 3.2.6. Es resultado del uso de `plot(TukeyHSD(...))`; vea Código R 4.6.3.

4.7. Comparaciones mediante contrastes

Un contraste es una combinación lineal de parámetros del mismo tipo, por ejemplo,

- Contraste de medias de tratamientos: $W = \sum_{i=1}^a c_i \mu_i$, con $\sum_{i=1}^a c_i = 0$.
- Contraste de efectos de tratamientos: $V = \sum_{i=1}^a b_i \alpha_i$, con $\sum_{i=1}^a b_i = 0$.

Muchas veces, el interés en la experimentación es obtener información para probar alguna hipótesis relativa a un contraste dado. Para hacer inferencias sobre un contraste de medias nos basamos en la distribución muestral del contraste estimado, bajo las suposiciones del DCA:

$$\widehat{W} = \sum_{i=1}^a c_i \bar{Y}_{i\bullet} \sim N \left(\sum_{i=1}^a c_i \mu_i, \sum_{i=1}^a c_i^2 \frac{\sigma^2}{n_i} \right) \quad (4.12)$$

4.7.1. Pruebas de hipótesis sobre un contraste de medias

Bajo los supuestos del DCA, el estadístico de prueba para los test relativos a contrastes de medias de tratamientos, es:

$$T_0 = \frac{\widehat{W} - W_0}{\sqrt{MSE \sum_{i=1}^a \frac{c_i^2}{n_i}}} \sim t_{N-a} \quad (4.13)$$

Tabla 4.1: Tests de hipótesis sobre contrastes de medias de tratamientos

Tests	Rechazar H_0 si	VP
$H_0 : W = W_0$ $H_1 : W \neq W_0$	$ T_0 > t_{\gamma/2, N-a}$	$P(t_{N-a} > T_0)$
$H_0 : W = W_0$ $H_1 : W > W_0$	$T_0 > t_{\gamma, N-a}$	$P(t_{N-a} > T_0)$
$H_0 : W = W_0$ $H_1 : W < W_0$	$T_0 < -t_{\gamma, N-a}$	$P(t_{N-a} < T_0)$

En particular, el test de significancia de un contraste, es decir cuando $H_0 : W = 0$ vs. $H_1 : W \neq 0$, también puede realizarse a través de un ANOVA, calculando la suma de cuadrados del contraste SSW:

$$SS_W = \frac{\widehat{W}^2}{\left(\sum_{i=1}^a c_i^2 / n_i \right)}, \quad (4.14)$$

que tiene un grado de libertad, y construyendo el estadístico F para la prueba como la razón de SSW sobre el MSE del modelo. Es decir:

$$F_0 = (SS_W / MSE) \stackrel{H_0}{\sim} f_{1, N-a} \quad (4.15)$$

H_0 es rechazada si $F_0 > f_{\gamma, 1, N-a}$, o bien si $VP = P(f_{1, N-a} > F_0) < \gamma$, con γ siendo el nivel de significancia usado.

4.7.2. Ejemplo, [2]

Un experimento fue llevado a cabo para determinar el desempeño relativo de siete filtros de membrana para soportar el crecimiento de colonias de bacterias. Los siete tipos de filtros son considerados como los siete niveles del factor y son codificados de 1 a 7. Los tipos de filtros 1, 4 y 7 fueron recibidos preesterilizados. Se recolectaron datos sobre los conteos de las colonias de coliformes fecales de una muestra de agua de río que crecieron en cada filtro. Tres filtros de cada tipo fueron observados y los conteos promedios por colonia fueron:

$$\bar{Y}_{1\bullet} = 36.0, \bar{Y}_{2\bullet} = 18.0, \bar{Y}_{3\bullet} = 27.7, \bar{Y}_{4\bullet} = 28.0, \bar{Y}_{5\bullet} = 28.3, \bar{Y}_{6\bullet} = 37.7, \bar{Y}_{7\bullet} = 30.3;$$

el $MSE = 21.6$. Suponga que se desea probar la hipótesis de que *los filtros preesterilizados no difieren de los no preesterilizados en términos de los conteos promedio de colonias*, contra la hipótesis alternativa bilateral de que sí difieren.

Solución: La hipótesis de interés involucra un contraste de medias dado por

$$W = \frac{1}{3}(\mu_1 + \mu_4 + \mu_7) - \frac{1}{4}(\mu_2 + \mu_3 + \mu_5 + \mu_6),$$

de donde $\widehat{W} = 3.508$ y $SSW = (3.508)^2/(7/36)$. Observe que $c_i = 1/3$ para $i = 1, 4, 7$, y $c_i = 1/4$ para $i = 2, 3, 5, 6$. La hipótesis a probar es

$$H_0 : W = 0, \text{ vs. } H_1 : W \neq 0,$$

y el estadístico de prueba $F_0 = SSW/MSE = 2.93$; el valor crítico al 0.05 de significancia es $F_{0.05,1,14} = 4.6$.

Por tanto, no se obtuvo evidencia suficiente en contra de H_0 , y se concluye que los filtros preesterilizados no difieren significativamente de los que no se preesterilizaron. Un intervalo de confianza del 95 % para el contraste dado es

$$\widehat{W} \pm t_{0.025,14} \sqrt{MSE \sum_{i=1}^7 \frac{c_i^2}{n_i}} = 3.508 \pm 2.144787 \sqrt{21.6 \times 7/36} = (-0.888, 7.904).$$

4.7.3. Método de Bonferroni para comparaciones preplaneadas

Si m intervalos simultáneos de confianza son calculados para contrastes preplaneados, y si cada intervalo tiene un nivel de confianza de $(1 - \gamma^*)100\%$, entonces el nivel de confianza total es mayor o igual a $(1 - m\gamma^*)100\%$. Así, un experimentador puede asegurar que el nivel de confianza total es al menos $(1 - \gamma)100\%$ ajustando $\gamma^* = \gamma/m$ para un intervalo de confianza individual. Esto se conoce como el método de Bonferroni para intervalos de confianza simultáneos. Una fórmula para un conjunto de intervalos de confianza simultáneos de $(1 - \gamma)100\%$, para m contrastes preplaneados $\sum_{i=1}^a c_i \mu_i$ en un DCA con un solo factor efectos fijos, es:

$$\sum_{i=1}^a c_i \mu_i \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet} \pm t_{\gamma/(2m), N-a} \sqrt{MSE \sum_{i=1}^a c_i^2 / n_i} \right) \quad (4.16)$$

Cuando m es muy grande, $\gamma/(2m)$ es muy pequeño, resultando posiblemente en intervalos de confianza extremadamente amplios. En este caso deben preferirse los métodos Scheffé o Tukey.

Si algunos de los m intervalos simultáneos son para respuestas medias, entonces los intervalos requeridos son de la forma:

$$\mu_i \in \left(\bar{Y}_{i\bullet} \pm t_{\gamma/(2m), N-a} \sqrt{MSE/n_i} \right). \quad (4.17)$$

Similarmente, reemplazando γ por γ/m en un conjunto de m hipótesis nulas cada una de la forma $H_0 : \sum_{i=1}^a c_i \mu_i = 0$, éstas pueden ser probadas contra sus respectivas hipótesis alternativas bilaterales a un nivel de significancia simultáneo de γ , usando el conjunto de reglas de decisión de la forma: rechazar H_0 si $SS_W/MSE > f_{\gamma/m, 1, N-a}$.

Nota 4.3. Según [2], el método de Bonferroni sólo puede usarse para pruebas preplaneadas sobre contrastes y medias. Un experimentador quien explora los datos y luego procede a calcular intervalos de confianza simultáneos para los pocos contrastes que parecen de interés, ha calculado efectivamente un número muy grande de intervalos, esto porque los contrastes interesantes son usualmente aquellos que parecen ser

significativamente diferentes de cero, y un cálculo mental aproximado de las estimaciones de un gran número de contrastes ha sido hecho para identificar estos contrastes interesantes. El método de Scheffé debería usarse para contrastes que fueron seleccionados después de que los datos fueron examinados.

4.7.4. Ejemplo

Considere de nuevo el Ejemplo 4.7.2. suponga que antes de que los datos fueran recogidos, los experimentadores habían planeado calcular un conjunto de intervalos de confianza simultáneos del 90 % para los siguientes $m = 3$ contrastes, los cuales han sido seleccionados de un estudio original adelantado previamente por otros investigadores:

$W_1 = \frac{1}{3}(\mu_1 + \mu_4 + \mu_7) - \frac{1}{4}(\mu_2 + \mu_3 + \mu_5 + \mu_6)$. Este contraste mide la diferencia en la respuesta promedio de los filtros preesterilizados y los no preesterilizados.

$W_2 = \frac{1}{2}(\mu_1 + \mu_7) - \frac{1}{5}(\mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$. Este contraste mide la diferencia en la respuestas promedio de los dos tipos de filtros con poros de tamaños graduados vs. los otros cinco tipo de filtros con poros de tamaños uniformes.

$W_3 = \frac{1}{6}(\mu_1 + \mu_2 + \mu_4 + \mu_5 + \mu_6 + \mu_7) - \mu_3$. Este contraste es la diferencia entre la respuesta promedio de los tipos de filtro que son recomendados por sus fabricantes para análisis bacteriológico del agua y el único tipo de filtro que es recomendado para pruebas de esterilidad de productos farmacéuticos o cosméticos.

Aplicando (4.16) a cada uno de los tres contrastes con $t_{0.10/(2 \times 3), 14} = 2.338$, se obtiene lo siguiente:

$$W_1 \in (-1.283, 8.299), \quad W_2 \in (-0.039, 10.459), \quad W_3 \in (-4.759, 8.793).$$

Note que los tres intervalos incluyen al cero, aunque el segundo está cerca de excluirlo. Así, a un nivel de significancia total de 0.10, concluimos que los contrastes no son significativos y no existen diferencias entre los respectivos tipos de filtros comparados en cada caso.

4.7.5. Método de Scheffé

De acuerdo a [2], este método está diseñado para probar todos los contrastes que pudieran interesarle al experimentador, sin el inconveniente de inflar el error tipo I. Scheffé Proporciona un conjunto de intervalos de confianza simultáneos del $(1 - \gamma)100\%$ cuyos anchos son sólo determinados por el número de tratamientos y el número de observaciones en el experimento, sin importar cuántos contrastes son de interés.

Para a tratamientos en un DCA con un solo factor, los intervalos de confianza simultáneos de $(1 - \gamma)100\%$ para todos los contrastes $\sum_{i=1}^a c_i \mu_i$ está dado por,

$$\sum_{i=1}^a c_i \mu_i \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet} \pm \sqrt{(a-1) f_{\gamma, a-1, N-a}} \times \sqrt{MSE \sum_{i=1}^a c_i^2 / n_i} \right) \quad (4.18)$$

Si interesan intervalos de confianza simultáneos para cualesquier número de las medias de nivel junto con cualesquier contrastes, la anterior expresión queda de la siguiente manera, [2]:

$$\sum_{i=1}^a c_i \mu_i \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet} \pm \sqrt{a f_{\gamma, a, N-a}} \times \sqrt{MSE \sum_{i=1}^a c_i^2 / n_i} \right), \quad (4.19)$$

junto con,

$$\mu_i \in \left(\bar{Y}_{i\bullet} \pm \sqrt{a f_{\gamma, a, N-a}} \times \sqrt{MSE / n_i} \right) \quad (4.20)$$

4.7.6. Ejemplo

Volvamos al Ejemplo 4.7.2 sobre los filtros. Examinando los resultados se observa que los filtros tipo 2 parecen tener un conteo promedio mucho más bajo que los otros tipos, por tanto, uno puede desear recalcular cada uno de los tres intervalos en el Ejemplo 4.7.4, excluyendo el filtro tipo 2. También puede ser de interés comparar los filtros tipo 1 y 6, los cuales muestran los conteos promedios de colonias más altos, con los otros filtros. Estos no son contrastes preplaneados, sino que han llegado a ser de interés sólo después de que los datos fueron examinados, y por tanto, necesitamos usar el método de Scheffé de comparaciones múltiples. En resumen, estamos interesados en los siguientes 12 contrastes:

$$W_1 = \frac{1}{3}(\mu_1 + \mu_4 + \mu_7) - \frac{1}{3}(\mu_3 + \mu_5 + \mu_6)$$

$$W_2 = \frac{1}{2}(\mu_1 + \mu_7) - \frac{1}{4}(\mu_3 + \mu_4 + \mu_5 + \mu_6)$$

$$W_3 = \frac{1}{5}(\mu_1 + \mu_4 + \mu_5 + \mu_6 + \mu_7) - \mu_3$$

$$W_4 = \mu_1 - \mu_3, W_5 = \mu_1 - \mu_4, W_6 = \mu_1 - \mu_5, W_7 = \mu_1 - \mu_6, W_8 = \mu_1 - \mu_7,$$

$$W_9 = \mu_6 - \mu_3, W_{10} = \mu_6 - \mu_4, W_{11} = \mu_6 - \mu_5, W_{12} = \mu_6 - \mu_7.$$

Usando un nivel de confianza del 90 % y aplicando la expresión en (4.18) con $a = 7$, $N = 21$, $MSE = 21.6$, se obtienen los siguientes intervalos:

$$W_1 \in (-7.83, 8.23), W_2 \in (-5.79, 11.24), W_3 \in (-6.42, 15.14),$$

$$W_4 \in (-5.61, 22.21), W_5 \in (-5.91, 21.91), W_6 \in (-6.21, 21.61),$$

$$W_7 \in (-15.61, 12.21), W_8 \in (-8.21, 19.61), W_9 \in (-3.91, 23.91),$$

$$W_{10} \in (-4.21, 23.61), W_{11} \in (-4.51, 23.31), W_{12} \in (-6.51, 21.31).$$

Los anteriores intervalos son todos bastante amplios e incluyen el cero (0). Consecuentemente, a la tasa de error total de 0.10, no se encuentran contrastes significativos.

4.7.7. Relación entre ANOVA y el método de Scheffé.

El test ANOVA y el método de Scheffé de comparaciones múltiples son equivalentes en el siguiente sentido: El test ANOVA rechazará la hipótesis nula $H_0 : \mu_1 = \mu_2 = \dots = \mu_a$ a un nivel de significancia γ , si al menos uno de los intervalos de confianza, entre el número infinito de intervalos de confianza simultáneos de Scheffé del $(1 - \gamma)100\%$, excluye al cero. Sin embargo, los intervalos que excluyen al cero pueden no estar entre aquellos para los contraste que interesan examinar.

Los otros métodos de comparaciones múltiples no se relacionan al test ANOVA en esta forma. Es posible que cuando se usa uno de los otros métodos de comparaciones múltiples, que uno o más intervalos en un conjunto simultáneo de $(1 - \gamma)100\%$ excluya al cero, mientras que el test ANOVA no rechaza H_0 al nivel de significancia γ . Por tanto, si contrastes específicos de interés han sido identificados antes de correr el experimento y un método de comparaciones múltiples distinto al de Scheffé va

a ser usado, entonces los datos deben analizarse usando sólo el procedimiento de comparación múltiple.

En resumen:

1. Método de Bonferroni para comparaciones preplaneadas: aplica para cualesquiera m contrastes estimables preplaneados. Da intervalos más cortos que los otros métodos si m es pequeño. Puede usarse con cualquier diseño. No se recomienda su uso con contrastes formulados post experimento.
2. Método Scheffé para todas las comparaciones: Aplica a cualesquiera m contrastes estimables. Da intervalos más cortos que Bonferroni si m es grande. Permite la exploración de los datos. Puede usarse para cualquier diseño.
3. Método de Tukey para todas las comparaciones por pares: Es el mejor para todas las comparaciones por pares. Puede usarse en DCA, diseños de bloques aleatorizados y de bloques incompletos balanceados. Se cree que es aplicable en otros diseños. Puede extenderse para incluir todos los contrastes pero Scheffé es generalmente mejor para estos.
4. Método de Dunnett para comparaciones de tratamientos vs. control: Es el mejor para todos los contrastes de tratamientos vs. control. Puede usarse en DCA, diseño de bloques aleatorizados y de bloques incompletos balanceados.

4.7.8. Ejemplo

Retome el Ejemplo 3.2.6. Formule, estime y pruebe la significancia del contraste apropiado para probar si el promedio de mejora en productividad a un nivel alto de gasto en investigación y desarrollo, es:

1. Significativamente diferente a los otros dos niveles
2. Significativamente mayor que el de los otros dos niveles.

Observe que el contraste pedido implica comparar la media en el nivel alto contra el promedio de las medias de los otros dos niveles, luego, denotando los niveles alto,

bajo y medio por $i = 1, 2, 3$, respectivamente:

$$W = \mu_1 - \frac{\mu_2 + \mu_3}{2}$$

En 1. el test de hipótesis a realizar es $H_0 : W = 0$ vs. $H_1 : W \neq 0$; en 2. el test es $H_0 : W = 0$ vs. $H_1 : W > 0$. En R podemos hallar la estimación del contraste, probar su significancia y construir intervalos de confianza del 95 % como se muestra a continuación, mediante la función R `fit.contrast()` de la librería `gmodels`, o bien combinado las funciones R `confint()` o `summary()` con la función `glht()` de la librería `multcomp`. Para el punto 2, usaremos la con la función `summary()` con `glht()`, en esta última usando el argumento `alternative="greater"`.

Código R 4.7.1. Programa R para ajuste modelo Anova y prueba de significancia del contraste nivel altos. vs. promedio de los otros dos niveles, ejemplo 3.2.6.

```
library(multcomp)
library(gmodels)
#ENTRADA DE DATOS
ni=c(6,9,12)
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"), times=ni)), mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
#AJUSTE MODELO ANOVA UN FACTOR
modelo=aov(mprod~ID)
#CONTRASTE ID: ALTO VS. DEMÁS, TEST T Y I.C DEL 95%:
fit.contrast(modelo, "ID", rbind(":ALTO VS. DEMAS"=c(1, -1/2, -1/2)), conf=0.95)
#OTRA FORMA PARA CÁLCULO DEL CONTRASTE ID. ALTO VS. DEMÁS, Y SU I.C DEL 95%
contrasteb=rbind("alto vs. demás " =c(1, -1/2, -1/2))
confint(glht(modelo, linfct=mcp(ID=contrasteb), alternative="two.sided"))
summary(glht(modelo, linfct=mcp(ID=contrasteb), alternative="two.sided"))

#TEST ID. ALTO MAYOR QUE DEMÁS,
summary(glht(modelo, linfct=mcp(ID=contrasteb), alternative="greater"))
detach(diseño)
```

Salida R 4.7.1. Resultados para prueba sobre contraste pedido para datos del Ejemplo 3.2.6.

```
> fit.contrast(modelo,"ID",rbind("ALTO VS. DEMAS"=c(1,-1/2,-1/2)),conf=0.95)
              Estimate Std. Error t value      Pr(>|t|) lower CI upper CI
ID:ALTO VS. DEMAS 1.694444   0.3712111 4.564638 0.0001256483 0.9283023 2.460587

> confint(glht(modelo,linfct=mcp(ID=contrasteb),alternative="two.sided"))
      Simultaneous Confidence Intervals
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Quantile = 2.0639
95% family-wise confidence level
Linear Hypotheses:
              Estimate lwr      upr
alto vs. demás == 0 1.6944   0.9283 2.4606

> summary(glht(modelo,linfct=mcp(ID=contrasteb),alternative="two.sided"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Linear Hypotheses:
              Estimate Std. Error t value Pr(>|t|)
alto vs. demás == 0   1.6944     0.3712   4.565 0.000126 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> summary(glht(modelo,linfct=mcp(ID=contrasteb),alternative="greater"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Linear Hypotheses:
              Estimate Std. Error t value      Pr(>t)
alto vs. demás <= 0   1.6944     0.3712   4.565 6.28e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Salida R 4.7.2. Resultados para prueba sobre contraste pedido para datos del Ejemplo 3.2.6.

```
> fit.contrast(modelo,"ID",rbind("ALTO VS. DEMAS"=c(1,-1/2,-1/2)),conf=0.95)
              Estimate Std. Error t value      Pr(>|t|) lower CI upper CI
ID:ALTO VS. DEMAS 1.694444   0.3712111 4.564638 0.0001256483 0.9283023 2.460587

> confint(glht(modelo,linfct=mcp(ID=contrasteb),alternative="two.sided"))
      Simultaneous Confidence Intervals
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Quantile = 2.0639
95% family-wise confidence level
Linear Hypotheses:
              Estimate lwr      upr
alto vs. demás == 0 1.6944   0.9283 2.4606

> summary(glht(modelo,linfct=mcp(ID=contrasteb),alternative="two.sided"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Linear Hypotheses:
              Estimate Std. Error t value Pr(>|t|)
alto vs. demás == 0   1.6944      0.3712   4.565 0.000126 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> summary(glht(modelo,linfct=mcp(ID=contrasteb),alternative="greater"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = mprod ~ ID)
Linear Hypotheses:
              Estimate Std. Error t value      Pr(>t)
alto vs. demás <= 0   1.6944      0.3712   4.565 6.28e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

4.7.9. Ejemplo

Pruebe en el Ejemplo 3.2.7 si

1. Aplicar nitrato produce una producción promedio diferente respecto a no aplicar nitrato
2. También pruebe si aplicar nitrato eleva la producción promedio respecto a no aplicar nitrato.

En este caso, denotando por $i = 1, 2, \dots, 5$, a los niveles de nitrato ordenados de menor a mayor, en 1. y 2. se quiere probar sobre la diferencia del promedio de las medias para los niveles 2 a 5 menos la media del nivel 1, es decir,

$$W = \frac{\sum_{i=2}^5 \mu_i}{4} - \mu_1,$$

de modo que en 1. el test a realizar es $H_0 : W = 0$ vs. $H_1 : W \neq 0$ y en 2. el test a realizar es $H_0 : W = 0$ vs. $H_1 : W > 0$. En R procedemos como se muestra a continuación.

Código R 4.7.2. Programa R para ajuste modelo Anova y prueba de significancia del contraste con nitrato vs. sin nitrato en Ejemplo 3.2.7.

```
library(multcomp)
library(gmodels)
#ENTRADA DE DATOS
ni=rep(4,5) #No. réplicas por nivel
#obseve que los renglones de datos corresponden a cada nivel desde 0 a 200, respectivamente
diseño2=data.frame(nitrato=factor(rep(c(0,50,100,150,200),times=ni)),nlechug=scan())
104 114 90 140
134 130 144 174
146 142 152 156
147 160 160 163
131 148 154 163

diseño2

attach(diseño2)
```

```
#AJUSTE MODELO DE ANÁLISIS DE VARIANZA
modelo2=aov(nlechug~nitrato)
#CÁLCULO DEL CONTRASTE CON NITRATO VS. SIN NITRATO, SU TEST T, Y I.C DEL 95%
contraste=fit.contrast(modelo2,"nitrato",
rbind(":con nitrato vs. sin nitrato"=c(-1,1/4,1/4,1/4,1/4)),conf=0.95)

#TEST CONTRASTE CON "NITRATO MAYOR QUE SIN NITRATO" Y SU I.C DEL 95%
contr=rbind("con nitrato vs. sin nitrato"= c(-1,1/4,1/4,1/4,1/4))
summary(glht(modelo2,linfct=mcp(nitrato=contr),alternative="greater"))
detach(diseño2)
```

Salida R 4.7.3. Resultados para prueba sobre contraste pedido para datos del Ejemplo 3.2.7.

```
> fit.contrast(modelo2,"nitrato",
+ rbind(":con nitrato vs. sin nitrato"=c(-1,1/4,1/4,1/4,1/4)),conf=0.95)
              Estimate Std. Error t value      Pr(>|t|) lower CI upper CI
nitrato:con nitrato vs. sin nitrato    38.25   8.339165 4.58679 0.0003561942 20.47549 56.02451

> summary(glht(modelo2,linfct=mcp(nitrato=contr),alternative="greater"))
      Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: User-defined Contrasts
Fit: aov(formula = nlechug ~ nitrato)
Linear Hypotheses:
              Estimate Std. Error t value      Pr(>t)
con nitrato vs. sin nitrato <= 0    38.250      8.339    4.587 0.000178 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Capítulo 5

Modelo ANOVA un solo factor efectos aleatorios: Experimentos para estudiar componentes de varianza

Tal como establecen [2], hay ocasiones en que el interés yace sobre una gran población de posibles niveles del factor de interés, y los niveles que son usados durante el experimento son una muestra aleatoria de esta población (generalmente una pequeña proporción). En este tipo de estudios interesa identificar las fuentes principales de variación en un sistema y estimar sus contribuciones a la varianza de la respuesta.

Debido a los objetivos perseguidos, la estructura de tratamientos, los protocolos experimentales y el tipo de inferencias realizadas, se tiene que los efectos de los niveles del factor en este modelo son considerados efectos aleatorios, es decir, variables aleatorias cuyas distribuciones son las distribuciones de valores en la población de niveles. El correspondiente modelo es llamado Modelo de Efectos Aleatorios. No estamos interesados sólo en los niveles que se dieron en la experimentación; en cambio, estamos interesados en la variabilidad de los efectos de todos los niveles en la población. Por lo tanto, los efectos aleatorios deben ser tratados de forma diferente a como lo hicimos con los efectos fijos [2, 6]. ***Se asume que el número total de posibles niveles del factor es al menos 100 veces más grande que el número de niveles que pueden ser observados en la experimentación.*** Por lo general la

población de posibles niveles cumplirá este requerimiento, y para propósitos de especificación del modelo, se asume que la población es infinita; de lo contrario, alguna corrección por población finita deberá realizarse en todas las fórmulas, y esto va más allá del alcance de este curso. Considere los siguientes ejemplos tomados de [2]:

Ejemplo 1: Suponga que un fabricante de sopa de tomate enlatada desea reducir la variabilidad en el espesura de la sopa. Suponga que las causas más probables de la variabilidad son la calidad de la harina de maíz recibida del proveedor y las acciones de los operadores de las máquinas. Considere dos escenarios diferentes:

Escenario 1: Los operadores de las máquinas son altamente cualificados y han estado en la empresa por largo tiempo. Así, la causa más probable de variabilidad es la calidad de la harina entregada a la compañía. El factor de tratamiento es “la harina de maíz”, y sus posibles niveles son todos los posibles lotes de harina que el proveedor podría entregar. Teóricamente, esta es una población infinita de lotes. Estamos interesados no sólo en los lotes de harina que hayan sido entregadas actualmente, sino también en todos aquellos que serían despachados en el futuro. Si asumimos que los lotes entregados a la empresa son una muestra aleatoria de todos los lotes que podrían ser entregados, y si tomamos una muestra aleatoria de los lotes entregados para ser observados en el experimento, entonces el efecto de la harina sobre la espesura es un efecto aleatorio y puede ser modelado por una variable aleatoria.

Escenario 2: Sabemos que la calidad de la harina es extremadamente consistente, así que la causa más probable de variabilidad es debida a las acciones diferentes de los operadores de las máquinas. La empresa es grande y los operarios cambian muy frecuentemente. Consecuentemente, aquellos que están disponibles para tomar parte en el experimento son sólo una pequeña muestra de todos los operarios empleados por la empresa en el presente o que podrían ser empleados en el futuro. Si asumimos que los operarios disponibles para el experimento son representativos de la población, entonces podemos asumir que ellos son similares a una muestra aleatoria de una población muy grande de posibles operarios, presentes y futuros. Dado que nos gustaría conocer la variabilidad de toda la población, modelamos los efectos de

los operarios como variables aleatorias, y los llamamos efectos aleatorios.

En ausencia de cualesquier factores de bloqueo, puede usarse un diseño completamente aleatorizado. Los niveles del factor son seleccionados primero aleatoriamente de la población de todos los posibles niveles, y luego las unidades experimentales son asignadas aleatoriamente a aquellos niveles seleccionados, como en el caso de efectos fijos. ***Como regla general, si la variabilidad de los efectos de tratamientos es mucho mayor que la variabilidad del error, entonces el número de niveles a debe ser mayor y el número de réplicas n pequeño, y viceversa.***

Ejemplo 2: Experimento del contenido de lana limpia: Este experimento fue reportado por Cameron, J. M. (1951), *Biometrics*, vol. 7(1): pp. 83-96.

a) Objetivos del experimento: La lana virgen contiene cantidades variables de grasa, suciedad y material extraño los cuales deben ser removidos antes de la manufactura. El precio de compra y los aranceles de un embarque están basados en la cantidad real de lana presente, es decir, en la cantidad de lana presente después de una limpieza minuciosa y denominada como “el contenido limpio”. El contenido limpio es expresado como el porcentaje del peso de la lana limpia respecto al peso original de la lana virgen. El experimento fue corrido con el fin de estimar la variabilidad en “el contenido limpio” de los fardos de lana en un embarque.

b) Identificación de todas las fuentes de variación:

i) Factores de tratamiento y sus niveles: El factor de tratamiento fue el fardo de lana y sus niveles fueron toda la población de fardos en un embarque particular. Siete fardos fueron observados en el experimento, y estos fueron seleccionados aleatoriamente del embarque. El embarque era suficientemente grande para permitir que los fardos usados en el experimento sean considerados como una muestra aleatoria de una población infinita de fardos. El factor fardo de lana fue por tanto considerado como de efectos aleatorios.

ii) Unidades experimentales: Las unidades experimentales fueron los tramos o intervalos de tiempo entre mediciones, de manera que la asignación de estos a

los niveles seleccionados del factor determinaron el orden en el cual los fardos de lana fueron observados.

III) Factores de bloqueo, de ruido y covariables: No se identificaron ninguno como fuentes principales de variación.

- c) Elección de una regla para asignar las unidades experimentales a los niveles del factor: Un diseño completamente aleatorizado fue seleccionado
- d) Especificación de la medición a ser hecha, el procedimiento experimental y anticipación de dificultades: Una máquina fue usada para perforar de un lado a otro un fardo de lana y extraer un núcleo de lana. Varios núcleos fueron tomados de cada uno de los siete fardos seleccionados de modo que varias observaciones sobre “el contenido limpio” pudieron hacerse de cada fardo. Cada núcleo de lana fue pesado y luego limpiado por restriegue, etc. Después de limpiar, la lana fue repesada y el contenido limpio fue calculado como la razón del peso de la lana limpia al peso inicial multiplicado por 100 %. Una dificultad anticipada fue que el proceso de restregado, el cual funciona bien con grandes cantidades de lana, dio dificultad con cantidades pequeñas de lana, de manera que el error experimental observado en el experimento puede ser mucho más grande de lo que sería normalmente observado en la producción de rutina.

Las observaciones recolectadas en este experimento se presentan a continuación:

Contenido limpio (%)						
Fardo						
1	2	3	4	5	6	7
52.33	56.99	54.64	54.90	59.89	57.76	60.27
56.26	58.69	57.48	60.08	57.76	59.68	60.30
62.86	58.20	59.29	58.72	60.26	59.58	61.09
50.46	57.35	57.51	55.61	57.53	58.08	61.45

5.1. Modelo de un sólo factor de efectos aleatorios

Para un diseño completamente aleatorizado, con a niveles seleccionados aleatoriamente de un factor A, el modelo es:

$$Y_{ij} = \mu + A_i + \varepsilon_{ij}, \text{ con} \quad (5.1)$$

$$\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2), A_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$$

los ε_{ij} y A_i son mutuamente independientes, $i = 1, 2, \dots, a$ y $j = 1, 2, \dots, n_i$. Este modelo es representado en la Figura 5.1.

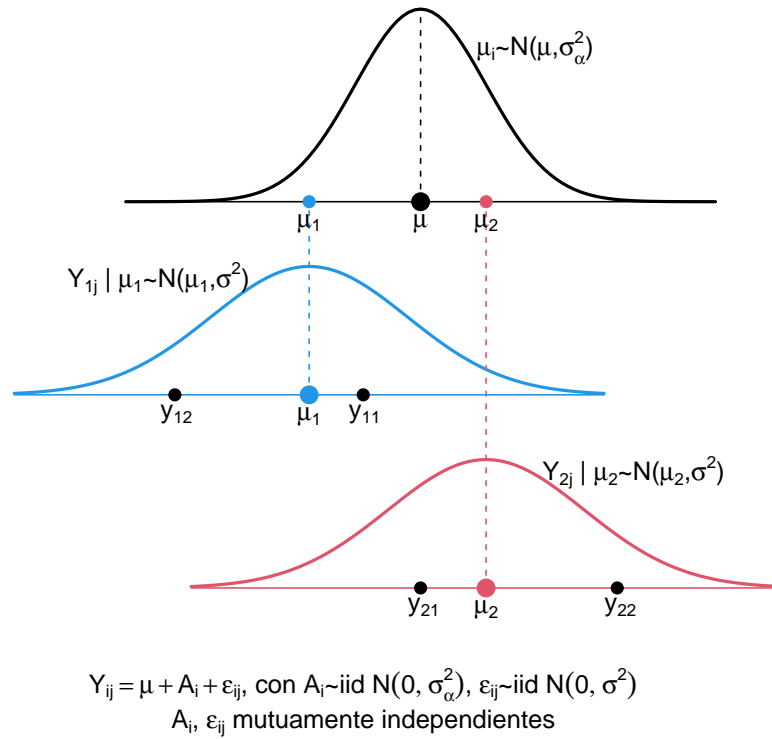


Figura 5.1: Representación del modelo ANOVA de un factor de efectos aleatorios. Adaptado de [7]. Note que las $\mu_i = \mu + A_i$, por tanto, $\mu_i \stackrel{iid}{\sim} N(\mu, \sigma_\alpha^2)$.

Dado que los niveles observados del factor han sido seleccionados aleatoriamente de una población “infinita”, los efectos observados son observaciones de variables aleatorias A_i , respectivamente. Se asume que las A_i son normales con varianza σ_α^2 , y

esta presunción necesitará ser chequeada junto con los supuestos sobre los errores. La media de la población de efectos de tratamiento ha sido absorbida en la constante μ , de modo que la distribución de A_i es normal de media cero y varianza σ_α^2 . Este último es el parámetro de interés, dado que si los efectos de todos los niveles del factor son iguales entonces σ_α^2 es cero. Si los efectos son muy diferentes, entonces σ_α^2 es muy grande.

Respecto al supuesto de independencia entre los A_i , este resulta razonable en tanto que los niveles del factor sean seleccionados aleatoriamente. Sin embargo, en el ejemplo 1 escenario 2, los niveles son una muestra “conveniente”, luego esta presunción debería investigarse cuidadosamente. La independencia entre los A_i y los ε_{ij} requiere que el factor no afecte ninguna fuente de variación que haya sido absorbida en la variable de error.

5.2. Distribución de las variables de respuesta

Bajo este modelo se cumple que

- El valor esperado de la respuesta es μ :

$$E[Y_{ij}] = E[\mu + A_i + \varepsilon_{ij}] = \mu \quad (5.2)$$

- La varianza de la respuesta es $\sigma_\alpha^2 + \sigma^2$:

$$V[Y_{ij}] = V[\mu + A_i + \varepsilon_{ij}] = V[A_i] + V[\varepsilon_{ij}] + 2 \times COV(A_i, \varepsilon_{ij}) = \sigma_\alpha^2 + \sigma^2 \quad (5.3)$$

- La distribución de Y_{ij} es normal (recuerde que la suma de variables normales independientes es una variable con distribución normal):

$$Y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2) \quad (5.4)$$

- Las observaciones sobre el mismo nivel o tratamiento están correlacionadas:

$$COV(Y_{ij}, Y_{is}) = COV(\mu + A_i + \varepsilon_{ij}, \mu + A_i + \varepsilon_{is}) = V[A_i] = \sigma_\alpha^2 \quad (5.5)$$

5.3. Análisis de varianza

La tabla ANOVA a construir es similar a la que se construye para el caso de efectos fijos, la diferencia yace en la interpretación de las fuentes de variación, los valores esperados y de las hipótesis asociadas.

Tabla 5.1: Tabla ANOVA en un DCA con un sólo factor de efectos aleatorios

Fuente de variación	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F ₀	Valor P
Factor	$a - 1$	SSA	$MSA = \frac{SSA}{a-1}$	$\sigma^2 + c\sigma_\alpha^2$	$\frac{MSA}{MSE}$	$P\left(f_{a-1, N-a} > F_0\right)$
Error	$N - a$	SSE	$MSE = \frac{SSE}{N-a}$	σ^2		
Total	$N - 1$	SST				
$SSA = \sum_i n_i \bar{Y}_{i\bullet}^2 - N\bar{Y}_{\bullet\bullet}^2$; $SST = \sum_i \sum_j Y_{ij}^2 - N\bar{Y}_{\bullet\bullet}^2$; $SSE = SST - SSA$; $c = \frac{1}{N(a-1)} \left[N^2 - \sum_{i=1}^a n_i^2 \right]$						

Nota 5.1. Si el diseño es balanceado, $N = na$ y $c = n$ en la expresión del valor esperado del MSA.

La hipótesis fundamental cuando el factor es de efectos aleatorios versa sobre la variabilidad de estos efectos en la población total de niveles, no sobre las medias de aquellos niveles observados en el experimento. Dado que la varianza de los efectos en la población es σ_α^2 , el test de hipótesis de interés asociado la Tabla ANOVA es:

$$H_0 : \sigma_\alpha^2 = 0, \text{ vs. } H_1 : \sigma_\alpha^2 > 0 \quad (5.6)$$

El estadístico de prueba es $F_0 = MSA/MSE$, veamos que aún bajo este modelo de efectos aleatorios, se distribuye como una variable aleatoria F:

Puede mostrarse que $SSA/(\sigma^2 + c\sigma_\alpha^2) \sim \chi_{a-1}^2$ y que $SSE/\sigma^2 \sim \chi_{N-a}^2$ y que ambas sumas de cuadrados son mutuamente independientes, por tanto:

$$\frac{MSA/(\sigma^2 + c\sigma_\alpha^2)}{MSE/\sigma^2} \sim \frac{\chi_{a-1}^2/(a-1)}{\chi_{N-a}^2/(N-a)} \sim f_{a-1, N-a}, \quad (5.7)$$

luego, cuando H_0 es cierta, $MSA/MSE \sim f_{a-1, N-a}$.

En este tipo de experimentos puede ser más interesante probar si la varianza debida a los efectos aleatorios es menor o igual que alguna proporción de la varianza del error [2], es decir:

$$H_0 : \sigma_\alpha^2 \leq \eta\sigma^2 \text{ vs. } H_1 : \sigma_\alpha^2 > \eta\sigma^2 \quad (5.8)$$

para alguna constante positiva η . Si H_0 es cierta, entonces:

$$\frac{MSA/(\sigma^2 + c\sigma_\alpha^2)}{MSE/\sigma^2} = \frac{MSA}{(1 + c\eta)MSE} \sim f_{a-1, N-a}. \quad (5.9)$$

Por tanto, la regla de decisión necesita una pequeña modificación para incluir la constante $(1 + c\eta)$, así: Rechazar H_0 si $\frac{MSA}{MSE} > (1 + c\eta) f_{\gamma, a-1, N-a}$, a un nivel de significancia γ dado.

5.4. Estimaciones

5.4.1. Estimaciones mediante el método ANOVA

De acuerdo a [12], este método de estimación consiste en encontrar los valores esperados de las sumas de cuadrados SSA y SSE (o de sus cuadrados medios MSA, MSE) a partir de sus respectivas definiciones y luego igualar estos valores esperados con los observados para luego despejar los estimadores. De esta forma aparecen ecuaciones que son lineales en las componentes de varianza desconocidas. Ver en la Tabla 5.2 el resumen de los estimadores e intervalos de confianza para los parámetros que interesan en este tipo de modelos.

Deducción de los estimadores puntuales

■ Desde que $E[MSE] = \sigma^2$, entonces tomamos $\hat{\sigma}^2 = MSE$.

■ Teniendo en cuenta que $E[MSA] = \sigma^2 + c\sigma_\alpha^2$, entonces

$$\sigma_\alpha^2 = \frac{E[MSA] - \sigma^2}{c} = \frac{E[MSA] - E[MSE]}{c},$$

por tanto, un estimador insesgado para σ_α^2 es $\hat{\sigma}_\alpha^2 = \frac{MSA - MSE}{c}$.

■ Siguiendo la fórmula de las proporciones de varianzas indicadas en la Tabla 5.2, se definen los estimadores para

1. **La razón de componentes de varianza:** $\frac{\sigma_\alpha^2}{\sigma^2}$

2. **La proporción de la varianza total debida al factor:** $\frac{\sigma_\alpha^2}{\sigma^2 + \sigma_\alpha^2}$,

simplemente sustituyendo cada varianza por su respectivo estimador puntual en la proporción de varianzas.

■ Dado que $E[\bar{Y}_{\bullet\bullet}] = \mu$, usamos $\hat{\mu} = \bar{Y}_{\bullet\bullet}$.

Tabla 5.2: Estimadores puntuales y por intervalos de confianza de $(1 - \gamma) 100 \%$,

Parámetro	Estimador puntual	Intervalo de confianza
σ^2	$\hat{\sigma}^2 = MSE$	$\frac{SSE}{\chi_{\gamma/2, N-a}^2} < \sigma^2 < \frac{SSE}{\chi_{1-\gamma/2, N-a}^2}$
σ_α^2	$\hat{\sigma}_\alpha^2 = \frac{MSA - MSE}{c}$	$\frac{\kappa \hat{\sigma}_\alpha^2}{\chi_{\gamma/2, \kappa}^2} < \sigma_\alpha^2 < \frac{\kappa \hat{\sigma}_\alpha^2}{\chi_{1-\gamma/2, \kappa}^2} (*)$ <p>con $\kappa = \frac{(MSA - MSE)^2}{\frac{MSA^2}{a-1} + \frac{MSE^2}{N-a}}$; o bien,</p> $\frac{(F_0 - f_{\gamma/2, a-1, N-a})}{cF_0\chi_{\gamma/2, a-1}^2} SSA < \sigma_\alpha^2 < \frac{(F_0 - f_{1-\gamma/2, a-1, N-a})}{cF_0\chi_{1-\gamma/2, a-1}^2} SSA (**)$
$\frac{\sigma_\alpha^2}{\sigma^2}$	$\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}^2} = \frac{MSA - MSE}{c \times MSE}$	$\frac{1}{c} \left[\frac{F_0}{f_{\gamma/2, a-1, N-a}} - 1 \right] < \frac{\sigma_\alpha^2}{\sigma^2} < \frac{1}{c} \left[\frac{F_0}{f_{1-\gamma/2, a-1, N-a}} - 1 \right]$
$\frac{\sigma_\alpha^2}{\sigma^2 + \sigma_\alpha^2}$	$\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}^2 + \hat{\sigma}_\alpha^2} = \frac{MSA - MSE}{(c-1)MSE + MSA}$	$\frac{L}{1+L} < \frac{\sigma_\alpha^2}{\sigma^2 + \sigma_\alpha^2} < \frac{U}{1+U},$ <p>donde $L = \frac{1}{c} \left(\frac{F_0}{f_{\gamma/2, a-1, N-a}} - 1 \right)$ y $U = \frac{1}{c} \left(\frac{F_0}{f_{1-\gamma/2, a-1, N-a}} - 1 \right)$</p>
μ	$\hat{\mu} = \bar{Y}_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}$	$\bar{Y}_{\bullet\bullet} \pm t_{\gamma/2, \nu^*} \times S_{\bar{Y}_{\bullet\bullet}}$ <p>con $S_{\bar{Y}_{\bullet\bullet}}^2 = \left(\sum_{i=1}^a \frac{n_i^2}{cN^2} \right) MSA + \left(\frac{1}{N} - \sum_{i=1}^a \frac{n_i^2}{cN^2} \right) MSE,$</p> $\nu^* = \frac{(a_1 MSA + a_2 MSE)^2}{\frac{(a_1 MSA)^2}{a-1} + \frac{(a_2 MSE)^2}{N-a}}, \text{ donde } a_1 = \sum_{i=1}^a \frac{n_i^2}{cN^2}, a_2 = \frac{1}{N} - \sum_{i=1}^a \frac{n_i^2}{cN^2}$
$F_0 = MSA/MSE, c = \frac{1}{N(a-1)} \left[N^2 - \sum_{i=1}^a n_i^2 \right]$ y para diseños balanceados $c = n$ simplemente. Para el I.C sobre μ , si el diseño es balanceado, $a_1 = (an)^{-1}$, $a_2 = 0$ y por tanto, los grados de libertad $\nu^* = a - 1$		

Para la deducción de los intervalos de confianza puede consultar Apéndice A.7. Las ecuaciones (*) y (**) para los I.C de la varianza σ_α^2 , corresponden a aproximaciones

pues no existe una forma cerrada para deducir un intervalo de confianza exacto para un estimador que es una combinación lineal de sumas de cuadrados (independientes). Las ecuaciones para los grados de libertad denotadas en la Tabla 5.2 como κ y ν^* corresponden a aproximaciones tipo Satterthwaite (Ver Apéndice A.8).

Acciones cuando la estimación de σ_α^2 es negativa

No hay nada que sea inherente al método de estimación ANOVA que evite una estimación negativa para σ_α^2 . De acuerdo a [6], hay que tener precaución con las estimaciones de la componente de varianza σ_α^2 . Puede suceder que dicha estimación resulte negativa como consecuencia de un $MSA < MSE$. Por definición una componente de varianza es positiva, pero si esto no sucede, existen varios cursos de acción sugeridos:

1. Aceptar tal estimación como evidencia de un valor verdadero igual a cero (en especial si es muy grande y negativo) y usar por tanto el valor de cero como valor estimado, reconociendo que el estimador será sesgado. Si esta solución es aplicada, tener en cuenta que el modelo debería replantearse así

$$Y_{ij} = \mu + \varepsilon_{ij}, \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2),$$

y en este caso, el estimador de la varianza del error sería $\hat{\sigma}^2 = MST$.

2. Conservar la estimación negativa, tomando en cuenta que es posible que los cálculos subsecuentes con tal valor no tendrán mucho sentido.
3. Interpretar la estimación negativa de la componente de varianza como una indicación de un modelo estadístico incorrecto, en especial si el MSE es considerablemente mayor que el MSA.
4. Utilizar un método diferente de análisis de varianza para estimar sus componentes. Ver: Searle S. R (1971, Capítulo 9). *Linear models*. New York, Wiley; Searle, S. R., Casella, G. and McCulloch, C. E. (1992, Capítulo 4). *Variance components*. John Wiley & Sons, Inc. Entre estos métodos se encuentran máxima

verosimilitud (estimadores MLE) y máxima verosimilitud restringida o residual (estimadores REML).

5. Reunir más datos y analizarlos por separado o junto con los existentes y esperar que el aumento de información conduzca a estimaciones positivas.

Según [12], en experimentos balanceados, en general, no parece necesario preocuparse con la posibilidad de tener un estimador $\hat{\sigma}_a^2$ negativo desde que el número de niveles o tratamientos no sea muy pequeño, además, tener muchos niveles es más importante que tener muchas observaciones por nivel. Parafraseando a [12], *“Esto es lo que uno esperaría: numerosas clases son necesarias si uno quiere estimar la componente de varianza de clases con cualquier grado de optimalidad. Trescientas observaciones en cada una de dos clases solo dan información sobre dos clases; es mejor información sobre esas clases que cinco observaciones sobre cada una de ellas, pero siguen siendo sólo dos clases”*.

5.4.2. Estimaciones por máxima verosimilitud (MLE) y máxima verosimilitud restringida o residual (REML)

De acuerdo a [12], en estos métodos se excluye explícitamente la posibilidad de estimaciones negativas. En la estimación de máxima verosimilitud se busca el vector de parámetros (incluyendo a la media global y las componentes de varianza) que maximizan la verosimilitud del vector completo de las N respuestas, en el espacio parametral ($-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$, $0 \leq \sigma_\alpha^2 < \infty$), donde la verosimilitud es la función de densidad de probabilidad conjunta de la Y_{ij} pero vista como función de los parámetros desconocidos. Por facilidad matemática, se trabaja sobre el log de la verosimilitud, que para el ANOVA un factor de efectos aleatorios en un DCA corresponde a,

$$l(\mu, \Sigma|Y) = -\frac{1}{2}N \log(2\pi) - \frac{1}{2}(N-a) \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^a \log(\sigma^2 + n_i \sigma_\alpha^2) - \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^a \frac{\sigma_\alpha^2}{\sigma^2 + n_i \sigma_\alpha^2} (Y_{i\bullet} - n_i \mu)^2, \quad (5.10)$$

donde Y es el vector de respuestas, Σ es la matriz de varianzas covarianzas del vector de respuestas y $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$, el total de las observaciones en el tratamiento i -ésimo.

Cuando el diseño es balanceado la ecuación (5.10), simplifica a

$$l(\mu, \Sigma|Y) = -\frac{1}{2}N \log(2\pi) - \frac{1}{2}a(n-1) \log(\sigma^2) - \frac{1}{2}a \log(\lambda) - \frac{\text{SSE}}{2\sigma^2} - \frac{\text{SSA}}{2\lambda} - \frac{an(\bar{Y}_{\bullet\bullet} - \mu)^2}{2\lambda} \quad (5.11)$$

donde $\lambda = \sigma^2 + n\sigma_\alpha^2$. En [12] puede verse que los estimadores de máxima verosimilitud son como sigue,

- En el caso balanceado,

$$\tilde{\mu} = \bar{Y}_{\bullet\bullet} \quad (5.12)$$

$$\tilde{\sigma}^2 = \begin{cases} \text{MSE}, & \text{si } \left(1 - \frac{1}{a}\right) \text{MSA} \geq \text{MSE} \\ \frac{\text{SST}}{an}, & \text{si } \left(1 - \frac{1}{a}\right) \text{MSA} < \text{MSE} \end{cases} \quad (5.13)$$

$$\tilde{\sigma}_\alpha^2 = \begin{cases} \frac{(a-1)\text{MSA} - a \cdot \text{MSE}}{an}, & \text{si } \left(1 - \frac{1}{a}\right) \text{MSA} \geq \text{MSE} \\ 0, & \text{si } \left(1 - \frac{1}{a}\right) \text{MSA} < \text{MSE} \end{cases} \quad (5.14)$$

Los valores esperados de los estimadores $\tilde{\sigma}^2$, $\tilde{\sigma}_\alpha^2$, no son funciones fácilmente tratables (ver [12]).

- En el caso desbalanceado, sea $\dot{\sigma}_\alpha^2$ el valor que toma σ_α^2 en la solución del llamado sistema de ecuaciones ML (derivadas del log verosimilitud con respecto a los parámetros del modelo, igualadas a cero). Si $\dot{\sigma}_\alpha^2$ es positivo, los estimadores de máxima verosimilitud son mucho más complejos de obtener pues no hay solución analítica, se requieren métodos iterativos para la solución de sistemas de ecuaciones no lineales, y se destaca el hecho de que la media muestral global no es el estimador MLE de μ bajo este escenario. En cambio, cuando $\dot{\sigma}_\alpha^2$ es negativo, los estimadores MLE son como sigue

$$\tilde{\mu} = \bar{Y}_{\bullet\bullet} \quad (5.15)$$

$$\tilde{\sigma}^2 = \frac{\text{SST}}{N} \quad (5.16)$$

$$\tilde{\sigma}_\alpha^2 = 0. \quad (5.17)$$

Ahora veamos un poco de los estimadores REML. De acuerdo a [12] y [2], los estimadores REML de las componentes de varianza son una adaptación de la máxima verosimilitud, en donde solo se maximiza la parte que es invariante a localización, lo que en términos del modelo ANOVA implica que se maximiza la parte de la verosimilitud que no involucra a μ . Uno de sus méritos es que es un método de máxima verosimilitud, que a pesar de no preocuparse de la estimación de los efectos fijos, tiene en cuenta los grados de libertad asociados a estos. Este método también puede considerarse como un método de estimación de verosimilitud marginal. La idea de este procedimiento es como sigue. Considere inicialmente el caso balanceado, es decir cuando $n_i = n$ para todo $i = 1, \dots, a$. Por manipulación algebraica puede mostrarse que la verosimilitud del vector de observaciones \mathbf{Y} factoriza de la siguiente manera:

$$L(\mu, \sigma^2, \sigma_\alpha^2 | \mathbf{Y}) = \underbrace{\frac{\exp \left[-\frac{(\bar{Y}_{\bullet\bullet} - \mu)^2}{2\lambda/(an)} \right]}{\sqrt{2\pi\lambda/(an)}}}_{L(\mu | \bar{Y}_{\bullet\bullet})} \times \underbrace{\frac{\exp \left[-\frac{1}{2} \left(\frac{\text{SSE}}{\sigma^2} + \frac{\text{SSA}}{\lambda} \right) \right]}{(2\pi)^{(an-1)/2} (\sigma^2)^{[a(n-1)/2]} \lambda^{(a-1)/2} (an)^{1/2}}}_{L(\sigma^2, \sigma_\alpha^2 | \text{SSE}, \text{SSA})} \quad (5.18)$$

es decir, $L(\mu, \sigma^2, \sigma_\alpha^2) = L(\mu | \bar{Y}_{\bullet\bullet}) \times L(\sigma^2, \sigma_\alpha^2 | \text{SSE}, \text{SSA})$,

con $\lambda = \sigma^2 + n\sigma_\alpha^2$, donde los factores $L(\mu | \bar{Y}_{\bullet\bullet})$ y $L(\sigma^2, \sigma_\alpha^2 | \text{SSE}, \text{SSA})$ son independientes, ya que puede demostrarse que las sumas de cuadrados SSA, SSE son independientes de $\bar{Y}_{\bullet\bullet}$ el promedio muestral global. El método toma entonces como verosimilitud residual o restringida a $L(\sigma^2, \sigma_\alpha^2 | \text{SSE}, \text{SSA})$ y los estimadores REML para σ^2 y σ_α^2 son tales que maximizan a esta verosimilitud (o equivalentemente a su logaritmo natural) en el espacio parametral $\sigma^2 > 0$, $\sigma_\alpha^2 \geq 0$, siendo estos estimadores,

$$\tilde{\sigma}_R^2 = \begin{cases} \text{MSE} & \text{si } \frac{1}{n} (\text{MSA} - \text{MSE}) \geq 0 \\ \frac{\text{SST}}{an-1} & \text{si } \frac{1}{n} (\text{MSA} - \text{MSE}) < 0 \end{cases} \quad (5.19)$$

$$\tilde{\sigma}_{\alpha,R}^2 = \begin{cases} \frac{1}{n}(\text{MSA} - \text{MSE}) & \text{si } \frac{1}{n}(\text{MSA} - \text{MSE}) \geq 0 \\ 0 & \text{si } \frac{1}{n}(\text{MSA} - \text{MSE}) < 0. \end{cases} \quad (5.20)$$

Note que en el caso balanceado, cuando $\frac{1}{n}(\text{MSA} - \text{MSE}) \geq 0$, los estimadores REML coinciden con los estimadores ANOVA. Ahora bien, comparados con los estimadores MLE en (5.13) y (5.14), vemos que los estimadores REML toman en cuenta los grados de libertad, del SSA y SST y en esa forma tiene en cuenta los grados de libertad usados en la estimación de μ , aún cuando este método no involucra explícitamente la estimación de μ .

En el caso desbalanceado la verosimilitud correspondiendo a

$$L(\mu, \sigma^2, \sigma_\alpha^2 | \mathbf{Y}) = \frac{\exp \left\{ - \left[\frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2}{2\sigma^2} - \sum_{i=1}^a \frac{n_i^2 \sigma_\alpha^2 (\bar{Y}_{i\bullet} - \mu)^2}{2\sigma^2 (\sigma^2 + n_i \sigma_\alpha^2)} \right] \right\}}{(2\pi)^{N/2} (\sigma^2)^{(N-a)/2} \prod_{i=1}^a (\sigma^2 + n_i \sigma_\alpha^2)^{1/2}}, \quad (5.21)$$

sin embargo, no es directa la factorización de esta verosimilitud en dos factores que permita separar una función de μ como en el caso balanceado. De acuerdo a [12], una construcción más elaborada de esta factorización es necesaria y es derivada del método REML general, en el cual, en lugar de usar directamente la verosimilitud del vector de respuestas \mathbf{Y} , se trabaja sobre combinaciones lineales de este vector elegidas de forma que no contengan los efectos fijos que hay en el modelo. Estas combinaciones lineales son equivalentes a los residuos obtenidos después de ajustar por mínimos cuadrados solo la parte de efectos fijos del modelo. Más detalles sobre el procedimiento, en el Capítulo 6 de [12].

Con relación a las propiedades distribucionales de los estimadores MLE y REML y los cálculos de intervalos de confianza para las componentes de varianza, las proporciones de varianza y la razón de varianzas, están basadas en propiedades de normalidad asintótica del vector de parámetros estimados, tema fuera del alcance de este curso. En R los estimadores REML pueden ser obtenidos a través de la función `lmer` de la librería `lme4`, o con la función del mismo nombre pero en la librería `lmerTest`,

en donde es posible tener valores P para la prueba de significancia de efectos fijos en el caso de experimentos factoriales de efectos mixtos.

5.5. Chequeo de supuestos

De acuerdo a [2], la manera más simple de chequear los supuestos en un modelo de un factor de efectos aleatorios es usar los gráficos de residuales en forma similar a como se hizo en el caso de efectos fijos. Necesitamos chequear que $\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ y también que los supuestos sobre los efectos aleatorios son válidos, esto es $A_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ y que todas las variables son mutuamente independientes.

Chequear las presunciones sobre los errores se hace directamente, dado que se procede de la misma forma como en el caso del modelo de un factor efectos fijos. Reemplazamos las variables A_i en el modelo, temporalmente, por efectos fijos, luego los residuales son definidos como $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_{i\bullet}$. Con estos chequeamos la independencia entre los errores, la normalidad y la varianza constante mediante gráficos. Si bien la no independencia entre los errores y los efectos aleatorios A_i es difícil de detectar, las varianzas desiguales de los errores indican una forma del problema.

De acuerdo a [2], el supuesto de normalidad sobre los efectos aleatorios A_i puede ser chequeado cuando los tamaños de muestra son iguales, a menos que el número de niveles a sea muy pequeño. Los promedios de tratamientos $\bar{Y}_{i\bullet}$ deberían tener una distribución $N(\mu, \sigma_\alpha^2 + \sigma^2/n)$. Por tanto, el gráfico de los promedios observados versus sus correspondientes scores normales, debería aproximar a una línea recta que corta el eje vertical aproximadamente en μ y con pendiente aproximada de $\sqrt{\sigma_\alpha^2 + \sigma^2/n}$. Es importante chequear el supuesto de normalidad dado que el análisis para los modelos de efectos aleatorios no es robusto a la no normalidad de los efectos aleatorios. Este gráfico también puede usarse para chequear outliers entre los tratamientos observados. En un experimento con tan sólo $a = 3$ niveles no se tienen suficientes datos para examinar la distribución de los A_i .

5.6. Ejemplo

Analice los resultados referentes a los datos del experimento de la lana dados en el ejemplo 2 sobre el experimento del contenido de lana limpia. El programa R es como sigue.

Código R 5.6.1. Programa R para ajuste modelo Anova un factor de efectos aleatorios en un DCA y la obtención de estimaciones y análisis de supuestos. Observe el uso del función R `lmer()` de la librería `lme4`, la cual ajusta y da estimaciones de las componentes de varianza y de la media global, especificando la fórmula del modelo como `respuesta~1+(1|factor)`, o bien `respuesta~1|factor`, donde el número 1 indica que la media global será ajustada, mientras que la barra vertical | en el segundo término separa los efectos fijos de los efectos aleatorios, siendo los fijos los que se especifiquen a la izquierda de la barra y los aleatorios los que se escriban a su derecha. También observe cómo en la función `aov` se da la especificación del factor como de tipo aleatorio: `aov(cont1~Error(fardos))`.

```
library(lme4)
library(MASS)
#INGRESO DE LOS DATOS
datos=data.frame(fardos=factor(rep(1:7,times=4)),cont1=scan())
52.33 56.99 54.64 54.90 59.89 57.76 60.27
56.26 58.69 57.48 60.08 57.76 59.68 60.30
62.86 58.20 59.29 58.72 60.26 59.58 61.09
50.46 57.35 57.51 55.61 57.53 58.08 61.45

datos
attach(datos)
#boxplots por niveles del factor
medias.fardos=sapply(split(cont1,fardos),mean)
plot(cont1~fardos,boxwex=0.5)
lines(1:7,medias.fardos,col=2,lty=2,type="b",pch=19)
#MODELANDO EFECTOS ALEATORIOS PARA OBTENER COMPONENTES DE VARIANZA
diseño=lmer(cont1~1|fardos)

#LO SIGUIENTE NOS DA ALGUNAS MEDIDAS DE AJUSTE Y LAS ESTIMACIONES DE LAS COMPONENTES DE VARIANZA
```

```
#LO QUE APARECE EN LA SALIDA COMO INTERCEPTO ES SIMPLEMENTE EL VALOR DE LA MEDIA GLOBAL MUESTRAL
summary(diseño)

#OBTENIENDO LA ANOVA EN MODELO DE EFECTOS ALEATORIOS
#summary sobre el objeto diseño no nos da la ANOVA; ésta se obtiene como sigue
summary(aov(cont1~Error(fardos)))

#RESIDUALES ESTANDARIZADOS Y VALORES AJUSTADOS CALCULADOS COMO SI EL MODELO FUESE DE EFECTOS FIJOS
res.estudent=stdres(aov(cont1~fardos))
Yhat=fitted(aov(cont1~fardos))
#GRÁFICO DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS VS. VALORES AJUSTADOS
plot(Yhat,res.estudent,xlab="Respuesta ajustada",ylab="Residuales internamente estudentizados",
     pch=1:7,col=c(1:6,"brown"),cex=1.5)
abline(h=c(-2,0,2),lty=2)
legend("topright",legend=c(1:7),col=c(1:6,"brown"),pch=c(1:7))
#GRÁFICO DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS VS. NIVELES FACTOR FARDO
stripchart(res.estudent~fardos,xlab="fardos",ylab="Residuales internamente estudentizados",vertical=T,
           pch=1:7,col=c(1:6,"brown"),cex=1.5)
abline(h=c(-2,0,2),lty=2)
legend("topright",legend=c(1:7),col=c(1:6,"brown"),pch=c(1:7))
#PROBAR NORMALIDAD SOBRE LOS ERRORES
shapiro.test(res.estudent)
qqnorm(res.estand,main="Gráfico probabilidad normal sobre residuos internamente estudentizados")
qqline(res.estudent,lty=2)

#PROBAR NORMALIDAD SOBRE LOS EFECTOS ALEATORIOS
#Calcular medias de tratamientos estandarizadas
medias.fardos.estand=scale(medias.fardos)
medias.fardos.estand
qqnorm(medias.fardos.estand,main="Gráfico de probabilidad normal para los efectos")
qqline(medias.fardos.estand,lty=2,col=2)
abline(h=0,lty=2)
abline(v=0,lty=2)
detach(datos)
```

Salida R 5.6.1. *Resultados numéricos del programa R anterior. Los resultados de `summary(diseño)` nos muestra las estimaciones de las componentes de varianza σ_{α}^2 y σ^2 , y de la media global μ (como un intercepto constituyendo efectos fijos), mientras que con la instrucción `R summary(aov(cont1~Error(fardos)))` obtenemos la ANOVA pero sin la información del valor P para el test de significancia de la componente de*

varianza σ_{α}^2 .

```
> summary(diseño)
Linear mixed model fit by REML ['lmerMod']
Formula: cont1 ~ 1 | fardos
REML criterion at convergence: 132.9
Scaled residuals:
      Min       1Q   Median       3Q      Max
-2.58770 -0.28827  0.01657  0.54711  2.36811
Random effects:
 Groups   Name      Variance Std.Dev.
 fardos   (Intercept) 1.183    1.088
 Residual                6.261    2.502
Number of obs: 28, groups: fardos, 7
Fixed effects:
              Estimate Std. Error t value
(Intercept)  58.0364    0.6266   92.62

> summary(aov(cont1~Error(fardos)))
Error: fardos
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  6  65.96   10.99
Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals 21  131.5    6.261

> shapiro.test(res.estand)
      Shapiro-Wilk normality test
data:  res.estand
W = 0.91068, p-value = 0.02054
```

Primero debemos chequear los supuestos sobre los errores calculando y graficando los residuales suponiendo como si los efectos fueran fijos. El gráfico de residuales internamente estudentizados vs. valores ajustados muestra como característica más notable que el fardo 1 tiene el residual internamente estudentizado más grande (también es notable del gráfico boxplot que en dicho fardo se presenta mayor variabilidad comparado con los demás. Esto puede significar una de varias cosas: quizás el dato correspondiente a tal observación es errónea, así que este valor es un outlier, o quizás el fardo 1 es extremadamente más variable que los otros fardos en la población, o

quizás los errores no se distribuyen normales. Si el dato es un outlier y se puede establecer que sucedió algo inusual durante la experimentación, podemos eliminar tal dato y replantear todos los cálculos. Eliminando tal observación, los demás residuales internamente estudentizados yacen dentro del rango esperado para errores distribuidos normalmente. El gráfico de residuales también nos indica que el promedio de

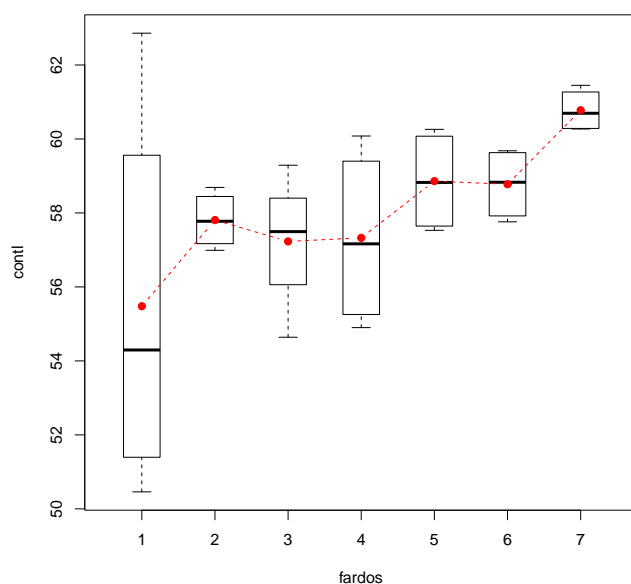


Figura 5.2: Boxplots para los datos del experimento del contenido de lana limpia

contenido limpio del fardo 1 está considerablemente más abajo al del resto de fardos y este es el mismo fardo que contiene el posible outlier. Podría sospecharse que este fardo no proviene del mismo embarque o fue contaminado en algún momento antes de ser medido. De otro lado, el embarque puede contener un número de “fardos raros” y esto debería ser investigado. En el otro extremo del rango, vemos que el fardo 7 tuvo el contenido limpio más alto y fue el menos variable. Quizás esto no es demasiado sorprendente, dado que un fardo con 100 % de contenido limpio probablemente no mostraría variabilidad en las mediciones tomadas sobre él. Así, uno puede sospechar que el modelo que incluye errores distribuidos normales no es ideal para esta situación. Sin embargo, la no normalidad no es severa.

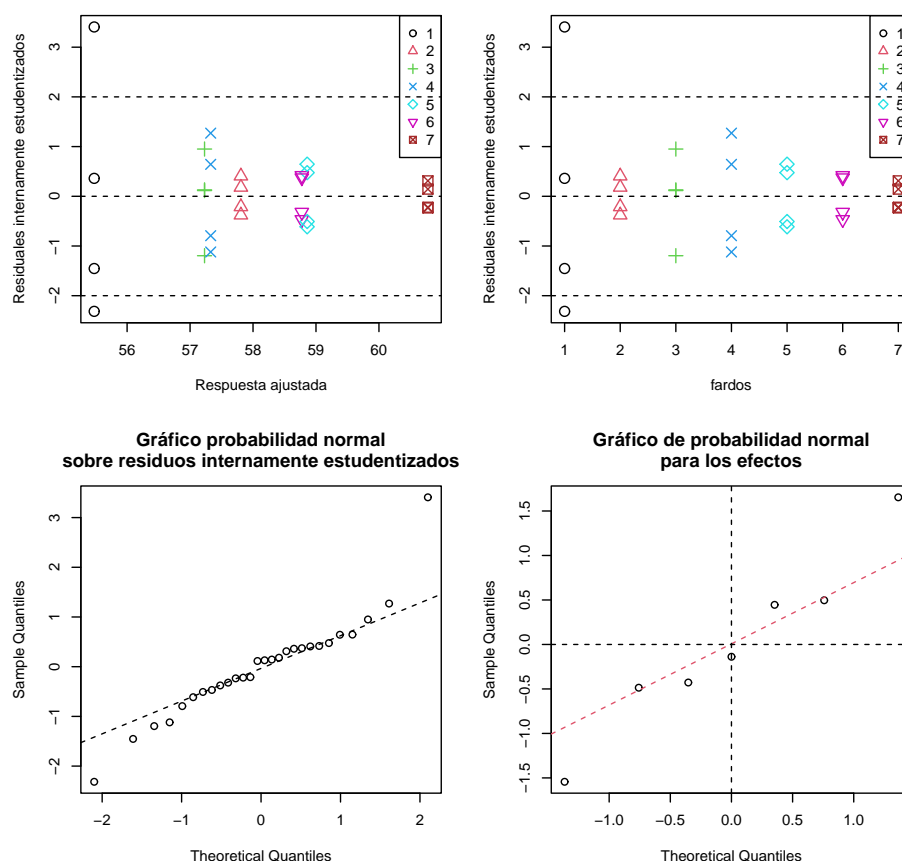


Figura 5.3: Residuos internamente estudentizados para el experimento del contenido de lana limpia. Gráficos de probabilidad normal también sobre residuos internamente estudentizados y efectos estimados (estandarizados).

Para chequear el supuesto de que los efectos aleatorios del modelo son normales, en este ejemplo cómo se cuenta con siete niveles, es posible evaluar el gráfico de probabilidad normal presentado para los promedios estandarizados de los tratamientos contra sus scores normales (esto no puede hacerse para modelos con más de un efecto aleatorio, dado que los promedios de tratamientos no son independientes). Según el gráfico no se observa evidencia en contra del supuesto de normalidad de la población de fardos. En la gráfica, los puntos pueden aproximadamente ser representados por una línea que pasa por el origen y de pendiente 1 (esto se espera para las medias estandarizadas de las observaciones por fardos.)

En resumen, el modelo de efectos aleatorios de un solo factor considerado para el ejemplo parece no ajustar bien a los datos, dado que las varianzas aparentemente no son constantes o hay un outlier. No obstante, se ha establecido que la población de fardos en este embarque es extremadamente variable, a pesar de que el ANOVA nos dice que la componente de varianza debido a los fardos no es significativa. Los fardos 1 y 7 parecen ser diferentes de los otros cinco fardos seleccionados. Quizás el embarque está constituido de subpoblaciones disímiles (de diferentes fuentes u orígenes). Esto debería chequearse, dado que puede dar una pista sobre cómo mejorar el contenido limpio de la lana en el futuro.

Capítulo 6

Potencia en ANOVA con un Factor en un DCA

De acuerdo con [2], los tamaños de muestra a usar en un experimento son críticos para la potencia sin embargo en la práctica, además de las consideraciones estadísticas en la determinación de los tamaños de muestra, aparecen otros aspectos importantes como son los costos tanto en tiempo y dinero, los cuales pueden diferir de nivel a nivel o de tratamiento a tratamiento. Probablemente habrá un presupuesto fijo para todo el experimento. En ese caso, hay que estimar el número máximo N de observaciones para la realización del experimento, descontando los costos de análisis, instrumentos, etc. Estas limitaciones deben llevarnos antes de ejecutar el experimento a reconsiderar si los objetivos previamente planteados en un estudio pueden ser alcanzados con los recursos limitados. Los cálculos de potencia nos dirán si con el número máximo de observaciones que se pueden obtener según el presupuesto disponible, es más que suficiente o si por el contrario nos quedaremos cortos para cumplir con todos los requerimientos de la experimentación, y en este caso debemos volver atrás en la planeación para intentar relajar algunos de los requerimientos u objetivos. De lo contrario, se requerirá un incremento del presupuesto.

Tiene poco sentido correr el experimento con tamaños de muestra más pequeñas que los requeridos sin haber estimado qué efecto tendrá esto sobre los análisis.

6.1. Potencia en experimentos con un factor de efectos fijos en un DCA

Consideraremos dos métodos para determinar los tamaños de muestra, estos son:

- Mediante la especificación de la longitud deseada de los intervalos de confianza,
- Mediante la especificación de la potencia requerida en el ANOVA.

6.1.1. Tamaños de muestra para alcanzar intervalos de confianza de longitudes especificadas

La longitud de un intervalo de confianza disminuye al incrementarse n . Consecuentemente, si la longitud de un intervalo de confianza es especificado, podría ser posible calcular los tamaños de muestra requeridos, específicamente cuando estos son iguales. Como se requiere conocer la varianza de los errores, es necesario una estimación preliminar bien sea a partir de un estudio piloto o bien, de un estudio anterior similar. Una aproximación de ensayo y error al problema puede ser aplicado para el cálculo de n . Veamos el siguiente ejemplo:

Ejemplo 1

Suponga que en un experimento se desea observar los efectos de un factor en $a = 5$ niveles y que la variable respuesta es medida en mm . Suponga también que un experimento piloto arroja que el MSE es a lo sumo igual a 10 mm^2 . Se desea construir intervalos de confianza simultáneos del 95 % para las diferencias de pares de medias, y que cada intervalo tenga una longitud no superior a 6 mm (es decir, el ancho medio o diferencia menos significativa no debería ser mayor que 3 mm). Si aplicamos Tukey para la obtención de tales intervalos, con tamaños iguales de muestra n , se está requiriendo que:

$$HSD_{ij} = \frac{1}{\sqrt{2}} q_{\gamma}(a, g.l) \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = \frac{1}{\sqrt{2}} q_{0.05}(5, 5(n-1)) \sqrt{10 \times \frac{2}{n}} = q_{0.05}(5, 5(n-1)) \sqrt{\frac{10}{n}} \leq 3$$

de donde se establece que debe cumplirse que $q_{0.05}^2(5, 5(n-1)) \leq 0.9n$.

Usando los valores críticos para el rango crítico estudentizado, los cuáles se pueden calcular con la función R: `qtukey(alpha, nmeans=a, df=gl, lower.tail=FALSE)`, se obtienen los resultados presentado en la Tabla 6.1. En la Figura 6.1 se ilustra el comportamiento de $q_{0.05}^2(5, 5(n-1)) - 0.9n$ vs. n .

Tabla 6.1: Evaluación de los tamaños de muestra satisfaciendo $q_{0.05}^2(5, 5(n-1)) \leq 0.9n$

n	$q_{0.05}^2(5, 5(n-1)) - 0.9n$	Acción
2	30.38	Incrementar n
3	18.96	Incrementar n
4	15.47	Incrementar n
5	13.41	Incrementar n
6	11.85	Incrementar n
7	10.53	Incrementar n
8	9.33	Incrementar n
9	8.21	Incrementar n
10	7.15	Incrementar n
11	6.12	Incrementar n
12	5.11	Incrementar n
13	4.12	Incrementar n
14	3.15	Incrementar n
15	2.18	Incrementar n
16	1.23	Incrementar n
17	0.28	Incrementar n
18	-0.66	No incrementar n

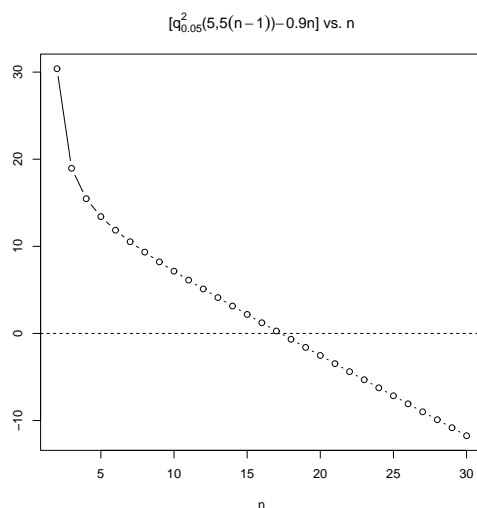


Figura 6.1: Comportamiento de la diferencia $q_{0.05}^2(5, 5(n-1)) - 0.9n$ vs. n , en la solución del tamaño de muestra óptimo según comparaciones de Tukey con un $HSD_{ij} \leq 3 \text{ mm}$, a un nivel de significancia de 0.05 y usando $\hat{\sigma}^2 = 10 \text{ mm}^2$ en un DCA de efectos fijos de un sólo factor con $a = 5$ niveles.

Si $n = 17$ observaciones son tomadas de cada nivel, y si el MSE es aproximadamente 10 mm^2 en el experimento principal, los intervalos de confianza simultáneos de Tukey del 95 % para las diferencias de pares de medias serán un poco más largos que el ancho requerido de 6 mm . Si $n = 18$ observaciones son tomadas de cada nivel, los intervalos serán un poco más cortos que los 6 mm requeridos. Si el costo del experimento es alto, entonces $n = 17$ debería seleccionarse; de lo contrario $n = 18$ debe ser preferido.

El método de ensayo y error presentado en el anterior ejemplo puede ser usado con cualquier otro método de comparaciones múltiples para obtener los tamaños de muestra requeridos para alcanzar los objetivos del experimento. El mismo tipo de cálculo puede hacerse para tamaños de muestra desiguales, si se especifican los tamaños relativos, por ejemplo $n_1 = 2n_2 = 2n_3 = 2n_4$.

A menos que se requiera más información en algunos tratamientos que en otros, o que hayan restricciones presupuestales de costo y/o tiempo, o varianzas desiguales, es generalmente preferible seleccionar tamaños de muestras iguales. Usar tamaños de muestras iguales es ventajoso porque los intervalos de confianza para comparaciones por pares son de la misma longitud, lo cual los hace más fáciles de comparar, y los procedimientos de comparación múltiple y los de análisis de varianza son menos sensibles a la violación de los supuestos de normalidad de los errores.

6.1.2. Tamaños de muestra usando la potencia de la prueba ANOVA

Siguiendo a [2], suponga que uno de los principales objetivos de un experimento es examinar si todos los tratamientos tienen efectos similares sobre la respuesta. La hipótesis nula asociada a esta pregunta es algo irrealista; los efectos de los tratamientos son ciertamente casi no exactamente iguales, y aún si lo fueran, la variabilidad debida a “ruido” en el experimento enmascara este hecho. En cualquier caso, si niveles diferentes producen sólo una muy pequeña diferencia en la variable respuesta, el investigador puede no estar interesado en descubrir tal hecho. Por ejemplo, una diferencia de cinco minutos de duración por peso (\$) entre dos baterías diferentes proba-

blemente no sería notada por muchos usuarios. Sin embargo una gran diferencia tal como 60 minutos de duración por peso (\$) puede ser bien notable. Así, el experimentador debe requerir que H_0 sea rechazada con alta probabilidad si $\mu_i - \mu_j > 60 \text{ min}/\$$ para algún $i \neq j$, pero puede no estar interesado en rechazar H_0 si $\mu_i - \mu_j \leq 5 \text{ min}/\$$ para todo $i \neq j$. En muchos experimentos hay algún valor Δ tal que si la diferencia en los efectos o en las medias de cualesquiera dos tratamientos excede a Δ , al experimentador le gustaría rechazar a H_0 en favor de H_1 con alta probabilidad.

Si en un experimento la componente de varianza debida al error experimental, σ^2 , es muy grande, es posible que el estadístico F para la respectiva prueba no sea lo suficientemente grande para rechazar H_0 , al nivel de significancia usado, a pesar de que en términos poblacionales, existan ciertas desviaciones de lo que se establece en H_0 . Se trata de un problema de potencia de la prueba, la cual está asociada a los tamaños muestrales. Para un nivel de significancia dado, *a mayor tamaño muestral, más sensible es la prueba a pequeñas desviaciones respecto a H_0 .*

Recuerde que potencia es:

- La probabilidad de rechazar H_0 cuando ésta es falsa, o bien, la probabilidad de rechazar una H_0 falsa.
- La probabilidad de detectar desviaciones respecto a H_0 .
- Potencia = $1 - \beta$, donde $\beta = P(\text{error tipo II})$
- La probabilidad de que el estadístico de la prueba pertenezca a la región de rechazo, cuando H_0 es falsa.

En términos prácticos, cuando no se rechaza la H_0 de un test de significancia sobre los parámetros de interés, eso no implica que dichos parámetros sean exactamente iguales a cero, sino que sus magnitudes (en términos absolutos) no son significativamente grandes para el nivel de significancia, y tamaños de muestra usados, y por tanto el aporte a la variabilidad total es pequeña frente al error experimental.

Basaremos los cálculos de potencia en el análisis de varianza en la siguiente expresión:

$$\text{Potencia} = P(F_0 > f_{\gamma, \nu_1, \nu_2} | H_0 \text{ es falsa}), \quad (6.1)$$

con

F_0	Estadístico de la prueba ANOVA
f_{γ, ν_1, ν_2}	Valor crítico para la prueba bajo la distribución f_{ν_1, ν_2}
γ	Nivel de significancia de la prueba.
ν_1	Grados de libertad del numerador
ν_2	Grados de libertad del denominador

Además, se supondrá que los diseños completamente aleatorizados son balanceados respecto al número de réplicas por nivel del factor, es decir $n_i = n$ para todo i , entonces $N = an$. Considere como test de hipótesis básico a probar que:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

$$H_1 : \text{algún } \alpha_i \neq 0, \quad i = 1, 2, \dots, a$$

con estadístico de prueba $F_0 = \frac{MSA}{MSE}$ y región crítica $F_0 > f_{\gamma, a-1, N-a}$ a un nivel de significancia γ especificado. Cuando H_0 es falsa se tiene que:

$$E(MSA) = \sigma^2 + n \frac{\sum_{i=1}^a \alpha_i^2}{a-1} \quad \text{con} \quad \sum_{i=1}^k \alpha_i^2 \neq 0 \quad (6.2)$$

Por tanto, F_0 ya no sigue la distribución f central que conocemos, sino una distribución f no central, con parámetro de no centralidad λ , dado por,

$$\lambda = \frac{n \sum_{i=1}^a \alpha_i^2}{\sigma^2}, \quad (6.3)$$

es decir bajo H_1 , $F_0 \sim f_{nc}(a-1, N-a, \lambda)$. Aplicando la expresión (6.1) y denotando la potencia por $\pi(\lambda)$ tenemos que:

$$\pi(\lambda) = P(f_{nc}(a-1, N-a, \lambda) > f_{\gamma, a-1, N-a}) \quad (6.4)$$

Nota 6.1. El parámetro de no centralidad λ en un DCA balanceado también puede

interpretarse de la siguiente manera

$$\lambda = n(a-1) \times \frac{\text{varianza entre grupos}}{\text{varianza dentro de grupos}} = n(a-1) \times \frac{\frac{1}{a-1} \sum_{i=1}^a \alpha_i^2}{\sigma^2} \quad (6.5)$$

Nota 6.2. Con respecto al test F sobre la significancia de un contraste $W = \sum_{i=1}^a c_i \alpha_i$, el parámetro de no centralidad en ese caso corresponde a

$$\lambda_W = \frac{n \left(\sum_{i=1}^a c_i \alpha_i \right)^2}{\sigma^2 \sum_{i=1}^a c_i^2} \quad (6.6)$$

Ejemplo 2

Considere el siguiente caso: Una compañía de alimentos deseaba probar 4 diseños de empaque para un nuevo cereal. 20 almacenes con aproximadamente igual volumen de ventas fueron seleccionados como las unidades experimentales. A cada almacén se le asignó aleatoriamente uno de los diseños de empaque, con cada diseño asignado a 5 almacenes. Se eligieron los almacenes de modo que resultaran comparables en ubicación, volumen de ventas, precios, cantidad y ubicación de espacios para estantes y esfuerzos promocionales especiales. Se observó las ventas durante el período de estudio y los resultados del análisis de varianza, y estimaciones de los efectos se presentan a continuación:

Fuente	df	Suma cuadrados	Cuadrados medios	F_0	$P(f_{3,16} > F_0)$
empaque	3	586.8	195.6	19.56	< 0.0001
error	16	160.0	10		
total	19	746.8			

Parámetro	Estimación	Error estándar	T_0	$P(t_{16} > T_0)$
α_1	-4.0	1.22475	-3.27	0.0049
α_2	-5.2	1.22475	-4.25	0.0006
α_3	0.6	1.22475	0.49	0.6309
α_4	8.6	1.22475	7.02	< 0.0001

¿Cuál fue la potencia alcanzada en este experimento, si se usó un nivel de significancia de 0.05?

En este caso $n = 5, a = 4, \hat{\alpha}_1 = -4, \hat{\alpha}_2 = -5.2, \hat{\alpha}_3 = 0.6, \hat{\alpha}_4 = 8.6, \hat{\sigma}^2 = MSE = 10$, con grados de libertad 3 y 16. Por tanto, el parámetro de no centralidad (aproximado) es

$$\lambda \approx \frac{n \sum_{i=1}^a \hat{\alpha}_i^2}{\hat{\sigma}^2} = \frac{5 \times 117.36}{10} = 58.68$$

Calculamos la potencia según (6.4), usando en R la función `power.anova.test()`, o bien la función de probabilidad para la distribución F, `pf()`, especificando parámetro de no centralidad, como se muestra a continuación

```
a=4 #Número de niveles
n=5 #tamaño de muestra por nivel
efectos=c(-4,-5.2,0.6,8.6)
var.entre=sum(efectos^2)/(a-1) #La varianza entre las medias
var.intra=10 #el MSE del modelo
power.anova.test(groups=a,n=n,between.var=var.entre,within.var=var.intra,sig.level = 0.05)

      Balanced one-way analysis of variance power calculation

      groups = 4
        n = 5
between.var = 39.12
within.var = 10
sig.level = 0.05
  power = 0.9999817
NOTE: n is number in each group
#o bien
df1=a-1
df2=a*(n-1)
nc=n*sum(efectos^2)/var.intra #Parámetro de no centralidad
fcrit=qf(0.05,df1=df1,df2=df2,lower.tail=F) #valor crítico criterio de rechazo de H0
pf(fcrit,df1=df1,df2=df2,ncp=nc,lower.tail=F) #cálculo de la potencia con la F no central
[1] 0.9999817
```

es decir, se alcanzó una potencia de 0.99998, muy buena.

Ejemplo 3

En el caso anterior, ¿qué tan grandes deben ser las muestras ($n = ?$) para obtener una potencia de 0.90, a un nivel de significancia de 0.05, cuando las verdaderas medias por nivel del factor son $\mu_1 = 15, \mu_2 = 13, \mu_3 = 19.5, \mu_4 = 27.5$?

Calculemos los efectos mediante la expresión $\alpha_i = \mu_i - \mu$, donde $\mu = \frac{1}{a} \sum_{i=1}^a \mu_i$, es decir:

$$\mu = \frac{1}{4} (15 + 13 + 19.5 + 27.5) = 18.75, \text{ luego,}$$

$$\alpha_1 = 15 - 18.75 = -3.75, \alpha_2 = 13 - 18.75 = -5.75$$

$$\alpha_3 = 19.5 - 18.75 = 0.75, \alpha_4 = 27.5 - 18.75 = 8.75$$

Por tanto $\sum_{i=1}^4 \alpha_i^2 = 124.25$.

Note que para el cálculo del parámetro de no centralidad no se conoce el tamaño de muestra ni la varianza; sin embargo, con los datos experimentales podemos aproximar la varianza con el MSE, por tanto $\lambda = 124.25 \times n/10 = 12.425n$, es decir, el parámetro de no centralidad es función del tamaño muestral. Por otra parte, el valor crítico para la prueba es $f_{0.05,3,4(n-1)}$, donde los grados de libertad del denominador son función del tamaño muestral. Aplicando la ecuación (6.4) y el valor pedido de potencia, obtenemos la siguiente ecuación la cual queda en función del número de réplicas a usar:

$$\pi(\lambda) = P(f_{nc}(3, 4(n-1), 12.425n) > f_{0.05,3,4(n-1)}) = 0.90.$$

Con la función `R power.anova.test()`, omitiendo en ésta el argumento `n=`, se puede calcular el tamaño de muestra necesario, como se ilustra a continuación.

```
#Calculando n necesario para una potencia de 0.90, para
#Rechazar en el test ANOVA con una significancia de 0.05
#cuando mu1=15, mu=13, mu3=19.5 y mu4=27.5
medias=c(15,13,19.5,27.5) #medias supuestas
a=length(medias)
var.entre=var(medias) #varianza entre medias es igual a suma de cuadrados de
                        #los efectos divididos por (a-1)
var.intra=10 #Estimación de sigma^2
power.anova.test(groups=a,between.var=var.entre,within.var=var.intra,power=0.90,sig.level = 0.05)

  Balanced one-way analysis of variance power calculation

    groups = 4
      n = 2.397409
between.var = 41.41667
within.var = 10
  sig.level = 0.05
    power = 0.9

NOTE: n is number in each group
```

Como podemos ver de los resultados R, el tamaño de muestra calculado no da exactamente un valor entero, sino que es $n = 2.397409$. De la Figura 6.2 es claro que para $n = 2$ la potencia es inferior a 0.90 mientras que con $n = 3$ la potencia es de 0.985 aproximadamente. Por tanto el número de réplicas a usar debe ser de $n = 3$ para cada nivel del factor.

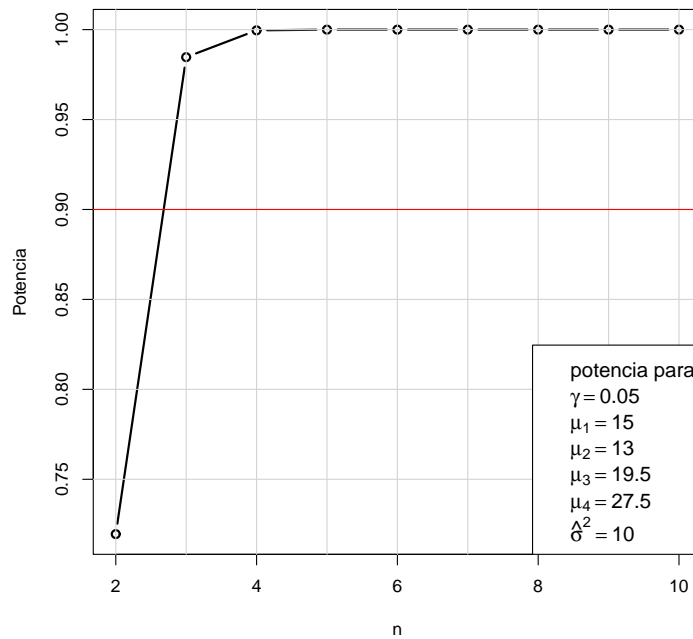


Figura 6.2: Comportamiento de la potencia del test ANOVA vs. n , en un DCA con un factor de efectos fijos, cuando se desea detectar diferencias entre medias a un nivel de significancia de 0.05, con las verdaderas medias sean iguales a $\mu_1 = 15$, $\mu_2 = 13$, $\mu_3 = 19.5$, $\mu_4 = 27.5$.

Sin embargo no siempre es fácil especificar la magnitud de los efectos individuales que se desean detectar. Es más sencillo especificar la magnitud $\Delta = \mu_i - \mu_j$ de la cual se habló previamente, entre cualesquiera dos medias. Denotemos por $\pi(\Delta)$ la potencia de una prueba en Δ . Luego $\pi(\Delta)$ es la probabilidad de rechazar H_0 cuando los efectos de al menos dos de los tratamientos difieren en Δ . $\pi(\Delta)$ es una función de Δ , del tamaño muestral n , del número de niveles del factor a , del nivel de significancia de

la prueba γ , y de la varianza del error σ^2 . Consecuentemente los tamaños de muestra pueden ser determinados si $\pi(\Delta)$, a , γ , y σ^2 son conocidos. Los valores de Δ , $\pi(\Delta)$, a y γ son elegidos por el experimentador, pero la varianza del error debe ser estimada usando datos de un estudio piloto u otro experimento similar. En general debe usarse el valor más grande probable de σ^2 . De acuerdo a [2], *si la varianza supuesta es demasiado pequeña, la potencia de la prueba será más baja que la $\pi(\Delta)$ especificada. Por el contrario, si la varianza supuesta es demasiado grande, la potencia será más grande de lo necesario, y diferencias entre los efectos (o las medias) menores a Δ ocasionarán que H_0 sea rechazada con alta probabilidad.* Lo anterior se deriva del hecho de que en el cálculo del tamaño de muestra con una varianza supuesta muy pequeña, obtendríamos un n más pequeño del necesario, y en consecuencia, menor posibilidad de detectar desviaciones de H_0 que son de interés detectar en el experimento, mientras que si la varianza supuesta es muy grande, entonces el n calculado será también muy grande y en consecuencia se puede alcanzar detección de desviaciones aún más pequeñas de las que interesan. Por tanto, el valor de σ^2 usado en los cálculos de tamaños de muestra es crucial, no puede ser definido de manera arbitraria.

De nuevo, si H_1 es cierta, F_0 sigue una distribución f no central, donde el parámetro de no centralidad bajo un diseño balanceado sigue siendo dado por (6.3). Los cálculos de los tamaños de muestra n requeridos para alcanzar la potencia $\pi(\Delta)$ en Δ para a , γ , y σ^2 dados, según [2], se basan en el hecho de que la situación más difícil de detectar es aquella en la cual los efectos de dos de los niveles del factor (por ejemplo, el primero y el último) difieren en Δ y los otros efectos en medio de estos son todos iguales, es decir, que para alguna constante c ,

$$\begin{aligned}\mu + \alpha_2 &= \mu + \alpha_3 = \cdots = \mu + \alpha_{a-1} = c, \\ \mu + \alpha_1 &= c + \Delta/2, \quad \mu + \alpha_a = c - \Delta/2\end{aligned}$$

En este caso,

$$\lambda = \frac{n \sum_{i=1}^a \alpha_i^2}{\sigma^2} = \frac{n\Delta^2}{2\sigma^2} \quad (6.7)$$

La potencia de la prueba ANOVA depende del tamaño de muestra n a través de la

distribución de MSA/MSE la cual a su vez depende del parámetro de no centralidad λ . La determinación de n de nuevo puede ser hecha iterativamente, dado que los grados de libertad del denominador del estadístico F, $\nu_2 = N - a = n(a - 1)$, depende del valor desconocido n , o valiéndose de la función R, `power.anova.test()`, omitiendo el uso del argumento `n=`.

Ejemplo 4 [2]

Se va realizar un experimento para comparar la extensión a la cual tres tipos particulares de jabón se disuelven en agua. El jabón se comprará en tiendas locales y se cortarán en cubos de peso y dimensión similares de aproximadamente de 1 pulgada y luego se pesarán en una balanza digital de laboratorio con una precisión de lectura de 10 mg. El peso de cada cubo se hará aproximadamente igual al peso del cubo más pequeño cepillando cuidadosamente capas delgadas de estos. Se hará un registro del peso pre experimental de cada cubo. Se asume que las barras de jabón compradas serán típicas de la población de barras de jabón disponibles en las tiendas. Cada cubo debería ser cortado de una barra distinta de jabón comprada a partir de una muestra aleatoria de tiendas con el fin de que el experimento sea tan representativo como pueda ser posible de la población de barras de jabón. Un número igual de observaciones será hecha sobre cada uno de los tres niveles del factor. El experimento se llevará a cabo usando moldes metálicos idénticos para ponqués. El agua será calentada a 100°F (aprox. Temperatura de baño caliente) y cada sección será rápidamente llenada con un cuarto de taza de agua. Las secciones llenas de agua de los moldes son las unidades experimentales y éstas serán asignadas aleatoriamente a los diferentes jabones. Esta regla de asignación define un diseño completamente aleatorizado con n observaciones en cada nivel del factor jabón. Después de 24 horas, los contenidos de los moldes serán vaciados sobre una pantalla y se dejarán escurrir y secar por un periodo de cuatro días con el fin de asegurar que el agua absorbida por cada cubo ha sido removida completamente. Suponga que un experimento piloto fue realizado para obtener una estimación de la varianza la cual fue de cerca de $0.007 \text{ } gmm^2$. En la prueba de hipótesis sobre la igualdad de las medias de pérdida de peso de los tres tipos de

jabones considerados, el experimentador desea detectar una diferencia de al menos $\Delta = 0.25 \text{ grms}$ entre cualesquier dos tipos de jabones, con una probabilidad de 0.90, y una probabilidad de 0.05 de cometer error tipo I. Esta diferencia fue considerada la discrepancia más pequeña en la pérdida de peso de los jabones que sería notable. ¿Cuál es el tamaño de muestra a usar?

Usando un modelo ANOVA de un solo factor efectos fijos en un diseño completamente aleatorizado, para $a = 3$ tratamientos, con $\Delta = 0.25 \text{ grms}$, $\sigma^2 = 0.007 \text{ grm}^2$, el parámetro de no centralidad según (6.7) es

$$\lambda = \frac{n\Delta^2}{2\sigma^2} = n \frac{(0.25)^2}{2 \times 0.007} = 4.4643n$$

además $\nu_2 = a(n-1) = 3(n-1)$, por tanto la potencia está dada por:

$$\pi(\lambda) = P(f_{nc}(2, 3(n-1), 4.4643n) > f_{0.05, 2, 3(n-1)}) = 0.90$$

De nuevo, mediante R, con la función `power.anova.test()`, hallamos n como se muestra a seguir.

```
a=3
Delta=0.25
var.intra=0.007 #Estimación de sigma^2
var.entre=Delta^2/(2*(a-1)) #varianza entre medias es igual a suma de cuadrados de
                           #los efectos divididos por (a-1)
power.anova.test(groups=a,between.var=var.entre,within.var=var.intra,power=0.90,sig.level = 0.05)

Balanced one-way analysis of variance power calculation

  groups = 3
    n = 4.038656
between.var = 0.015625
within.var = 0.007
sig.level = 0.05
  power = 0.9

NOTE: n is number in each group
```

En los resultados R se observa que el tamaño de muestra calculado es $n = 4.038656$, es decir aproximadamente de 4. De la Figura 6.3, vemos que para $n = 4$ la potencia es prácticamente igual a la deseada de 0.90, mientras que para $n = 5$ la potencia es 0.967 aproximadamente. Por tanto el tamaño de muestra a usar en cada nivel es 4.

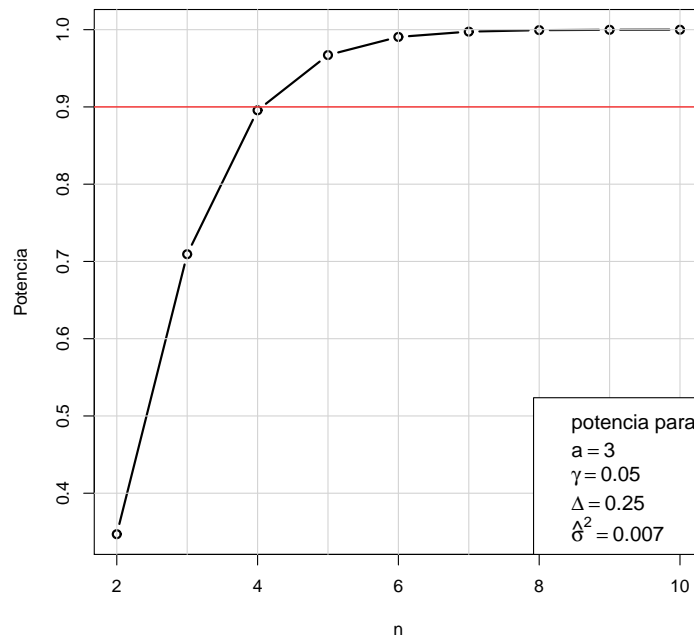


Figura 6.3: Comportamiento de la potencia del test ANOVA vs. n , en un DCA con un factor de efectos fijos con $a = 3$, cuando se desea detectar diferencias $\Delta = \mu_i - \mu_j = 0.25$, entre cualesquiera dos medias, a un nivel de significancia de 0.05.

Resumiendo, el procedimiento para el cálculo del tamaño de muestra n , para satisfacer un requerimiento de potencia a un nivel de significancia dado, es el siguiente:

1. Obtener una estimación de la varianza del error, bien sea de un estudio piloto o de un estudio similar previo.
2. Hallar la expresión para los grados de libertad del denominador de la F y del parámetro de no centralidad (y según la magnitud de los efectos que interesan detectar) en función de n .
3. Calcular iterativamente al nivel de significancia establecido, la potencia para n desde 2 incrementando en cada paso el tamaño muestral en 1.
4. Para el primer valor de n para el cual la potencia es mayor o igual al valor requerido, parar, y dicho valor de n es el buscado.

6.2. Potencia en ANOVA un factor de efectos aleatorios en un DCA y tamaños de muestra

En este caso necesitamos determinar tanto el número de niveles del factor a ser observados y el número n de observaciones a tomar en cada nivel. Una mirada a las fórmulas para los intervalos de confianza para σ_α^2/σ^2 , σ_α^2 y para $\sigma_\alpha^2/(\sigma^2 + \sigma_\alpha^2)$ muestra que para los cálculos de a y n basados en la longitud de intervalos de confianza no será posible actuar de manera directa, dado que tales fórmulas dependen del MSA y del MSE, los cuales son desconocidos antes del experimento. Sin embargo, consideraciones de las varianzas de los estimadores usados para desarrollar los intervalos de confianza nos ayuda a determinar un balance apropiado entre “más tratamientos” y “más replicaciones”. Con este enfoque, si el número total de observaciones N es fijado, y si esperamos que $\sigma_\alpha^2 \geq \sigma^2$, [2] sugieren usar $a = N/2$ y $n = 2$. De otro lado, en el caso más inusual cuando se espera que σ_α^2 sea más pequeño que σ^2 , entonces sugieren a tan pequeña como sea posible comparativamente con n .

Un criterio para determinar los tamaños de muestra, con a fijo, procede como sigue: Sea un diseño balanceado,

$$H_0 : \sigma_\alpha^2 = 0 \text{ vs. } H_1 : \sigma_\alpha^2 > 0, \quad F_0 = \frac{MSA}{MSE}, \text{ y región crítica } F_0 > f_{\gamma, a-1, N-a};$$

cuando H_0 es falsa se tiene que $E(MSA) = \sigma^2 + n\sigma_\alpha^2$, con $\sigma_\alpha^2 > 0$, y puede mostrarse que

$$F^* = \frac{MSA/E(MSA)}{MSE/E(MSE)} \sim f_{a-1, N-a} \quad (6.8)$$

Reemplazando los correspondientes valores esperados en la expresión anterior se obtiene,

$$F^* = \frac{MSA}{MSE \left(1 + n \frac{\sigma_\alpha^2}{\sigma^2}\right)} \sim f_{a-1, N-a} \quad (6.9)$$

Aplicando la ecuación (6.1), y teniendo en cuenta lo anterior, se tiene que cuando $\frac{\sigma_\alpha^2}{\sigma^2} = \Delta$,

$$\text{Potencia} = P\left(\frac{MSA}{MSE} > f_{\gamma, a-1, N-a} \mid \frac{\sigma_\alpha^2}{\sigma^2} = \Delta\right) = P\left(f_{a-1, N-a} > \frac{f_{\gamma, a-1, N-a}}{1 + n\Delta}\right) \quad (6.10)$$

Por tanto, con las tablas f tradicionales, se puede calcular fácilmente la potencia en un ANOVA de efectos aleatorios en un diseño completamente aleatorizado, usando la expresión en (6.10).

Ejemplo 5

Calcular la probabilidad de rechazar $H_0 : \sigma_\alpha^2 = 0$ cuando $\sigma_\alpha^2 = 0.75\sigma^2$, a un nivel de significancia de 0.05, en un experimento ANOVA de efectos aleatorios con un factor usando $a = 4$ niveles y $n = 5$ réplicas por nivel.

Aplicando la expresión en (6.10) tenemos lo siguiente:

$$\text{Pot.} = P\left(f_{3,16} > \frac{f_{0.05,3,16}}{1 + 5 \times 0.75}\right) = P\left(f_{3,16} > \frac{3.24}{1 + 5 \times 0.75}\right) = P(f_{3,16} > 0.682) = 0.58$$

es decir, sólo el 58 % de las veces que se repita tal experimento, se detectará una componente de varianza de $0.75\sigma^2$; o bien, con el diseño experimental planteado, sólo se tiene una probabilidad de 0.58 para detectar dicha componente de varianza.

Nota 6.3. Para calcular la potencia alcanzada en un experimento dado, en (6.10) se utilizan las estimaciones de las componentes de varianza. Tenga en cuenta que si en los cálculos se obtiene $\hat{\sigma}_\alpha^2 < 0$, debe considerar que tal componente es cero y en tal caso la potencia alcanzada es el mismo nivel de significancia de la prueba!

Suponga ahora que deseamos probar la hipótesis

$$H_0 : \sigma_\alpha^2 \leq \eta\sigma^2$$

vs.

$$H_1 : \sigma_\alpha^2 > \eta\sigma^2$$

para una constante η positiva, a un nivel de significancia γ . Se vio previamente con la ecuación (5.9), que la regla de decisión es

$$\text{rechazar } H_0 \text{ si } MSA/MSE > (1 + c\eta) f_{\gamma, a-1, N-a};$$

Recuerde que si el diseño es balanceado, entonces $c = n$. ¿Cuál es la probabilidad de rechazar H_0 si el verdadero valor de σ_α^2/σ^2 es Δ ?, es decir ¿cuál es la probabilidad de

6.2. Potencia en ANOVA un factor de efectos aleatorios en un DCA y tamaños de muestra 151

que $MSA/MSE > k$ cuando $\sigma_\alpha^2/\sigma^2 = \Delta$, donde $k = (1 + c\eta) f_{\gamma, a-1, N-a}$? La potencia en este caso se calcula como: $P\left(MSA/MSE > k \mid \frac{\sigma_\alpha^2}{\sigma^2} = \Delta\right)$. Por (6.9) y tomando $\sigma_\alpha^2/\sigma^2 = \Delta$, y el criterio de rechazo señalado en este caso, se tiene que:

$$P\left(\frac{MSA}{MSE} > k \mid \frac{\sigma_\alpha^2}{\sigma^2} = \Delta\right) = P\left(\frac{MSA}{MSE(1 + n \times \Delta)} > \frac{k}{1 + n \times \Delta}\right) = P\left(f_{a-1, a(n-1)} > \frac{k}{1 + n \times \Delta}\right), \quad (6.11)$$

con $k = (1 + n \times \eta) f_{\gamma, a-1, a(n-1)}$. Suponga que deseamos que la potencia sea igual a π cuando $\sigma_\alpha^2/\sigma^2 = \Delta$, entonces en (6.11), tendríamos que bajo la distribución f con grados de libertad $a - 1, a(n - 1)$, el cuantil de cola superior de nivel π , $f_{\pi, a-1, a(n-1)}$ satisface:

$$f_{\pi, a-1, a(n-1)} = \frac{k}{1 + n \times \Delta}$$

Por tanto

$$\frac{(1 + n \times \eta) f_{\gamma, a-1, a(n-1)}}{1 + n \times \Delta} = f_{\pi, a-1, a(n-1)}$$

de donde:

$$\frac{f_{\gamma, a-1, a(n-1)}}{f_{\pi, a-1, a(n-1)}} = \frac{1 + n \times \Delta}{1 + n \times \eta}$$

Necesitamos seleccionar γ y η para probar H_0 junto con Δ y π . Podemos intentar determinar a y n por ensayo y error, probando hasta hallar valores que satisfagan que:

$$f_{\gamma, a-1, a(n-1)} \times f_{1-\pi, a(n-1), a-1} \leq \frac{1 + n \times \Delta}{1 + n \times \eta} \quad (6.12)$$

Ejemplo 6

Suponga que en un experimento de un solo factor de efectos aleatorios deseamos rechazar hipótesis $H_0 : \sigma_\alpha^2 \leq \eta\sigma^2$ a favor de $H_1 : \sigma_\alpha^2 > \eta\sigma^2$ con $\eta = 1$, a un nivel de significancia de 0.05 y que además deseamos rechazar la hipótesis con alta probabilidad, digamos $\pi = 0.95$, si la verdadera razón $\sigma_\alpha^2/\sigma^2 \geq \Delta = 2.0$. Cuántos niveles del factor y cuántas réplicas deberían tomarse de cada una?

De (6.12), necesitamos hallar a y n tal que $f_{0.05, a-1, a(n-1)} \times f_{0.95, a(n-1), a-1} \leq \frac{1+2n}{1+n}$. Esto puede ser dispendioso sin la ayuda de un programa estadístico. Veamos una rutina

en R para este ejemplo en particular,

```
#La siguiente función de usuario supone vector x con: a=x[1] y n=x[2]
#es diseñada para evaluar la desigualdad en la ecuación (6.12)
potDCAaleat=function(x,eta,delta,pot,alpha){
  v1=x[1]-1;
  v2=x[1]*(x[2]-1)
  f1=qf(alpha,v1,v2,lower.tail=F);
  f2=qf(1-pot,v2,v1,lower.tail=F)
  c=(1+x[2]*delta)/(1+x[2]*eta)
  p=f1*f2
  N=ifelse(p<=c,x[1]*x[2],Inf)
  N
}

#La siguiente Función de usuario determinará los valores óptimos de a y n
#que minimicen N y satisfaciendo la desigualdad en ecuación (6.12)
optimoa.n=function(rangoa,rangon,eta,delta,pot,alpha){
  mini=c()
  for(i in rangoa){
    for(j in rangon){
      x=c(i,j)
      mini=rbind(mini,c(i,j,potDCAaleat(x,eta=eta,delta=delta,pot=pot,alpha=alpha)))
    }
  }
  sol=mini[which.min(mini[,3]),]
  cat("eta", eta,"\\n")
  cat("delta", delta,"\\n")
  cat("Potencia deseada",pot,"\\n")
  cat("significancia",alpha,"\\n")
  cat("Tamaños para mínimo N","\\n")
  cat("Número de niveles a",sol[1],"\\n")
  cat("Número de réplicas n",sol[2],"\\n")
  cat("Tamaño total N",sol[3],"\\n")
}
```

6.2. Potencia en ANOVA un factor de efectos aleatorios en un DCA y tamaños de muestra153

```
#Usamos la función optimoa.n( ) evaluando para a y n desde 2 a 100, respectivamente,  
#con los valores de eta, delta, potencia y nivel significancia del problema  
optimoa.n(c(2:100),c(2:100),eta=1,delta=2,pot=0.95,alpha=0.05)  
eta                1  
delta              2  
Potencia deseada   0.95  
significancia      0.05  
Tamaños para mínimo N  
Número de niveles a 86  
Número de réplicas n 4  
Tamaño total N     344
```

De lo anterior, tenemos que los valores a usar para cumplir los requerimientos de detección y potencia, al nivel de significancia de 0.05, son $a = 86$ niveles y $n = 4$ réplicas por nivel, para un total de 344 observaciones!

Parte II

DISEÑOS EN BLOQUES COMPLETOS CON UN SÓLO FACTOR DE TRATAMIENTOS

Capítulo 7

Diseños en bloques completos aleatorizados con un sólo factor de tratamientos de efectos fijos

Un experimento puede involucrar uno o más factores “nuisance” (de ruido) que aunque no son de interés para el experimentador podría tener un efecto sobre la respuesta. Clasificamos estos factores de ruido en tres tipos:

1. Factores de bloqueo
2. Factores de ruido propiamente
3. Covariables

Cada uno de los anteriores tipos de factores conduce a distintos tipos de análisis.

7.1. Distinción entre factores de bloque, de ruido y covariables

De acuerdo a [2], no siempre es obvio o fácil clasificar un factor “nuisance” como factor de bloque, de ruido o como covariable. La decisión a menudo está gobernada por el objetivo del experimento. [2] establecen la siguiente distinción entre estos elementos.

Un factor “nuisance” es clasificado **como factor de ruido** si el objetivo del experimento es hallar ajustes de los factores de tratamiento cuyas respuestas son al menos afectadas por la variación de niveles de los factores “nuisance”. El ajuste de factores de ruido usualmente puede controlarse durante un experimento pero son incontrolables fuera del laboratorio. Este tipo de factores puede provenir de fuentes internas

o de fuentes externas de variación, por ejemplo, el clima en el cual un producto es usado es una fuente externa de ruido. Las fuentes internas incluyen la variación de los materiales usados en la fabricación de un producto, el uso de las componentes y materiales a lo largo de la vida del producto, entre otros.

Las covariables son factores “nuisance” que no pueden ser controlados pero que pueden ser medidos antes de, o durante el experimento. Algunas veces las covariables son de interés en sí mismas, pero cuando se incluyen en el modelo como variables “nuisance” sus efectos son usados para ajustar las respuestas de modo que los tratamientos puedan ser comparados como si todas las unidades experimentales fueran idénticas.

Un **diseño de bloques** es apropiado cuando el objetivo del experimento es comparar los efectos de diferentes tratamientos promediados sobre un rango de condiciones diferentes. Las unidades experimentales son agrupadas en conjuntos de tal forma que dos unidades experimentales en el mismo grupo son similares y pueden ser medidas bajo condiciones experimentales similares, pero dos unidades experimentales en diferentes grupos probablemente ocasionarán mediciones muy diferentes aún cuando sean asignadas al mismo tratamiento. Los grupos de unidades experimentales similares son llamados **bloques**, y las condiciones que varían de bloque a bloque forman los niveles del **factor de bloqueo**. El análisis de un diseño de bloques incluye la comparación de tratamientos aplicados a las unidades experimentales dentro del mismo bloque. Así, el propósito de bloquear es prevenir que diferencias grandes en las unidades experimentales enmascaren las diferencias entre los niveles del factor de tratamiento, mientras que al mismo tiempo permita que los tratamientos sean examinados bajo diferentes condiciones experimentales.

Los niveles de un factor de bloqueo pueden ser los valores de una covariable que ha sido medida antes del experimento y cuyos valores son usados para agrupar a las unidades experimentales. Muy a menudo sin embargo, los niveles de un factor de bloqueo son agrupamientos de características que no pueden ser medidas convenientemente. Por ej., agrupar los rangos de tiempo en el mismo día dentro del mismo

bloque, puede asegurar que las condiciones ambientales dentro de un bloque sean bastante similares sin necesidad de medirlas.

El diseño de bloques es muy popular debido a que los niveles del factor de bloqueo no necesariamente necesitan ser medidos. Los experimentadores agrícolas pueden saber que parcelas cercanas en un campo son similares, en tanto que las parcelas lejanas no lo son. Los experimentadores industriales pueden saber que dos unidades producidas por una misma máquina tienen características similares, mientras que aquellas producidas por dos máquinas distintas son algo diferentes. Los experimentadores clínicos pueden saber que las mediciones tomadas sobre el mismo sujeto serán similares, mientras que aquellas tomadas sobre sujetos distintos no lo serán. Por tanto los bloques pueden ser formados sin conocer con exactitud los valores de covariables o características de las unidades experimentales o de las condiciones bajo las cuales se realizará el experimento y sobre las cuales se definen los niveles del factor de bloqueo.

7.2. Tamaño de los bloques

Aunque es posible que el número de unidades experimentales en cada bloque sea distinto, se examinará sólo diseños de bloques con igual tamaño de bloque. Representaremos el número de bloques por b y el tamaño de bloque por k .

Algunas veces los tamaños de los bloques son definidos en forma natural, y algunas veces estos deben ser seleccionados específicamente por el experimentador. Por ej., en un experimento de fabricación de un ponqué, las unidades experimentales son los moldes de mezcla batida colocados en distintas posiciones dentro del horno. Si la temperatura no puede ser controlada cuidadosamente, es probable que haya un gradiente de temperatura desde el estante superior al estante inferior del horno, aunque la temperatura en todas las posiciones dentro de un estante puede ser más o menos constante. Si la variable respuesta medida es afectada por la temperatura, las unidades experimentales en el mismo estante son similares, pero aquellas en diferentes estantes son diferentes. Hay un agrupamiento natural de las unidades ex-

perimentales en bloques definidos por el estante del horno. Así, los estantes son los bloques de las unidades experimentales y representan los niveles del factor de bloqueo “temperatura”. El número b de bloques es el número de estantes en el horno. El tamaño de bloque k es el número de moldes de mezcla batida que pueden ser acomodadas en cada estante.

Sin embargo, no siempre el tamaño de bloque es definido naturalmente, sino que es definido por el equipo experimental o definido a juicio del experimentador. En algunos experimentos los equipos usados pueden causar tendencias en los datos (por ejemplo, desgaste, sobrecalentamiento, fatiga, etc.), lo mismo que la variabilidad de lotes de materias primas, etc.

Es común que en experimentos industriales haya una división automática en bloques de acuerdo a intervalos de tiempo en un día, como protección contra condiciones experimentales cambiantes. Un experimento piloto puede ser realizado para determinar la necesidad de bloquear.

Nota 7.1. *Debe tenerse en cuenta que cuando se bloquea innecesariamente, las pruebas de hipótesis serán menos potentes y los intervalos de confianza serán más anchos que los que se obtendrían con un diseño completamente aleatorizado.*

Una vez se haya determinado el número y el tamaño de bloques y se hayan agrupado a las unidades experimentales en tales grupos, a continuación se asignan las unidades a los tratamientos. La peor asignación posible de unidades a tratamientos es asignar todas las unidades de un bloque a un tratamiento, todas las unidades de otro bloque a otro tratamiento, y así sucesivamente. Esta asignación no permite distinguir en el análisis entre diferencias debidas a bloques y las diferencias entre tratamientos. Los efectos de los tratamientos y los del factor de bloqueo quedan confundidos (**confusión de efectos**).

La mejor asignación posible es aquella que asigna a cada tratamiento el mismo número de unidades experimentales en cada bloque. Esto puede lograrse sólo cuando el tamaño del bloque k sea un múltiplo del número de niveles o tratamientos del

factor, a . Este diseño es llamado **diseño de bloques completos**, y el caso especial en cual $k = a$, es el que se conoce usualmente como **diseño de bloques completos aleatorizados**, o simplemente **diseño de bloques aleatorizados**.

Si el tamaño de bloque no es múltiplo de a , entonces el diseño es conocido como un **diseño de bloques incompletos**. Este nombre es algunas veces reservado para el diseño en el cual $k < a$. Estos diseños son más complicados para analizar que los diseños completos.

7.3. El diseño de bloques completos aleatorizados o DBCA

Este diseño con a tratamientos y con $N = ab$ unidades experimentales agrupadas en b bloques de $k = a$ unidades de forma que las unidades dentro de un bloque son similares y unidades en bloques diferentes son sustancialmente distintas. Las $k = a$ unidades experimentales en cada bloque son asignadas aleatoriamente a los a tratamientos o niveles del factor de interés de modo que cada tratamiento es asignado a una unidad por bloque. Así, cada tratamiento aparece una vez en cada bloque y $n = b$ veces en el diseño.

7.3.1. Análisis de diseños de bloques completos aleatorizados

Una tabla de entrada de datos típica de este modelo es como sigue

Tabla 7.1: Tabla de datos experimentales de un DOE en un DBCA

	Bloques						
Tratamientos	1	2	...	j	...	b	Medias de tratamientos
1	Y_{11}	Y_{12}	...	Y_{1j}	...	Y_{1b}	$\bar{Y}_{1\bullet}$
2	Y_{21}	Y_{22}	...	Y_{2j}	...	Y_{2b}	$\bar{Y}_{2\bullet}$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
i	Y_{i1}	Y_{i2}	...	Y_{ij}	...	Y_{ib}	$\bar{Y}_{i\bullet}$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
a	Y_{a1}	Y_{a2}	...	Y_{aj}	...	Y_{ab}	$\bar{Y}_{a\bullet}$
Medias de bloques	$\bar{Y}_{\bullet 1}$	$\bar{Y}_{\bullet 2}$...	$\bar{Y}_{\bullet j}$...	$\bar{Y}_{\bullet b}$	Media global $\bar{Y}_{\bullet\bullet}$

Modelo y ANOVA

El modelo estándar del diseño de bloques completos aleatorizados es:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad (7.1)$$

$$\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad \sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = 0$$

donde μ es la respuesta promedio global, α_i es el efecto del i -ésimo tratamiento o nivel del factor de interés, β_j es el efecto del j -ésimo bloque, ε_{ij} es el error aleatorio asociado, y Y_{ij} es la variable aleatoria que representa la medición en el bloque j del tratamiento i y su media es $E[Y_{ij}] = \mu + \alpha_i + \beta_j$. *Este modelo asume que no existe interacción entre el factor de interés y el factor de bloqueo*; en muchas situaciones esta presunción será razonable, sin embargo, si se sospecha que hay interacción, el tamaño del bloque deberá ser incrementado para permitir estimar estos efectos.

El ANOVA para el modelo en (7.1) es similar al que veremos en el caso de dos factores sin interacción en un diseño completamente aleatorizado (Ver ANOVA en Tabla 7.2), sin embargo varias diferencias deben ser notadas:

- Las combinaciones de tratamientos de un modelo de dos factores en un diseño completamente aleatorizado son asignadas aleatoriamente a las unidades, en cambio en un diseño de bloques, aunque las observaciones son tomadas sobre todas las combinaciones de tratamientos y bloques, sólo los niveles de tratamientos del factor son asignados aleatoriamente a las unidades experimentales (dentro de cada bloque).
- Los niveles del factor de bloque representan agrupamientos intencionales de las unidades experimentales. Esto último desata una controversia sobre si es válido o no realizar el test de igualdad de medias de los bloques. Por otra parte considerando al factor de bloqueo como un factor “nuisance”, no estamos interesados propiamente en sus medias porque en muchos casos es improbable que si repetimos el experimento podamos usar bloques idénticos. Así que en vez

de probar la igualdad de las medias de bloque, más bien se compara la suma de cuadrados medios debida al factor de bloqueo con el MSE del modelo para determinar si bloquear fue o no benéfico en el experimento.

La hipótesis de interés

En el modelo ANOVA dado en (7.1), denotamos las medias de tratamientos por $\mu_{i\bullet} = \mu + \alpha_i$. El objetivo principal es evaluar la igualdad de medias de tratamientos o bien la significancia de los efectos de tratamientos (Ver Tabla 7.2),

$$\begin{array}{ll} H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 & \text{o bien,} \\ H_1 : \text{algún } \alpha_i \neq 0 & \end{array} \quad \begin{array}{l} H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \cdots = \mu_{a\bullet} \\ H_1 : \text{algún } \mu_{i\bullet} \neq \mu_{j\bullet} \end{array} \quad (7.2)$$

El estadístico de prueba y su distribución bajo H_0 , son:

$$F = \frac{MSA}{MSE} \stackrel{H_0}{\sim} f_{a-1, (a-1)(b-1)}, \quad (7.3)$$

y se rechaza para valores estadísticamente grandes.

Nota 7.2. Recuerde que no hay conclusión estadística acerca de la igualdad de medias de bloques o de significancia de efectos de bloques; la comparación meramente se hace como una forma de establecer la utilidad de haber creado bloques en el experimento, y proporciona información adicional para la planeación futura de experimentos similares.

Nota 7.3. Si el MSB es considerablemente mayor que el MSE, esto sugiere que la creación de bloques valió la pena en el sentido de que sirvió para reducir el tamaño del error experimental. De lo contrario, si el MSB es menor que el MSE, entonces la creación de bloques ha bajado la potencia de las pruebas de hipótesis y ha incrementado la longitud de los intervalos de confianza para los contrastes de tratamientos.

Nota 7.4. Si el MSB es menor que el MSE, no es válido pretender que el experimento fue diseñado como un diseño completamente aleatorizado y remover los efectos de bloques del modelo. La aleatorización no es la correcta para un diseño completamente aleatorizado.

Tabla 7.2: ANOVA en un DBCA

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F ₀	Valor P
Factor	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + \frac{b \sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, (a-1)(b-1)} > F_0)$
Bloques	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + \frac{a \sum_{j=1}^b \beta_j^2}{b - 1}$		
Error	$(a - 1)(b - 1)$	SSE	$MSE = \frac{SSE}{(a - 1)(b - 1)}$	$E[MSE] = \sigma^2$		
Total	$ab - 1$	SST				
Fórmulas						
$SSA = \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = b \sum_{i=1}^a \bar{Y}_{i\bullet}^2 - ab \bar{Y}_{\bullet\bullet}^2$ $SSB = \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2 = a \sum_{j=1}^b \bar{Y}_{\bullet j}^2 - ab \bar{Y}_{\bullet\bullet}^2$ $SST = \sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b Y_{ij}^2 - ab \bar{Y}_{\bullet\bullet}^2$ $SSE = SST - SSA - SSB$ <p>donde,</p> $\bar{Y}_{i\bullet} = \frac{1}{b} \sum_{j=1}^b Y_{ij} \text{ Media muestral en el nivel } i \text{ del factor de tratamientos}$ $\bar{Y}_{\bullet j} = \frac{1}{a} \sum_{i=1}^a Y_{ij} \text{ Media muestral en el bloque } j \text{ del factor de bloqueo}$ $\bar{Y}_{\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b Y_{ij} \text{ Media muestral de todas las } ab \text{ observaciones.}$						

7.3.2. Comparaciones múltiples

Se tiene que el estimador de mínimos cuadrados para $\mu + \alpha_i + \beta_j$ es $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{Y}_{i\bullet} + \bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$. Por tanto cualquier contraste de medias de tratamientos $\sum_{i=1}^a c_i \mu_{i\bullet}$ con $\sum_{i=1}^a c_i = 0$, donde $\mu_{i\bullet}$ es la media del i -ésimo nivel del factor de tratamientos, es estimable y su estimador de mínimo cuadrados es $\sum_{i=1}^a c_i \hat{\mu}_{i\bullet} = \sum_{i=1}^a c_i \bar{Y}_{i\bullet}$, cuya varianza es $V[\sum_{i=1}^a c_i \bar{Y}_{i\bullet}] = \sigma^2 \sum_{i=1}^a c_i^2 / b$. Todos los procedimientos de comparaciones múltiples de medias de tratamientos vistos previamente en el caso del diseño completamente aleatorizado, son válidos para los contrastes de medias en el diseño de bloques completos aleatorizados.

La fórmula general para los intervalos de confianza simultáneos del $(1-\gamma)100\%$,

para estas comparaciones, es:

$$\sum_{i=1}^a c_i \mu_{i\bullet} \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet} \pm \varpi \sqrt{MSE \sum_{i=1}^a c_i^2 / b} \right) \quad (7.4)$$

donde ϖ toma los valores correspondientes a los coeficientes críticos para los métodos de Tukey, Bonferroni, Scheffé, LSD, Dunnett, respectivamente, para el nivel de significancia dado y con grados de libertad iguales a $(a-1)(b-1)$.

7.3.3. Chequeo de supuestos del modelo

Como siempre, es necesario corroborar los supuestos acerca de los errores del modelo: Independencia, normalidad y varianza constante, además, también es importante chequear la forma del modelo (carencia de ajuste). Los residuales del modelo son calculados como $\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij}$:

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet} \quad (7.5)$$

Los supuestos sobre los errores son chequeados mediante gráficos de residuales, como se resume a continuación:

Para chequear	Graficar residuales contra:
Independencia	Orden de las observaciones (según espacio o tiempo)
Varianza igual y outliers	Valores predichos, niveles del factor de tratamientos, niveles del factor de bloqueo
Normalidad	Scores normales (también graficar por separado para cada tratamiento, si el número de observaciones n para cada uno es grande, y para cada bloque, si k es grande)

7.3.4. Determinación del número de observaciones por tratamiento

Si el tamaño de bloque k , es predeterminado, podemos calcular el número de bloques que son requeridos para alcanzar un intervalo de confianza de longitud dada, o una prueba de hipótesis de potencia deseada a un nivel de significancia dado, de la misma forma que se procedió con el cálculo del número de réplicas en el modelo de un factor efectos fijos en un diseño completamente aleatorizado. Si el número de

bloques b está limitado, pero los tamaños de bloques k pueden ser muy grandes, pueden usarse las mismas técnicas para calcular el tamaño de bloque k requerido para un diseño general de bloques completos (ver más adelante este diseño).

Nota 7.5. En el diseño de bloques completos aleatorizados, la ecuación del parámetro de no centralidad de la distribución F bajo H_1 para el estadístico MSA/MSE , usa b en lugar de n .

7.4. Diseño general de bloques completos (sección complementaria)

7.4.1. Modelo y ANOVA

Suponga que se tienen $s > 1$ observaciones de cada tratamiento en cada bloque (es decir, el tamaño de cada bloque es $k = as$). Teniendo cada nivel del factor de tratamientos observado más de una vez en cada bloque, se tienen suficientes grados de libertad para medir una interacción bloque \times tratamiento, si tal efecto es esperado. Por tanto, hay dos modelos estándar para el diseño de bloques completos:

Modelo tratamiento - bloque (sin interacción)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (7.6)$$

$$i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, s, \quad \sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = 0$$

Modelo de interacción tratamiento - bloque

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (7.7)$$

$$i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, s, \quad \sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$$

En (7.7), $(\alpha\beta)_{ij}$ representa el efecto de interacción del bloque j con el nivel i del factor de tratamientos. En ambos modelos se supone que $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$; estos supuestos deben ser chequeados.

En la Tabla 7.3 se muestra la ANOVA para el modelo sin interacción y en la Tabla 7.4 el caso con interacción. Por el momento no se explica el significado de los efectos de interacción, esto será tratado al presentar los modelos ANOVA de efectos fijos con dos factores de tratamientos.

Tabla 7.3: ANOVA en un DBCA con bloques de tamaño $k = sa$, sin interacción factor-bloques

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F ₀	Valor P
Factor	a - 1	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + \frac{bs \sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
Bloques	b - 1	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + \frac{as \sum_{j=1}^b \beta_j^2}{b - 1}$		
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	abs - 1	SST				

Fórmulas

dfe = (abs - 1) - (a - 1) - (b - 1) = abs - a - b + 1, los grados de libertad del error.

$$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bs \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$$

$$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = as \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$$

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$$

$$SSE = SST - SSA - SSB$$

donde,

$\bar{Y}_{i\bullet\bullet} = \frac{1}{bs} \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}$ Media muestral en el nivel i del factor de tratamientos

$\bar{Y}_{\bullet j\bullet} = \frac{1}{as} \sum_{i=1}^a \sum_{k=1}^s Y_{ijk}$ Media muestral en el bloque j del factor de bloqueo

$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abs} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}$ Media muestral de todas las abs observaciones.

Tabla 7.4: ANOVA en un DBCA con bloques de tamaño $k = sa$, con interacción factor-bloques

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F ₀	Valor P
Factor	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + \frac{bs \sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
Bloques	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + \frac{as \sum_{j=1}^b \beta_j^2}{b - 1}$		
Interacción	dfi	$SSAB$	$MSAB = \frac{SSAB}{dfi}$	$E[MSAB] = \sigma^2 + \frac{s \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta)_{ij}^2}{dfi}$	$\frac{MSAB}{MSE}$	$P(f_{dfi, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abs - 1$	SST				
Fórmulas						
<p>dfi = $(a - 1)(b - 1)$, los grados de libertad de la interacción factor de tratamientos con bloques</p> <p>dfe = $(abs - 1) - (a - 1) - (b - 1) - (a - 1)(b - 1) = ab(s - 1)$, los grados de libertad del error.</p> <p>$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bs \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = as \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}^2 - abs \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s (Y_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = s \sum_{i=1}^a \sum_{j=1}^b (Y_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$</p> <p>$SSE = SST - SSA - SSB - SSAB$</p> <p>donde,</p> <p>$\bar{Y}_{i\bullet\bullet} = \frac{1}{bs} \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}$ Media muestral en el nivel i del factor de tratamientos</p> <p>$\bar{Y}_{\bullet j\bullet} = \frac{1}{as} \sum_{i=1}^a \sum_{k=1}^s Y_{ijk}$ Media muestral en el bloque j del factor de bloque</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abs} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^s Y_{ijk}$ Media muestral de todas las abs observaciones.</p>						

Nota 7.6. Si en el modelo con interacción el test sobre la significancia de los efectos de interacción rechaza H_0 (es decir, se concluye que al menos uno de los efectos de interacción es significativo), una prueba global sobre la igualdad de los efectos de tratamientos puede no ser de interés.

Comparaciones múltiples en el diseño general de bloques completos

En el modelo dado en (7.6), pueden usarse los métodos de comparaciones múltiples de medias de tratamientos $\mu_{i\bullet}$ donde las fórmulas para los intervalos de confianza simultáneos del $(1 - \gamma)100\%$ son de la forma:

$$\sum_{i=1}^a c_i \mu_{i\bullet} \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet\bullet} \pm \varpi \sqrt{MSE \sum_{i=1}^a \frac{c_i^2}{bs}} \right) \quad (7.8)$$

donde ϖ toma los valores correspondientes a los coeficientes críticos para los métodos de Tukey, Bonferroni, Scheffé, LSD, Dunnett, respectivamente, para el nivel de significancia dado y con grados de libertad iguales a $(abs - a - b + 1)$.

Si se trata del modelo en (7.7), en presencia de interacción significativa no se deben hacer comparaciones globales de las medias de tratamiento promediando sobre todos los bloques, sino que las comparaciones múltiples deben ser evaluadas dentro de cada bloque, es decir, se comparan las medias de celdas $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ con j fijo, luego las fórmulas para los intervalos de confianza simultáneos del $(1 - \gamma)100\%$ son de la forma:

$$\sum_{i=1}^a c_i \mu_{ij} = \sum_{i=1}^a c_i (\alpha_i + (\alpha\beta)_{ij}) \in \left(\sum_{i=1}^a c_i \bar{Y}_{ij\bullet} \pm \varpi \sqrt{MSE \sum_{i=1}^a \frac{c_i^2}{s}} \right) \quad (7.9)$$

donde ϖ toma los valores correspondientes a los coeficientes críticos para los métodos de Tukey, Bonferroni, Scheffé, LSD, Dunnett, respectivamente, para el nivel de significancia dado y con grados de libertad iguales a $ab(s - 1)$.

Validación de supuestos

Se realiza de la misma forma como en el diseño de bloques completos aleatorizados.

Determinación del número de observaciones

Un diseño general de bloques completos tiene $N = abs$ unidades experimentales divididas en b bloques de tamaño $k = as$. k y b deben ser elegidos de acuerdo a las condiciones experimentales, las restricciones presupuestales y los requerimientos de

longitud de los intervalos de confianza que se quieren evaluar o según las potencias para las pruebas de hipótesis, como se explicó en el caso del modelo de un factor efectos fijos diseño completamente aleatorizado. Para calcular el número de réplicas s de cada tratamiento en cada bloque para alcanzar una potencia en la prueba de que no existen diferencias entre medias del factor de tratamientos, s debe satisfacer la siguiente desigualdad (vea nuevamente la ecuación (6.7) y haga $n = bs$):

$$s \geq \frac{2\sigma^2\lambda}{b\Delta^2} \quad (7.10)$$

donde Δ es la mínima diferencia que se desea detectar entre cualesquiera dos medias de tratamientos y λ es el parámetro de no centralidad, que se supone especificado. Podemos resolver para s o para b dependiendo de si el tamaño de bloque o si el número de bloques puede ser grande.

7.5. Efectividad en la construcción de bloques: Eficiencia relativa (ER) y precisión de un experimento

Según [8], Yates (1935) introdujo la noción de eficiencia relativa (ER) en el contexto de la estimación de las comparaciones de tratamientos. Para dos diseños, D_1 , y D_2 la eficiencia relativa de D_1 a D_2 es definida como

$$ER(D_1 \text{ a } D_2) = \frac{\text{eficiencia } D_1}{\text{eficiencia } D_2} = \frac{1/\text{Var } D_1}{1/\text{Var } D_2} \quad (7.11)$$

Donde $\text{Var } D_i$ se refiere a $\text{Var}(\sum c_i \hat{\mu}_{i\bullet})$ para el diseño D_i , $i = 1, 2$. En particular, D_1 es un diseño de bloques completos aleatorizados (DBCA) con a tratamientos y b bloques y D_2 es un DCA con b replicaciones por cada tratamiento. La eficiencia relativa depende de la verdadera varianza para los dos diseños, sin embargo, en la práctica se desconocen estas varianzas y sólo se tienen disponibles los datos del experimento realizado bajo el DBCA para estimar su correspondiente σ^2 . Lo mejor que se puede hacer es obtener un estimación de la eficiencia relativa, que se denotará como EER, [8].

De acuerdo a [6], la eficiencia relativa se determina según el concepto de información formulado por Fisher (1960), concepto que mide la información que proporciona

7.5. Efectividad en la construcción de bloques: Eficiencia relativa (ER) y precisión de un experimento 17

la diferencia estimada entre dos medias de tratamientos respecto a la diferencia real entre las respectivas medias poblacionales, y es calculada como

$$I = \frac{(f+1)}{(f+3)} \frac{1}{s^2} \quad (7.12)$$

con s^2 la varianza estimada del error experimental y f sus grados de libertad. Si σ^2 fuera conocido entonces la información sería $I = 1/\sigma^2$. De acuerdo con [6], al reducir la variabilidad muestral debe crecer la información que se tiene en el experimento sobre la diferencia de medias de tratamientos poblacionales y esto es medido a través de la eficiencia relativa. El procedimiento de Fisher estima la eficiencia del diseño 1 en relación con el diseño 2 mediante ([3, 6]),

$$\text{EER (DBCA a DCA)} = \frac{I_1}{I_2} = \frac{(f_1+1)(f_2+3)s_2^2}{(f_2+1)(f_1+3)s_1^2} \quad (7.13)$$

Donde s_1^2 y s_2^2 son los cuadrados medios del error (los MSE) de los diseños 1 (DBCA) y 2 (DCA), respectivamente, y f_1 y f_2 son los grados de libertad. Aunque s_2^2 no se conoce, éste puede ser estimado con sólo los datos del DBCA, [3]:

$$\begin{aligned} s_2^2 &\simeq \frac{f_b MSB + (f_a + f_e) s_1^2}{f_b + f_a + f_e} \\ &\simeq \frac{(b-1) MSB + b(a-1) s_1^2}{(ab-1)}, \end{aligned} \quad (7.14)$$

con:

f_b = grados de libertad de los bloques

f_a = grados de libertad de los tratamientos

f_e = grados de libertad del error.

De acuerdo a [3, 8], otra manera de calcular la eficiencia relativa es (sin considerar el factor de corrección de Fisher)

$$\begin{aligned} \text{ER} &= k + (1-k) H \\ k &= \frac{b(a-1)}{ab-1} \text{ y } H = \frac{MSB}{s_1^2} \end{aligned} \quad (7.15)$$

Lentner, Arnold and Hinkelman (1989), citado en [6], señalan que no existe una prueba legítima para $H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0$, es decir H en la ecuación (7.15) no es un estadístico de prueba adecuado; sin embargo, hay una relación entre H y ER que

permite interpretar a H . Reemplazando H y k por sus valores específicos en (7.15), tenemos que

$$ER = \frac{(b-1)MSB + b(a-1)s_1^2}{(ab-1)s_1^2}, \quad (7.16)$$

de donde se observa que (7.16) es igual a s_2^2 dado en (7.14), dividido por s_1^2 . Se puede mostrar que:

$$ER = 1 \Leftrightarrow H = 1$$

$$ER < 1 \Leftrightarrow H < 1$$

$$ER > 1 \Leftrightarrow H > 1$$

Nota 7.7. Si bien H no proporciona información completa sobre la eficiencia relativa, permite una verificación rápida de la efectividad del uso de bloques, [6].

Interpretación de los valores de E.R: Si

$ER = 1$ (100 %) El bloqueo no ha ayudado ni perjudicado

$ER < 1$ (100 %) El bloqueo no ha sido útil (fue ineficiente)

$ER > 1$ (100 %) El bloqueo ha sido útil (fue eficiente)

Para intepretar estos resultados en términos de economía en la experimentación, tenga en cuenta que se requieren $n = b \times ER$ replicaciones por tratamiento para que un DCA sea tan efectivo como un DBCA con b bloques. Por ejemplo:

Si $ER = 0.70$ (70 %), no es útil el DBCA, ya que 7 réplicas de un DCA darían tanta información como 10 bloques o réplicas de un DBCA.

Si $ER = 3.0$ (300 %), es útil el DBCA ya que el DCA para obtener la misma eficiencia que el DBCA, necesitaría 3 réplicas por cada bloque en el DBCA, o en otras palabras, el DCA requeriría 2 réplicas más por cada bloque en el DBCA.

Nota 7.8. De acuerdo con [8], la ER sólo habla de la precisión de las estimaciones y no de la potencia (sensibilidad del experimento). Por ésta razón se recomienda considerar un DBCA mejor que un DCA cuando $E.R=125\%$.

Nota 7.9. Compare la ER con la ecuación para EER dada en (7.13) y note que

$$EER = \frac{(f_1 + 1)(f_2 + 3)}{(f_2 + 1)(f_1 + 3)} \times ER$$

7.6. Ejemplo: Comparación de cuatro métodos de ensamble

Un equipo de mejora investiga el efecto de cuatro métodos de ensamble A, B, C, y D, sobre el tiempo de ensamble en minutos. En primera instancia la estrategia experimental fue aplicar cuatro veces los cuatro métodos de ensamble en orden aleatorio en un diseño completamente aleatorizado. Sin embargo, los experimentadores se dan cuenta que hay cuatro operadores con distintas destrezas y consideran que esto puede afectar de manera significativa los tiempos de ensamble, y por ende la comparación de los métodos, entonces debe utilizarse un diseño de bloques completos aleatorizados para controlar activamente en el experimento a los operadores que realizarán el ensamble. Los datos recolectados en este diseño fueron los siguientes:

Operador	Método			
	A	B	C	D
1	6	7	10	10
2	9	10	16	13
3	7	11	11	11
4	8	8	14	9

Se desea determinar si existen diferencias entre los cuatro métodos y cómo se diferencian, además elegir el mejor método, si lo hay.

En primera instancia se analizan los gráficos boxplots de los tiempos vs. métodos y tiempos vs. operadores, como se ilustra en la Figura 7.1 y el gráfico de medias como se muestra en la Figura 7.2, donde la línea horizontal representa la gran media $\bar{Y}_{\bullet\bullet}$, respecto a la cual vemos la posición de las medias de tratamientos $\bar{Y}_{i\bullet}$ y las medias de bloques $\bar{Y}_{\bullet j}$. Se observan diferencias en los tiempos promedios de ensamble entre los cuatro métodos. A primera vista parece que los dos mejores métodos son el A y B. Parece que con el método C los tiempos alcanzados tienen mayor dispersión que en los otros dos casos. Por su parte, el operador 1 parece tener en promedio el menor tiempo de ensamble, en tanto que los operadores 3 y 4 en promedio y tienen aproximadamente los mismos tiempos. El operador 2 es más lento y opera con una mayor variabilidad lo que nos dice que es menos consistente en sus acciones.

El modelo estadístico es

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2),$$

con $i = 1, 2, 3, 4$ para denotar los métodos A, B, C y D respectivamente, y $j = 1, 2, 3, 4$ operadores.

α_i representa el efecto sobre el tiempo promedio de ensamble debido al método i -ésimo

β_j es el efecto del j -ésimo operador sobre el tiempo promedio de ensamble.

Restricciones: $\sum_{i=1}^4 \alpha_i = \sum_{j=1}^4 \beta_j = 0$.

Salida R 7.6.1. Resultados Test ANOVA.

```
> anova(diseño)
Analysis of Variance Table

Response: Tiempo

      Df Sum Sq Mean Sq F value    Pr(>F)
Método  3   61.5    20.5    10.25 0.002919 **
Operador 3   28.5     9.5     4.75 0.029846 *
Residuals 9   18.0     2.0
---
```

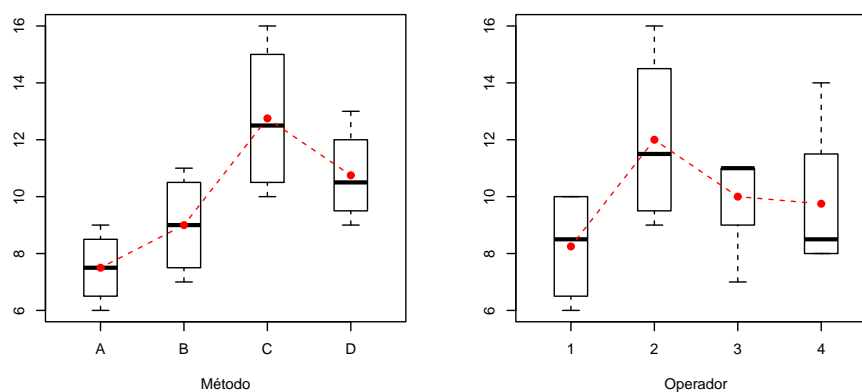


Figura 7.1: Boxplots tiempos de ensamble según método y según operador

El test de hipótesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$, es rechazado con un valor p de $P(f_{3,9} > F_0) = P(f_{3,9} > 10.25) = 0.0029$ a un nivel de significancia de 0.05, es decir, los tiempos medios de ensamble no son iguales para todos los métodos.

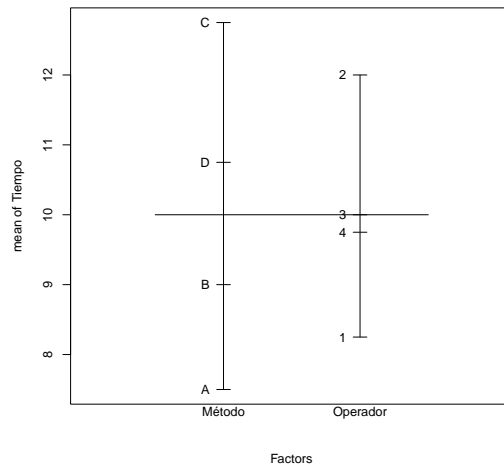


Figura 7.2: Medias de tiempos de ensamble según método y según operador

Para evaluar la eficiencia del diseño tenemos que la *seudo - prueba sobre los efectos de bloque* (operador) tiene un $H = F_0 = 4.75$ y un valor P: $P(f_{3,9} > F_0) = 0.029846$, que nos indica que el bloqueo por operario resultó útil o eficiente para reducir el error experimental y por tanto un ahorro en cuanto al número de réplicas necesarias para detección de diferencias entre los métodos. También podemos evaluar la eficiencia del diseño usando las ecuaciones (7.13) y (7.16), tenemos que

$$ER = \frac{(b-1)MSB + b(a-1)s_1^2}{(ab-1)s_1^2} = \frac{3 \times 9.5 + 4 \times 3 \times 2}{(4 \times 4 - 1) \times 2} = 1.75$$

$$EER \times 100\% = \frac{(f_1 + 1)(f_2 + 3)}{(f_2 + 1)(f_1 + 3)} \times ER \times 100\% = \frac{(9 + 1)(12 + 3)}{(12 + 1)(9 + 3)} \times 1.75 \times 100\% = 168\%$$

luego, el DBCA fue altamente eficiente, ya que para un DCA se requeriría $n = b \times EER = 4 \times 1.68 = 6.72 \approx 7$ (o bien, $n = b \times ER = 4 \times 1.75 \approx 7$) réplicas por tratamiento para obtener la misma información que este DBCA.

Comparemos ahora los tiempos medios de ensamble para los cuatro métodos mediante el procedimiento de Tukey. Con R se obtuvieron los siguientes resultados:

Salida R 7.6.2. Resultados para comparaciones pareadas por Tukey sobre las medias de tratamientos (Métodos de ensamble).

```
> TukeyHSD(diseño,"Método",conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Tiempo ~ Método + Operador)
$'Método'
      diff      lwr      upr    p adj
B-A  1.50 -1.6217987 4.621799 0.4758801
C-A  5.25  2.1282013 8.371799 0.0024211
D-A  3.25  0.1282013 6.371799 0.0412298
C-B  3.75  0.6282013 6.871799 0.0195634
D-B  1.75 -1.3717987 4.871799 0.3548246
D-C -2.00 -5.1217987 1.121799 0.2566550

> HSD.test(diseño,"Método", group=TRUE,console=TRUE) #Comparaciones de Tukey
Study: diseño ~ "Método"
HSD Test for Tiempo
Mean Square Error: 2
Método, means
      Tiempo      std r Min Max
A       7.50 1.290994 4    6    9
B       9.00 1.825742 4    7   11
C      12.75 2.753785 4   10   16
D      10.75 1.707825 4    9   13
Alpha: 0.05 ; DF Error: 9
Critical Value of Studentized Range: 4.41489
Honestly Significant Difference: 3.121799
Means with the same letter are not significantly different.
Groups, Treatments and means
a      C      12.75
ab     D      10.75
bc     B       9
c      A       7.5
```

Como podemos ver de lo anterior, Tukey define tres grupos de medias que se trasladan: grupo A: métodos C y D, grupo B: métodos D y B y grupo C: métodos A y B. Hay un problema de potencia del procedimiento debido a la gran dispersión en los datos en algunos operarios y métodos como veremos a seguir.

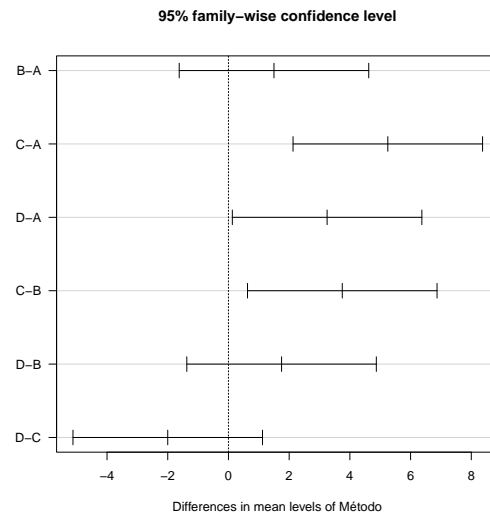


Figura 7.3: I.C Tukey del 95 % de confianza, para las diferencias de medias de tratamientos (Métodos de ensamble)

Validemos los supuestos de normalidad y de varianza constante mediante gráficos de residuales, ver : La normalidad parece un supuesto razonable para los errores en este experimento. Sin embargo, hay problemas con el supuesto de varianza constante. Con los métodos B y C la dispersión de los residuales es mayor comparativamente a la dispersión en los métodos A y D. Con el método A hay menos dispersión. A nivel del operador 3 la dispersión de los residuales es mayor aparentemente debido a una observación que es subestimada. A pesar de estos inconvenientes es posible llegar a una conclusión: “Elegir como mejor método el A”, dado que tiene menor tiempo promedio observado y menor dispersión”(esto último indica que los operadores con este método son más consistentes).

Adicionalmente, se estimaron las medias y los efectos de cada método de ensamble, y sus intervalos de confianza (cuyos niveles de confianza pueden ser un poco diferentes a los esperados dados los problemas de varianza). Se deja al lector la interpretación de estos resultados.

Salida R 7.6.3. Estimaciones de medias y efectos de tratamientos (Métodos de ensamblaje).

```
> lsmeans(diseño, "Método")
```

Método	lsmean	SE	df	lower.CL	upper.CL
A	7.50	0.7071068	9	5.900413	9.099587
B	9.00	0.7071068	9	7.400413	10.599587
C	12.75	0.7071068	9	11.150413	14.349587
D	10.75	0.7071068	9	9.150413	12.349587

Results are averaged over the levels of: Operador
Confidence level used: 0.95

```
> rbind(efect.métodoA, efect.métodoB, efect.métodoC, efect.métodoD)
```

	Estimate	Std. Error	t value	Pr(> t)	lower CI	upper CI
Método:efecto método A	-2.50	0.6123724	-4.082483	0.002747777	-3.8852827	-1.1147173
Método:efecto método B	-1.00	0.6123724	-1.632993	0.136904126	-2.3852827	0.3852827
Método:efecto método C	2.75	0.6123724	4.490731	0.001508930	1.3647173	4.1352827
Método:efecto método D	0.75	0.6123724	1.224745	0.251759476	-0.6352827	2.1352827

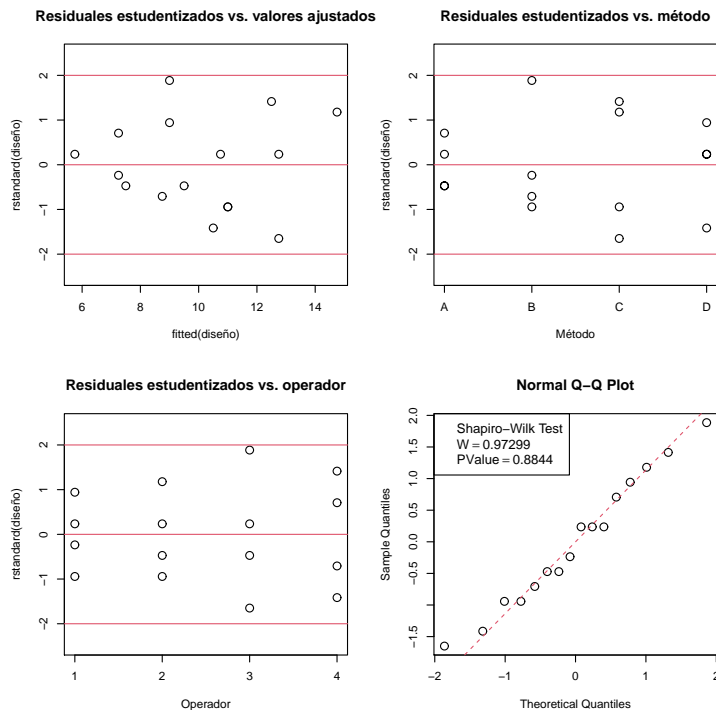


Figura 7.4: Gráficos de residuales internamente estudentizados y test de Normalidad en ejemplo 7.6.

EL programa R que fue usado para el análisis de estos datos, el ajuste del modelo ANOVA, estimación de medias y efectos de tratamientos y análisis de residuales, se muestra a continuación.

Código R 7.6.1. Programa R para ajuste modelo Anova en el DBCA del Ejemplo 7.6.

```
library(gmodels)
library(multcomp)
library(lsmeans)
library(agricolae)

ensamble=data.frame(Método=factor(rep(c("A","B","C","D"),times=4)),Operador=factor(rep(1:4,each=4)),
                    Tiempo=scan())

6 7 10 10
9 10 16 13
7 11 11 11
8 8 14 9

ensamble

attach(ensamble)
#Gráficos descriptivos
mediasMét=sapply(split(Tiempo,Método),mean)
mediasOpe=sapply(split(Tiempo,Operador),mean)

win.graph(width=7,height=3.5)
layout(matrix(c(1,1,2,2),nrow=1))
boxplot(Tiempo~Método,boxwex=0.4,xlab="Método")
lines(mediasMét,type="b",pch=19,col=2,lty=2)
boxplot(Tiempo~Operador,boxwex=0.4,xlab="Operador")
lines(mediasOpe,type="b",pch=19,col=2,lty=2)

#Gráfico de Medias
plot.design(Tiempo~Método+Operador)

#AJUSTANDO EL MODELO ANOVA Y OBTENCIÓN DE LA TABLA ANOVA
diseño=aov(Tiempo~Método+Operador)
anova(diseño)

#OBTENCIÓN DE MEDIAS DE TRATAMIENTO CON SUS I.C DEL 95%
lsmeans(diseño,"Método")
```

```

#OBTENIENDO EFECTOS DE TRATAMIENTOS, RESULTADOS PARA TEST DE SIGNIFICANCIA Y SUS I.C DEL 95%
efect.métodoA=fit.contrast(diseño,"Método",rbind(":efecto método A"=c(3/4,-1/4,-1/4,-1/4)),conf=0.95)
efect.métodoB=fit.contrast(diseño,"Método",rbind(":efecto método B"=c(-1/4,3/4,-1/4,-1/4)),conf=0.95)
efect.métodoC=fit.contrast(diseño,"Método",rbind(":efecto método C"=c(-1/4,-1/4,3/4,-1/4)),conf=0.95)
efect.métodoD=fit.contrast(diseño,"Método",rbind(":efecto método D"=c(-1/4,-1/4,-1/4,3/4)),conf=0.95)
rbind(efect.métodoA,efect.métodoB,efect.métodoC,efect.métodoD)

#INTERVALOS DE TUKEY PARA LAS DIFERENCIAS DE MEDIAS DEL FACTOR DE TRATAMIENTOS
TukeyHSD(diseño,"Método",conf.level=0.95)
HSD.test(diseño,"Método", group=TRUE,console=TRUE) #Comparaciones de Tukey

#GRÁFICOS DE INTERVALOS DE TUKEY
plot(TukeyHSD(diseño,"Método",conf.level = 0.95),cex.lab=0.8,las=1)

#OBTENIENDO TEST DE NORMALIDAD SOBRE RESIDUALES INTERNAMENTE ESTUDENTIZADOS
shapiro.test(rstandard(diseño))

#OBTENIENDO DE GRÁFICOS PARA VALIDACIÓN DE SUPUESTOS CON RESIDUOS INTERNAMENTE ESTUDENTIZADOS
nf=layout(rbind(c(1,1,2,2),c(3,3,4,4)))
plot(fitted(diseño),rstandard(diseño),ylim=c(-2.5,2.5),cex=1.5,
      main="Residuales estudentizados vs. valores ajustados")
abline(h=c(-2,0,2),col=2)

stripchart(rstandard(diseño)~Método,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Método",
            main="Residuales estudentizados vs. método")
abline(h=c(-2,0,2),col=2)

stripchart(rstandard(diseño)~Operador,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Operador",
            main="Residuales estudentizados vs. operador")
abline(h=c(-2,0,2),col=2)

qqnorm(rstandard(diseño),cex=1.5)
qqline(rstandard(diseño),col=2,lty=2)
legend("topleft",legend=c("Shapiro-Wilk Test",expression(W==0.97299),expression(PValue== 0.8844)),
      cex=1.1)
detach(ensamble)

```

7.7. Problema para análisis

Un supervisor de producción quiere estudiar la eficiencia de 6 operarios que realizan el montaje de un equipo electrónico. Como hay muchas alternativas de ensamble, controló la habilidad del operario seleccionando tres montajes: A, B, C. Luego se asignó aleatoriamente cada método a todos los operarios y se registró el tiempo en minutos. Los datos obtenidos aparecen en la siguiente tabla.

Montaje	Operador					
	1	2	3	4	5	6
A	20.2	22.6	19.2	22.5	18.7	21.5
B	23.7	24.1	22.6	24.3	20.2	21.5
C	21.4	23.0	22.9	22.0	19.8	20.1

1. ¿Qué tipo de diseño DCA o DBCA describe el enunciado (cuál sería el factor de bloqueo si fuese este caso)?
2. ¿Cuál es la estructura de tratamientos y qué tipo de efectos (fijos o aleatorios)?
¿Cuál es la variable respuesta?
3. ¿Modelo Anova apropiado?
4. Enuncie y decida sobre la prueba de hipótesis relacionada a los tratamientos.
5. ¿Se justifica el diseño que se utilizó?
6. Los resultados de la prueba Tukey para Montaje y Operario se encuentran en la Salida R 7.7.2. Con estos resultados, ¿Qué recomendaría usted si se desea minimizar el tiempo?
7. Estime también los efectos principales del factor de tratamientos y sus I.C del 95 %. Interprete los resultados.
8. Valide supuestos.

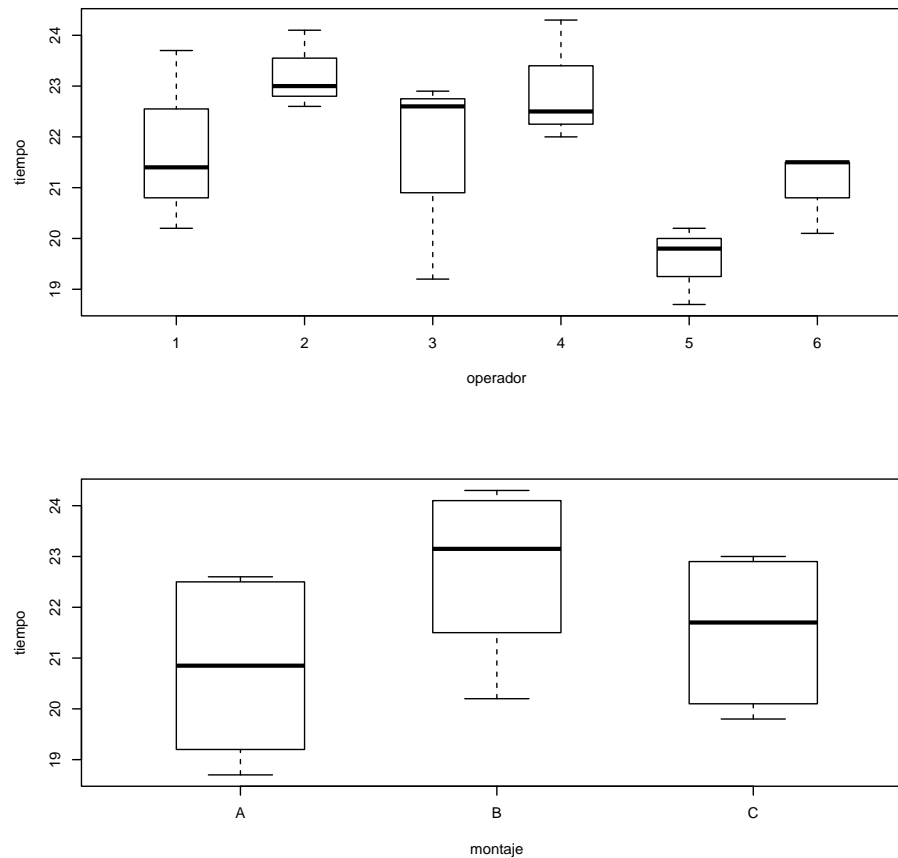


Figura 7.5: Boxplots tiempos de montaje según operador y según tipo de montaje

Salida R 7.7.1. Resultados Test ANOVA.

```
> anova(diseño2)
Analysis of Variance Table

Response: tiempo

      Df Sum Sq Mean Sq F value    Pr(>F)
operador  5 26.665   5.333   5.4363 0.01129 *
montaje   2 11.610   5.805   5.9174 0.02015 *
Residuals 10  9.810   0.981
---
```

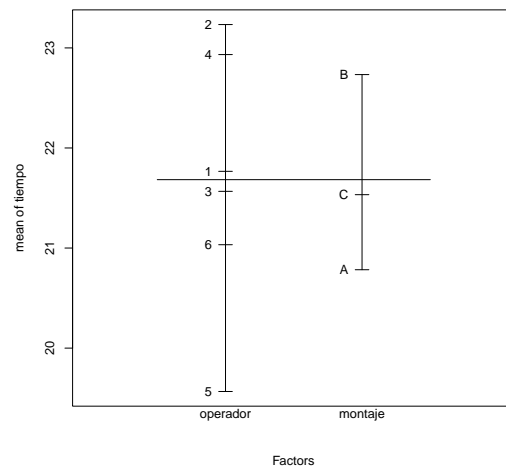


Figura 7.6: Medias de tiempos de ensamble según operador y según tipo de montaje

Salida R 7.7.2. Resultados para comparaciones pareadas por Tukey sobre las medias según operador y según montaje.

```
> TukeyHSD(diseño2, "operador", conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = tiempo ~ operador + montaje)
$operador
      diff      lwr      upr    p adj
2-1  1.4666667 -1.342216  4.2755495 0.4985635
3-1 -0.2000000 -3.008883  2.6088829 0.9998282
4-1  1.1666667 -1.642216  3.9755495 0.7037371
5-1 -2.2000000 -5.008883  0.6088829 0.1543419
6-1 -0.7333333 -3.542216  2.0755495 0.9360751
3-2 -1.6666667 -4.475550  1.1422162 0.3756824
4-2 -0.3000000 -3.108883  2.5088829 0.9987745
5-2 -3.6666667 -6.475550 -0.8577838 0.0101615
6-2 -2.2000000 -5.008883  0.6088829 0.1543419
4-3  1.3666667 -1.442216  4.1755495 0.5658844
5-3 -2.0000000 -4.808883  0.8088829 0.2191947
6-3 -0.5333333 -3.342216  2.2755495 0.9827533
5-4 -3.3666667 -6.175550 -0.5577838 0.0176029
6-4 -1.9000000 -4.708883  0.9088829 0.2594758
6-5  1.4666667 -1.342216  4.2755495 0.4985635
```

```
> TukeyHSD(diseño2,"montaje",conf.level=0.95)
Tukey multiple comparisons of means
95% family-wise confidence level
Fit: aov(formula = tiempo ~ operador + montaje)
$montaje
      diff      lwr      upr    p adj
B-A  1.95  0.3824202 3.5175798 0.0167000
C-A  0.75 -0.8175798 2.3175798 0.4208163
C-B -1.20 -2.7675798 0.3675798 0.1398316
```

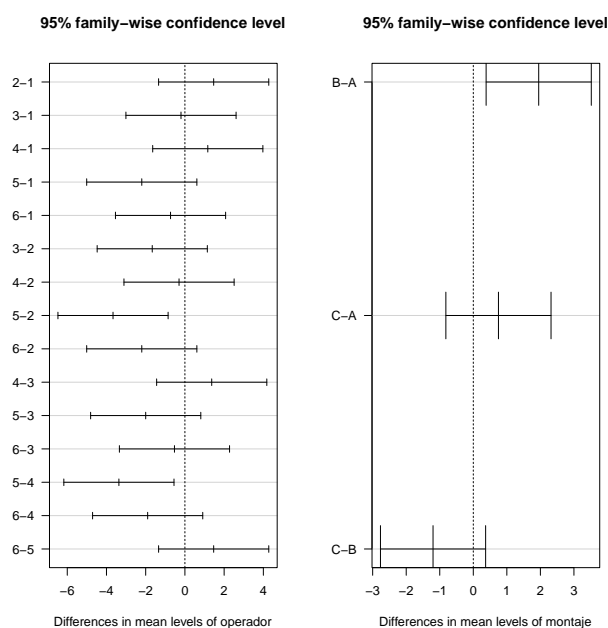


Figura 7.7: Gráficas de los intervalos de confianza de Tukey

Salida R 7.7.3. Estimaciones efectos de los operadores.

	Estimate	Std. Error	t value	Pr(> t)	lower CI	upper CI
operador:efecto operador 1	0.08333333	0.5220153	0.1596377	0.876344811	-1.07978929	1.2464560
operador:efecto operador 2	1.55000000	0.5220153	2.9692615	0.014063570	0.38687737	2.7131226
operador:efecto operador 3	-0.11666667	0.5220153	-0.2234928	0.827651475	-1.27978929	1.0464560
operador:efecto operador 4	1.25000000	0.5220153	2.3945657	0.037663712	0.08687737	2.4131226
operador:efecto operador 5	-2.11666667	0.5220153	-4.0547979	0.002305739	-3.27978929	-0.9535440
operador:efecto operador 6	-0.65000000	0.5220153	-1.2451742	0.241455469	-1.81312263	0.5131226

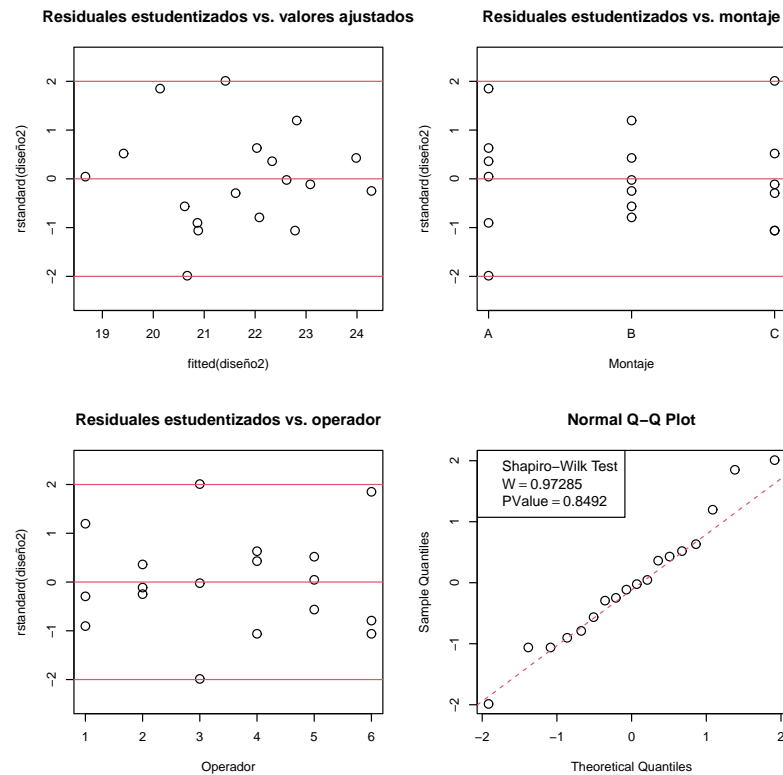


Figura 7.8: Gráficas de residuos internamente estudentizados Para análisis de supuestos

El programa R usado se presenta a continuación.

Código R 7.7.1. Programa R usado en Problema propuesto en Sección 7.7.

```
library(gmodels)
library(multcomp)
library(lsmmeans)
library(agricolae)

ensamble2=data.frame(operador=factor(rep(1:6,times=3)),montaje=factor(rep(c("A","B","C"),each=6)),
                      tiempo=scan())

20.2 22.6 19.2 22.5 18.7 21.5
23.7 24.1 22.6 24.3 20.2 21.5
21.4 23.0 22.9 22.0 19.8 20.1

attach(ensamble2)
layout(c(1,1,2,2))
plot(tiempo~operador+montaje,boxwex=0.5) #Boxplots según operadores y según montaje
win.graph()
```

```

plot.design(tiempo~operador+montaje) #Gráfico de medias

diseño2=aov(tiempo~operador+montaje) #Modelo ANOVA
anova(diseño2)

#INTERVALOS DE TUKEY PARA LAS DIFERENCIAS DE MEDIAS SEGÚN OPERADOR Y SEGÚN MONTAJE
TukeyHSD(diseño2,"operador",conf.level=0.95)
TukeyHSD(diseño2,"montaje",conf.level=0.95)

#GRÁFICOS DE INTERVALOS DE TUKEY
layout(cbind(c(1,1),c(2,2)))
plot(TukeyHSD(diseño2,"operador",conf.level = 0.95),cex.lab=0.8,las=1)
plot(TukeyHSD(diseño2,"montaje",conf.level = 0.95),cex.lab=0.8,las=1)

#OBTENIENDO EFECTOS DE OPERARIOS, RESULTADOS PARA TEST DE SIGNIFICANCIA Y SUS I.C DEL 95%
efect.operador1=fit.contrast(diseño2,"operador",rbind(":efecto operador 1"=c(5/6,-1/6,-1/6,-1/6,-1/6,-1/6)),
                             conf=0.95)
efect.operador2=fit.contrast(diseño2,"operador",rbind(":efecto operador 2"=c(-1/6,5/6,-1/6,-1/6,-1/6,-1/6)),
                             conf=0.95)
efect.operador3=fit.contrast(diseño2,"operador",rbind(":efecto operador 3"=c(-1/6,-1/6,5/6,-1/6,-1/6,-1/6)),
                             conf=0.95)
efect.operador4=fit.contrast(diseño2,"operador",rbind(":efecto operador 4"=c(-1/6,-1/6,-1/6,5/6,-1/6,-1/6)),
                             conf=0.95)
efect.operador5=fit.contrast(diseño2,"operador",rbind(":efecto operador 5"=c(-1/6,-1/6,-1/6,-1/6,5/6,-1/6)),
                             conf=0.95)
efect.operador6=fit.contrast(diseño2,"operador",rbind(":efecto operador 6"=c(-1/6,-1/6,-1/6,-1/6,-1/6,5/6)),
                             conf=0.95)
rbind(efect.operador1,efect.operador2,efect.operador3,efect.operador4,efect.operador5,efect.operador6)

#OBTENIENDO TEST DE NORMALIDAD SOBRE RESIDUALES INTERNAMENTE ESTUDENTIZADOS
shapiro.test(rstandard(diseño2))

#OBTENIENDO DE GRÁFICOS PARA VALIDACIÓN DE SUPUESTOS CON RESIDUOS INTERNAMENTE ESTUDENTIZADOS
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
plot(fitted(diseño2),rstandard(diseño2),ylim=c(-2.5,2.5),cex=1.5,
     main="Residuales estudentizados vs. valores ajustados")
abline(h=c(-2,0,2),col=2)

stripchart(rstandard(diseño2)~montaje,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Montaje",
           main="Residuales estudentizados vs. montaje")
abline(h=c(-2,0,2),col=2)

```

```
stripchart(rstandard(diseño2)~operador,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Operador",
          main="Residuales estudentizados vs. operador")
abline(h=c(-2,0,2),col=2)

qqnorm(rstandard(diseño2),cex=1.5)
qqline(rstandard(diseño2),col=2,lty=2)
legend("topleft",legend=c("Shapiro-Wilk Test",expression(W=0.97285),expression(PValue== 0.8492)),cex=1.1)

detach(ensamble2)
```

7.8. Diseños en cuadrados latinos DCL

Este diseño resulta cuando se tienen dos factores de bloqueo, el primero denominado factor de bloque columna y el segundo denominado factor de bloque fila. Se llama cuadrado porque tanto el número de niveles del factor de tratamientos como los niveles de los dos factores de bloqueo, son iguales, digamos r , y el término “latino” es debido a que los tratamientos son representados por letras latinas en mayúscula: A, B, C, etc. En un cuadrado latino, cada letra latina aparece sólo una vez por fila y por columna. Vea la Tabla 7.5.

Tabla 7.5: Cuadrado latino 4x4 inicial (izq.); sorteo de columnas 3, 2, 4, 1 (centro); sorteo de las filas 4, 2, 1, 3 (der.)

A	B	C	D	C	B	D	A	B	A	C	D
B	C	D	A	D	C	A	B	D	C	A	B
C	D	A	B	A	D	B	C	C	B	D	A
D	A	B	C	B	A	C	D	A	D	B	C

El cuadrado latino debe ser elegido antes de obtener los datos y luego se procede a aleatorizar las columnas y renglones (ver Tabla 7.5). Las letras latinas deben ser luego asignadas aleatoriamente entre los k tratamientos. En R, en la librería agricolae, se puede efectuar esto con la función `design.lsd()`. En el código R siguiente se ilustra el uso de esta función, suponiendo el experimento que se describe en la Sección 7.8.1, donde el factor de tratamientos es “marcas” de llantas con cuatro niveles, el

factor bloque fila es “Posición” de las llantas, con niveles Pos1, Pos2, Pos3 y Pos4, y el factor bloque columna es “Carro” con niveles Car1, Car2, Car3, Car4:

Código R 7.8.1. Programa R para crear y aleatorizar un plan experimental bajo un diseño cuadrado latino de 4x4.

```
marcas=1:4
diseñolat=design.lsd(trt=marcas,seed=23)
cuadlat=diseñolat$book
levels(cuadlat$row)=c("Pos1","Pos2","Pos3","Pos4")
levels(cuadlat$col)=c("Car1","Car2","Car3","Car4")
cuadlat
```

Salida R 7.8.1. Resultados del código R anterior.

```
> cuadlat
  plots row col marcas
1   101 Pos1 Car1     1
2   102 Pos1 Car2     3
3   103 Pos1 Car3     4
4   104 Pos1 Car4     2
5   201 Pos2 Car1     2
6   202 Pos2 Car2     4
7   203 Pos2 Car3     1
8   204 Pos2 Car4     3
9   301 Pos3 Car1     4
10  302 Pos3 Car2     2
11  303 Pos3 Car3     3
12  304 Pos3 Car4     1
13  401 Pos4 Car1     3
14  402 Pos4 Car2     1
15  403 Pos4 Car3     2
16  404 Pos4 Car4     4
```

En un diseño cuadrado latino se supone además que no hay interacciones entre el factor de tratamientos y los bloques y tampoco entre estos dos últimos, de modo que las diferencias entre niveles de tratamientos pueden generalizarse sobre filas y

columnas del diseño. Veremos mediante un ejemplo el modelo, inferencias y análisis en este tipo de diseños.

El modelo estadístico del DCL de dimensión $r \times r$, con un factor de efectos fijos y bloques no aleatorios (no hay selección aleatoria de los niveles de tratamientos ni de bloques) es

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \quad \text{con } \sum_{i=1}^r \alpha_i = \sum_{j=1}^r \beta_j = \sum_{k=1}^r \gamma_k = 0, \quad \varepsilon \stackrel{iid}{\sim} N(0, \sigma^2), \quad (7.17)$$

en donde α_i denota el efecto fijo del factor de tratamientos en su i -ésimo nivel, β_j corresponde al efecto fijo del factor bloque fila en su j -ésimo nivel y γ_k es el efecto fijo del factor bloque columnas en su k -ésimo nivel, con $i, j, k = 1, 2, \dots, r$. La tabla ANOVA de este diseño se muestra en la Tabla 7.6.

Para la eficiencia relativa, como ahora hay dos factores de bloqueo, se compara el DCL con el DBCA y para esto se puede proceder mediante dos enfoques [3]: tomando como bloques el factor de bloqueo fila o tomando como bloques el factor de bloqueo columna.

1. Considerando como bloques el factor de bloqueo fila: En este caso se comparan los diseños con y sin el factor bloques columnas en presencia del factor bloques filas. Para esto se aproxima el MSE del DBCA mediante, la siguiente ecuación

$$MSE(DBCA) \approx \frac{f_c \times MSC + (f_a + f_e)MSE}{f_c + f_a + f_e}, \quad (7.18)$$

donde MSC y MSE son los cuadrados medios del factor bloques columnas y del error en el DCL, respectivamente, y f_c , f_a y f_e corresponden a los grados de libertad del factor bloques columna, del factor de tratamientos y del error en el DCL, respectivamente. Entonces la eficiencia relativa del DCL respecto al DBCA, corregido por grados de libertad, corresponde a

$$EER = \frac{(f_1 + 1)(f_2 + 3)}{(f_1 + 3)(f_2 + 1)} \times \frac{MSE(DBCA)}{MSE}, \quad (7.19)$$

en donde $f_1 = f_e$, los grados de libertad del error en el DCL, y f_2 los grados de libertad del error en el DBCA usando sólo como factor de bloques al factor bloque fila.

Tabla 7.6: ANOVA en un DCL $r \times r$

Fuente	g.l	SC	CM	CME	F ₀	Valor P
Factor	$r - 1$	SSA	$MSA = \frac{SSA}{r - 1}$	$E[MSA] = \sigma^2 + \frac{r \sum_{i=1}^r \alpha_i^2}{r - 1}$	$\frac{MSA}{MSE}$	$P(f_{r-1, (r-1)(r-2)} > F_0)$
Filas	$r - 1$	SSF	$MSF = \frac{SSF}{r - 1}$	$E[MSF] = \sigma^2 + \frac{r \sum_{j=1}^r \beta_j^2}{r - 1}$		
Columnas	$r - 1$	SSC	$MSC = \frac{SSC}{r - 1}$	$E[MSC] = \sigma^2 + \frac{r \sum_{k=1}^r \gamma_k^2}{r - 1}$		
Error	$(r - 1)(r - 2)$	SSE	$MSE = \frac{SSE}{(r - 1)(r - 2)}$	$E[MSE] = \sigma^2$		
Total	$r^2 - 1$	SST				
Fórmulas						
$SSA = r \sum_{i=1}^r (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = r \sum_{i=1}^r \bar{Y}_{i\bullet\bullet}^2 - r^2 \bar{Y}_{\bullet\bullet\bullet}^2$ $SSF = r \sum_{j=1}^r (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = r \sum_{j=1}^r \bar{Y}_{\bullet j\bullet}^2 - r^2 \bar{Y}_{\bullet\bullet\bullet}^2$ $SSC = r \sum_{k=1}^r (\bar{Y}_{\bullet\bullet k} - \bar{Y}_{\bullet\bullet\bullet})^2 = r \sum_{k=1}^r \bar{Y}_{\bullet\bullet k}^2 - r^2 \bar{Y}_{\bullet\bullet\bullet}^2$ $SST = \text{SUMA}_{i,j,k} (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \text{SUMA}_{i,j,k} (Y_{ijk}^2) - r^2 \bar{Y}_{\bullet\bullet\bullet}^2$ $SSE = SST - SSA - SSF - SSC$ donde, SUMA _{i,j,k} indica que se suma sobre todas las r^2 celdas del DCL $\bar{Y}_{i\bullet\bullet}$ Media muestral en el nivel i del factor de tratamientos en toda la tabla del diseño $\bar{Y}_{\bullet j\bullet}$ Media muestral en el nivel j del factor de bloque filas $\bar{Y}_{\bullet\bullet k}$ Media muestral en el nivel k del factor de bloque columnas $\bar{Y}_{\bullet\bullet\bullet}$ Media muestral de todas las r^2 observaciones.						

2. Considerando como bloques al factor de bloqueo columna: Se comparan los diseños con y sin la presencia del factor de bloques fila, en presencia del factor de bloques columna. En este caso se aproxima el MSE del DBCA mediante, la siguiente ecuación

$$MSE(\text{DBCA}) \approx \frac{f_f \times MSF + (f_a + f_e)MSE}{f_f + f_a + f_e}, \quad (7.20)$$

donde MSF y MSE son los cuadrados medios del factor bloques fila y del error

en el DCL, respectivamente, y f_f , f_a y f_e corresponden a los grados de libertad del factor bloques fila, del factor de tratamientos y del error en el DCL, respectivamente. Entonces la eficiencia relativa del DCL respecto al DBCA, corregido por grados de libertad, en su forma es la misma ecuación (7.19), sin embargo con $MS(DBCA)$ calculado mediante ecuación (7.20) y con $f_1 = f_e$, los grados de libertad del error en el DCL, y f_2 los grados de libertad del error en el DBCA usando sólo como factor de bloques al factor bloque columna.

7.8.1. Ejemplo [4]

Una empresa de transportes terrestre estaba interesada en determinar cuál entre cuatro marcas de llantas tiene mayor duración, medida ésta en términos del desgaste. Para ello se planeó un experimento con un diseño cuadrado latino en el que se compararon las cuatro marcas de llantas sometiéndolas a una prueba de 32000 kms. de recorrido, utilizando cuatro diferentes tipos de vehículos y las cuatro posiciones posibles de las llantas en el vehículo. Así el factor de interés es el tipo de llanta o marca, y se controlan dos factores de bloqueo: el tipo de carro y la posición de la llanta en el carro. Se controlan estos dos factores porque por experiencia se sabe que el tipo de carro (factor bloques columna) y la posición de la llanta (factor bloques fila) tienen efectos en el desgaste de la misma.

La elección del cuadrado latino se hizo antes de obtener los datos. Para ello, a partir de un cuadrado latino inicial, se aleatorizaron las columnas y las filas (se asignaron aleatoriamente para cada factor de bloqueo sus respectivas etiquetas de niveles a sus bloques); después las diferentes marcas de llanta se asignaron de manera aleatoria a las letras latinas que denotan los niveles del factor marca. Las pruebas se hicieron al mismo tiempo con choferes a quienes se les instruyó para que manejaran de manera similar, sobre el mismo terreno para los cuatro carros. Al hacer las pruebas de los cuatro carros al mismo tiempo se evitó el efecto del ambiente en el desgaste; así mismo, el conductor y el tipo de terreno podrían influir, pero se consideró suficiente mantenerlos lo más homogéneo posible durante el experimento. Los datos observa-

dos se muestran a continuación. Se midió la diferencia máxima entre el grosor de la llanta nueva y el grosor de la llanta después de haber recorrido los 32000 kms. Obviamente a mayor diferencia en grosor, mayor desgaste. Las unidades de medición son milésimas de pulgadas:

Posición	Carro			
	1	2	3	4
1	C=12	D=11	A=13	B=8
2	B=14	C=12	D=11	A=13
3	A=17	B=14	C=10	D=9
4	D=13	A=14	B=13	C=9

Se desea determinar si existen diferencias entre las cuatro marcas de llantas y cómo se diferencian, además elegir la mejor marca.

Primero veamos los boxplots por Marca, por carro y por posición de la llanta. De la Figura 7.9 se establece que las marcas si difieren en cuanto al desgaste. Las marcas de menor desgaste son la C y D, aparentemente con estas dos se obtienen resultados similares. La marca A es la de mayor desgaste. A nivel de marcas la dispersión es aproximadamente similar (aunque con la marca 2 es un poco mayor). Los datos comparados según carros exhiben también diferencias, los carros pueden ordenarse de mayor a menor desgaste así, 1, 2, 3, y 4. Los datos según carros exhiben aproximadamente la misma dispersión. Por el contrario, la posición de la llanta no ofrece mucha diferencias en los promedios. Los desgastes en las posiciones 2 a 4 en promedio son prácticamente iguales; la posición 1 ofrece un desgaste promedio un poco inferior a las otras posiciones, pero es muy probable que estadísticamente no se diferencie de las otras. Pero hay que tener en cuenta que las pruebas se hicieron en un circuito, las vueltas siempre eran en el mismo sentido, lo que puede generar cierto desgaste mayor e una posición determinada. Además, la varianza parece no ser igual en particular los datos en la posición 3 ofrecen una dispersión mucho más grande que en las otras tres posiciones.

El modelo estadístico en este caso es

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

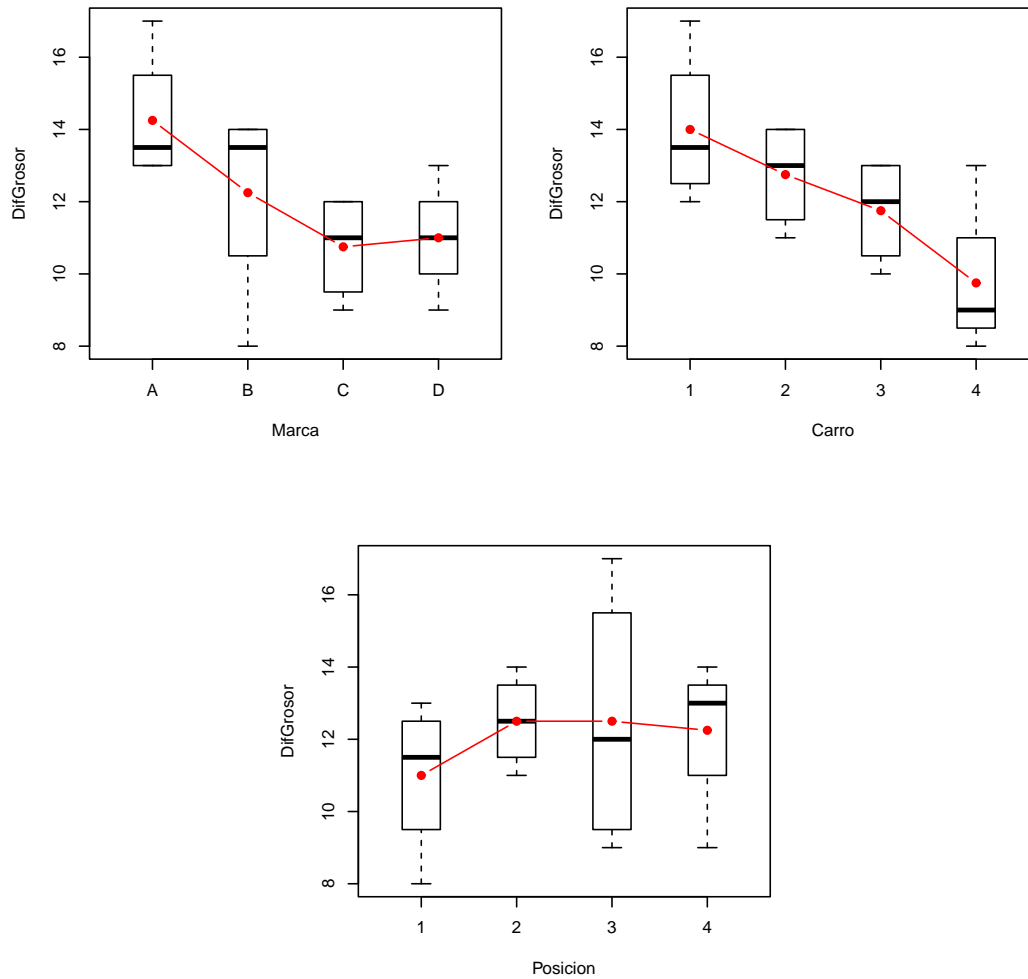


Figura 7.9: Boxplots datos Ejemplo 7.8.1

con $i = 1, 2, 3, 4$ para denotar las marcas A, B, C y D respectivamente, y $j = 1, 2, 3, 4$ posiciones y $k = 1, 2, 3, 4$ para denotar los carros. representa el efecto de la marca i -ésima, es el efecto de la j -ésima posición y el efecto de k -ésimo carro sobre el promedio de la diferencia máxima del grosor. Restricciones y supuestos son $\sum_{i=1}^4 \alpha_i = \sum_{j=1}^4 \beta_j = \sum_{k=1}^4 \gamma_k = 0$ los errores $\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$. A continuación se presenta la tabla ANOVA:

Salida R 7.8.2. *Tabla ANOVA en el DCL aplicado para comparar las Marcas de llantas.*

```

Analysis of Variance Table

Response: DifGrosor

      Df Sum Sq Mean Sq F value    Pr(>F)
Marca    3 30.688 10.2292 11.4186 0.006825 **
Posicion  3  6.187  2.0625  2.3023 0.176947
Carro     3 38.688 12.8958 14.3953 0.003784 **
Residuals 6  5.375  0.8958
---

```

La hipótesis de interés es $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$, y con un valor P de 0.0068 se concluye a un nivel de significancia de 0.05 que el desgaste promedio no es igual para todas las marcas. Si fuese válido realizar el test de significancia de los efectos del factor de bloque posición (el factor bloque filas) llegaríamos a la conclusión de que éste no es significativo. Realicemos el cálculo de la eficiencia relativa del DCL respecto a un DBCA si el único factor de bloque fueran los carros (vea las ecuaciones (7.19) y (7.20)):

$$ER = \frac{(6+1)(9+3)}{(9+1)(6+3)} \times \frac{3 \times 2.0625 + (3+6) \times 0.89583333}{(3+3+6) \times 0.89583333} = 1.23$$

con lo cual podemos decir que a pesar que es mayor que 1, el DCL no fue mejor que un DBCA usando como factor de bloqueo sólo el tipo de carro. Sin embargo, no podemos simplemente eliminar del modelo al factor de bloqueo posición y recalcular la ANOVA, dado que no se realizó un diseño de DBCA sino un DCL.

Veamos que resulta del procedimiento de Tukey en la comparación del factor Marca de llantas (el factor de tratamientos, el cuál es el único sobre el que interesa inferir sobre sus efectos y sus medias):

Salida R 7.8.3. *I.C y comparaciones de Tukey para las diferencias de medias según factor Marca de llantas.*

```
> TukeyHSD(modelo, "Marca", conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = DifGrosor ~ Marca + Posicion + Carro)
$Marca
      diff      lwr      upr    p adj
B-A -2.00 -4.316805  0.3168049 0.0872429
C-A -3.50 -5.816805 -1.1831951 0.0078229
D-A -3.25 -5.566805 -0.9331951 0.0112170
C-B -1.50 -3.816805  0.8168049 0.2144635
D-B -1.25 -3.566805  1.0668049 0.3315951
D-C  0.25 -2.066805  2.5668049 0.9805997

> HSD.test(modelo, "Marca", group=TRUE, console=TRUE) #Comparaciones de Tukey
Study: modelo ~ "Marca"
HSD Test for DifGrosor
Mean Square Error:  0.8958333
Marca, means
  DifGrosor      std r Min Max
A      14.25 1.892969 4   13  17
B      12.25 2.872281 4    8  14
C      10.75 1.500000 4    9  12
D      11.00 1.632993 4    9  13

Alpha: 0.05 ; DF Error: 6
Critical Value of Studentized Range: 4.895599
Honestly Significant Difference: 2.316805
Means with the same letter are not significantly different.
Groups, Treatments and means
a      A      14.25
ab     B      12.25
b      D      11
b      C      10.75
```

El método de comparación de Tukey arroja dos grupos de medias traslapados: Grupo A las marcas A y B; grupo B las marcas D y C, las cuales pueden considerarse estadísticamente iguales. A la empresa podría recomendarse cualquiera de las llantas

marcas C ó D, si son de igual costo.

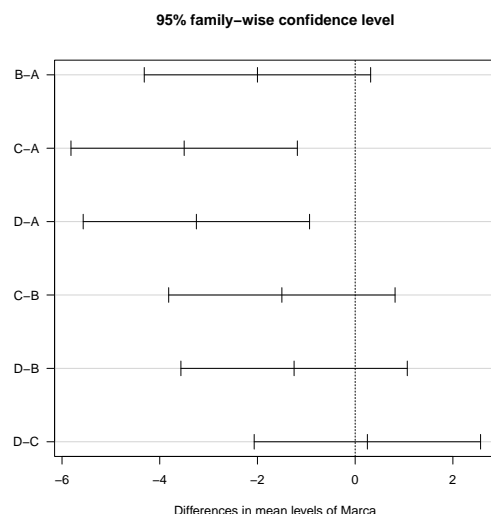


Figura 7.10: Gráfica de los I.C de Tukey para las diferencias entre medias del factor Marca, en ejemplo 7.8.1.

Para la verificación de los supuestos sobre los errores se considerarán las gráficas en la Figura 7.11. Del gráfico de q-q plot, el supuesto de normalidad muestra desvío ¿severo? por otro lado el supuesto de varianza constante pareciera no cumplirse a nivel de marcas, pues en la marca B existe una dispersión mayor de los residuales en relación a las otras marcas, mientras que en la marca D hay menor dispersión. A nivel de carro y posición se observa menor dispersión respectivamente, en los niveles 1, pero en general puede decirse que la falta de homogeneidad de varianza es moderada.

A continuación se dan los $\hat{\alpha}_i$ y los I.C respectivos. Se deja al lector su interpretación (OJO: α_i es estimado mediante contraste de medias de tratamientos).

Salida R 7.8.4. Efectos estimados de la Marca de llantas comparadas en el DCL aplicado.

	Estimate	Std. Error	t value	Pr(> t)	lower CI	upper CI
Marca:efecto Marca A	2.1875	0.4098399	5.3374500	0.001765566	1.1846579	3.19034213
Marca:efecto Marca B	0.1875	0.4098399	0.4574957	0.663413192	-0.8153421	1.19034213
Marca:efecto Marca C	-1.3125	0.4098399	-3.2024700	0.018542202	-2.3153421	-0.30965787
Marca:efecto Marca D	-1.0625	0.4098399	-2.5924757	0.041075111	-2.0653421	-0.05965787

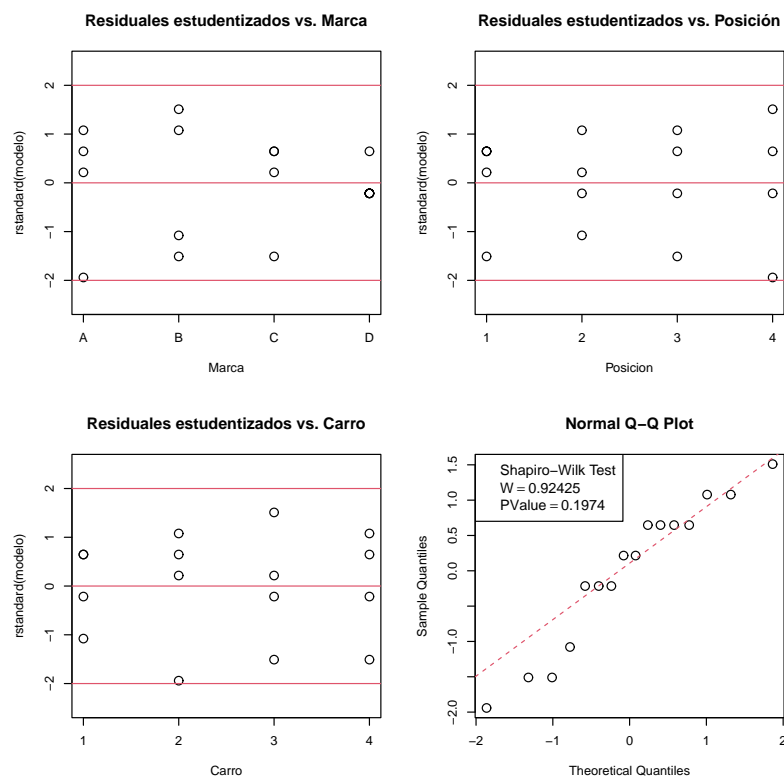


Figura 7.11: Gráfica de residuos internamente estudentizados y qqplot normal, en ejemplo 7.8.1.

Código R 7.8.2. Programa R usado e Ejemplo 7.8.1.

```
library(gmodels)
library(multcomp)
library(lsmeans)
library(agricolae)

datos=data.frame(scan(what=list(Posicion="",Carro="",Marca="",DifGrosor=0)))

1 1 C 12
1 2 D 11
1 3 A 13
1 4 B 8
2 1 B 14
2 2 C 12
2 3 D 11
2 4 A 13
3 1 A 17
```

```

3 2 B 14
3 3 C 10
3 4 D 9
4 1 D 13
4 2 A 14
4 3 B 13
4 4 C 9

```

```
datos
```

```
attach(datos)
```

```
#MEDIAS POR NIVELES DE TRATAMIENTOS Y DE BLOQUES
```

```
mediasM=sapply(split(DifGrosor,Marca),mean)
```

```
mediasC=sapply(split(DifGrosor,Carro),mean)
```

```
mediasP=sapply(split(DifGrosor,Posicion),mean)
```

```
#BOXPLOTS COMPARATIVOS
```

```
layout(rbind(c(1,1,2,2),c(0,3,3,0)))
```

```
boxplot(DifGrosor~Marca,boxwex=0.4,xlab="Marca",ylab="DifGrosor")
```

```
lines(1:4,mediasM,col=2,type="b",pch=19)
```

```
boxplot(DifGrosor~Carro,boxwex=0.4,xlab="Carro",ylab="DifGrosor")
```

```
lines(1:4,mediasC,col=2,type="b",pch=19)
```

```
boxplot(DifGrosor~Posicion,boxwex=0.4,xlab="Posicion",ylab="DifGrosor")
```

```
lines(1:4,mediasP,col=2,type="b",pch=19)
```

```
#AJUSTE MODELO ANOVA
```

```
modelo=aov(DifGrosor~Marca+Posicion+Carro)
```

```
anova(modelo)
```

```
#INTERVALOS DE TUKEY PARA LAS DIFERENCIAS DE MEDIAS SEGÚN MARCA LLANTAS
```

```
TukeyHSD(modelo,"Marca",conf.level=0.95)
```

```
HSD.test(modelo,"Marca", group=TRUE,console=TRUE) #Comparaciones de Tukey
```

```
#GRÁFICOS DE INTERVALOS DE TUKEY
```

```
plot(TukeyHSD(modelo,"Marca",conf.level = 0.95),cex.lab=0.8,las=1)
```

```
#OBTENIENDO EFECTOS DE TRATAMIENTOS, RESULTADOS PARA TEST DE SIGNIFICANCIA Y SUS I.C DEL 95%
```

```
efect.MarcaA=fit.contrast(modelo,"Marca",rbind(":efecto Marca A"=c(3/4,-1/4,-1/4,-1/4)),conf=0.95)
```

```
efect.MarcaB=fit.contrast(modelo,"Marca",rbind(":efecto Marca B"=c(-1/4,3/4,-1/4,-1/4)),conf=0.95)
```

```
efect.MarcaC=fit.contrast(modelo,"Marca",rbind(":efecto Marca C"=c(-1/4,-1/4,3/4,-1/4)),conf=0.95)
```

```
efect.MarcaD=fit.contrast(modelo,"Marca",rbind(":efecto Marca D"=c(-1/4,-1/4,-1/4,3/4)),conf=0.95)
```

```
rbind(efect.MarcaA,efect.MarcaB,efect.MarcaC,efect.MarcaD)

#OBTENIENDO TEST DE NORMALIDAD SOBRE RESIDUALES INTERNAMENTE ESTUDENTIZADOS
shapiro.test(rstandard(modelo))

#OBTENIENDO DE GRÁFICOS PARA VALIDACIÓN DE SUPUESTOS CON RESIDUOS INTERNAMENTE ESTUDENTIZADOS
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
stripchart(rstandard(modelo)~Marca,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Marca",
           main="Residuales estudentizados vs. Marca")
abline(h=c(-2,0,2),col=2)

stripchart(rstandard(modelo)~Posicion,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Posición",
           main="Residuales estudentizados vs. Posicion")
abline(h=c(-2,0,2),col=2)

stripchart(rstandard(modelo)~Carro,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1.5,xlab="Carro",
           main="Residuales estudentizados vs. Carro")
abline(h=c(-2,0,2),col=2)

qqnorm(rstandard(modelo),cex=1.5)
qqline(rstandard(modelo),col=2,lty=2)
legend("topleft",legend=c("Shapiro-Wilk Test",expression(W=0.92425),expression(PValue== 0.1974)),cex=1.1)

detach(datos)
```

Parte III

DISEÑOS COMPLETAMENTE ALEATORIZADOS CON TRATAMIENTOS FACTORIALES

Capítulo 8

Experimentos factoriales

8.1. Introducción

Los experimentos factoriales son aquellos en los cuales el interés se centra en el estudio de los efectos individuales y de interacción de las combinaciones de dos o más factores sobre una variable respuesta. En la práctica cuando el **número de factores a considerar es mayor que 3**, se suelen utilizar experimentos 2^k (k factores con dos niveles cada uno), experimentos 3^k (k factores con tres niveles cada uno), experimentos factoriales fraccionados, etc. Con el diseño factorial completo se corren aleatoriamente en el proceso todas las posibles combinaciones que pueden formarse con los niveles seleccionados de los factores de interés.

Definición 1: Llamaremos **tratamiento** a cada combinación posible entre los niveles de los factores. Así, para el caso de dos factores, se tendrá un total de ab posibles tratamientos. Por ejemplo, sea $a = 3$ y $b = 3$, entonces hay nueve tratamientos posibles como se ilustra en la Tabla 8.1.

Definición 2: Llamaremos **efecto de un factor** al cambio observado en la variable respuesta debido a un cambio de nivel de tal factor.

Definición 3: Los **efectos principales** denotan los cambios en la respuesta promedio debidos a la acción individual de cada factor.

Definición 4: Dos factores **interactúan** significativamente sobre la variable respuesta cuando el efecto de un factor depende del nivel en que esté el otro. Asociados a este fenómeno aparecen los **efectos de interacción** que miden la magnitud y dirección de

tales interacciones.

Tabla 8.1: Tratamientos en un experimento con dos factores A y B. cada combinación $A_i B_j$, $i, j = 1, 2, 3$, define un tratamiento, entonces hay 9 posibles tratamientos

	Niveles de B		
Niveles de A	B_1	B_2	B_3
A_1	$A_1 B_1$	$A_1 B_2$	$A_1 B_3$
A_2	$A_2 B_1$	$A_2 B_2$	$A_2 B_3$
A_3	$A_3 B_1$	$A_3 B_2$	$A_3 B_3$

8.1.1. Experimentos factoriales vs. la estrategia de mover un factor a la vez

Los experimentos factoriales son más eficientes que el tradicional experimento de “mover un factor a la vez”, que utilizan algunas personas cuando no tienen conocimiento del diseño de experimentos. Probar un factor a la vez significa probar primeros niveles de un factor mientras se mantienen fijos los otros factores del estudio. Una vez se encuentre el mejor nivel para el factor que se ha movido, éste es fijado en tal nivel, y luego se elige otro de los factores del estudio para moverlo, en tanto que los demás están fijos, y así sucesivamente se procede con cada factor. Sin embargo cuando existen interacciones significativas entre dos o más factores, este método lleva a conclusiones erróneas, alejadas de las condiciones óptimas de operación del proceso, en cambio con experimentos factoriales, probando las diferentes combinaciones de al menos dos niveles de los factores de estudio, es posible encontrar desde el primer diseño mejores condiciones de operación, cercanas al punto óptimo con menos corridas experimentales y mayor eficacia.

La experimentación empírica en apariencia más simple y sencilla resulta cara y poco eficaz, lo cual se agrava a medida que se tienen más factores de estudio.

8.1.2. Ventajas de los diseños factoriales

1. Se pueden aumentar para formar diseños compuestos en caso que se requiera una exploración más completa. Por ejemplo, si el comportamiento de la variable

respuesta no es lineal en algunos de factores controlados, es útil aumentar el diseño.

2. Se pueden correr **fracciones de los diseños factoriales** (sólo se corren algunos de los posibles tratamientos resultantes de las combinaciones de niveles de los factores de estudio), las cuales son de gran utilidad en las primeras etapas de una investigación con muchos factores, cuando interesa descartar de manera económica los menos importantes, antes de hacer un estudio más detallado con los factores que si son importantes.
3. Pueden utilizarse en combinación con diseños de bloques en situaciones en las que no puede correrse todo el diseño factorial completo bajo las mismas condiciones o circunstancias. Por ejemplo, cuando cada lote de material sólo alcanza para correr la mitad del experimento, éste se puede correr en dos bloques (dos lotes), lo que implica repartir las pruebas en los dos lotes de la manera más conveniente posible.

Nota 8.1. *Consideraremos el caso de diseños factoriales balanceados, es decir, igual número de replicaciones en cada tratamiento.*

8.2. Diseños factoriales con dos factores

Denotaremos a los factores por las letras A y B y sus números de niveles por a y b respectivamente.

8.2.1. Tipos de experimentos factoriales con dos factores

1. **De efectos fijos:** Los a niveles del factor A como los b niveles del factor B son definidos a propósito por el investigador.
2. **De efectos aleatorios:** Los a niveles del factor A como los b niveles del factor B son elegidos aleatoriamente dentro de una población de posibles niveles, respectivamente.

3. **De efectos mixtos:** Uno de los dos factores es de efectos fijos en tanto que el otro es de efectos aleatorios. Dos posibles modelos pueden postularse en este caso: **modelo mixto restringido** y **modelo mixto no restringido**.

Las estructuras de diseño pueden ser tanto completamente aleatorizadas como en bloques y en estos últimos, bajo el supuesto de que los factores de bloque no interactúan entre sí ni con los dos factores de estudio.

En cualquiera de los anteriores casos se debe considerar el fenómeno conocido como **interacción de los factores**, es decir, los modelos deben iniciarse considerando la existencia de interacción y luego a través de la evidencia muestral, determinar si dicha interacción es o no significativa. Sólo si en cada tratamiento se toman réplicas ($n > 1$ observaciones por tratamiento), será posible estimar y evaluar la interacción entre los factores.

8.3. Modelo factorial dos factores efectos fijos con interacción, en un DCA

8.3.1. Modelo ANOVA

Consideraremos que a cada **tratamiento** se asignan en forma completamente aleatorizada, n unidades experimentales obteniéndose por tanto de cada tratamiento n observaciones de la variable respuesta. En total habrán $N = abn$ observaciones de la variable respuesta. Un tabla de entrada de datos del experimento será como se presenta en la Tabla 8.2, y el modelo con interacción, para la respuesta k -ésima en el tratamiento $A_i B_j$ (nivel i del factor A combinado con nivel j del factor B), corresponde a:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \text{ con } \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (8.1)$$

$$i = 1, 2, \dots, a, \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, n, \quad \sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$$

donde,

- Y_{ijk} es la respuesta en la k -ésima réplica del tratamiento $A_i B_j$:

- μ es la media global sin distinguir tratamientos;
- α_i es el efecto fijo (efecto principal) del nivel i del factor A;
- β_j es el efecto fijo (efecto principal) del nivel j del factor B;
- $(\alpha\beta)_{ij}$ es el efecto fijo de la interacción de los factores A y B en sus niveles i, j , respectivamente.

Cuando ambos factores son cuantitativos, el comportamiento de la variable respuesta se puede describir con el modelo de regresión asociado dado por ([4]):

$$Y_{ijk} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{1,2} X_1 X_2 + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (8.2)$$

Donde β_0 es la media global, β_1 es el efecto del factor $X_1=A$, β_2 es el efecto del factor $X_2=B$ y $\beta_{1,2}$ representa el efecto de interacción $X_1 X_2=AB$. Se recomienda que las variables X_i se ajusten en la escala codificada: $-1, +1$ para dos niveles (con -1 para el nivel de valor más bajo y $+1$ para el nivel de valor mayor); o bien $-1, 0, +1$ para tres niveles (con -1 para el nivel de valor más bajo, 0 para el nivel medio y $+1$ para el nivel de valor más alto), esta codificación se hace restando a cada nivel original la media de los valores de los niveles del respectivo factor y dividiendo por la distancia entre niveles extremos, es decir dividiendo por 0.5 (Nivel Alto – Nivel Bajo) (esto implica niveles a igual distancia). En el ajuste y diagnóstico de este modelo se utilizan las técnicas de regresión múltiple. Cabe aclarar sobre los modelos en la ecuación (8.1) y (8.2) que aunque los correspondientes parámetros miden ambos los mismos efectos, no lo hacen en la misma escala. Los β_i representan la pendiente de la superficie de respuesta que describe el modelo en (8.2) y no las desviaciones respecto a la media global como sí lo hacen los α_i y β_j del modelo en la ecuación (8.1). Sin embargo al hacer las pruebas estadísticas la significancia de los efectos correspondientes definidos por cualquiera de los dos modelos debe ser igual. En algunos casos, como en la familia de factoriales 2^k , es fácil establecer la relación entre los dos grupos de parámetros. En cuanto a interpretación, los dos modelos predicen la misma respuesta sobre los puntos experimentales (es decir las combinaciones de niveles o tratamientos); pero

el segundo modelo puede usarse para predecir el comportamiento de la respuesta en cualquier punto dentro de los rango de prueba, formando de esta manera una superficie de respuesta. Si los factores tienen al menos tres niveles de prueba, el modelo de efectos en la ecuación (8.1) permite determinar el mejor tratamiento, mientras que el modelo en (8.2) es útil para encontrar el punto óptimo de operación.

En un DCA balanceado de dos factores, donde cada tratamiento $A_i B_j$ es replicado n veces, la tabla de datos luce como se muestra en Tabla 8.2.

Tabla 8.2: Tabla de datos experimentales de un DOE en un DCA balanceado, dos factores

	Niveles Factor B						
Niveles Factor A	B_1	B_2	...	B_j	...	B_b	Medias Factor A
A_1	Y_{111}	Y_{121}	...	Y_{1j1}	...	Y_{1b1}	$\bar{Y}_{1\bullet\bullet}$
	Y_{112}	Y_{122}	...	Y_{1j2}	...	Y_{1b2}	
	Y_{113}	Y_{123}	...	Y_{1j3}	...	Y_{1b3}	
	\vdots	\vdots	...	\vdots	...	\vdots	
	Y_{11n}	Y_{12n}	...	Y_{1jn}	...	Y_{1bn}	
A_2	Y_{211}	Y_{221}	...	Y_{2j1}	...	Y_{2b1}	$\bar{Y}_{2\bullet\bullet}$
	Y_{212}	Y_{222}	...	Y_{2j2}	...	Y_{2b2}	
	Y_{213}	Y_{223}	...	Y_{2j3}	...	Y_{2b3}	
	\vdots	\vdots	...	\vdots	...	\vdots	
	Y_{21n}	Y_{22n}	...	Y_{2jn}	...	Y_{2bn}	
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_i	Y_{i11}	Y_{i21}	...	Y_{ij1}	...	Y_{ib1}	$\bar{Y}_{i\bullet\bullet}$
	Y_{i12}	Y_{i22}	...	Y_{ij2}	...	Y_{ib2}	
	Y_{i13}	Y_{i23}	...	Y_{ij3}	...	Y_{ib3}	
	\vdots	\vdots	...	\vdots	...	\vdots	
	Y_{i1n}	Y_{i2n}	...	Y_{ijn}	...	Y_{ibn}	
\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_a	Y_{a11}	Y_{a21}	...	Y_{aj1}	...	Y_{ab1}	$\bar{Y}_{a\bullet\bullet}$
	Y_{a12}	Y_{a22}	...	Y_{aj2}	...	Y_{ab2}	
	Y_{a13}	Y_{a23}	...	Y_{aj3}	...	Y_{ab3}	
	\vdots	\vdots	...	\vdots	...	\vdots	
	Y_{a1n}	Y_{a2n}	...	Y_{ajn}	...	Y_{abn}	
Medias Factor B	$\bar{Y}_{\bullet 1\bullet}$	$\bar{Y}_{\bullet 2\bullet}$...	$\bar{Y}_{\bullet j\bullet}$...	$\bar{Y}_{\bullet b\bullet}$	Media global $\bar{Y}_{\bullet\bullet\bullet}$

en donde,

- Y_{ijk} es la respuesta en la k -ésima réplica del tratamiento A_iB_j
- $\bar{Y}_{i..}$ es la media muestral de las nb réplicas en el nivel i del factor A, es decir

$$\bar{Y}_{i..} = \frac{1}{nb} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij.}$$
- $\bar{Y}_{.j.}$ es la media muestral de las na réplicas en el nivel j del factor B, es decir,

$$\bar{Y}_{.j.} = \frac{1}{na} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij.}$$
- con $\bar{Y}_{ij.} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, la media muestral en el tratamiento A_iB_j ,
- $\bar{Y}_{...} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$, la media muestral global de las abn observaciones.

8.3.2. Parámetros de interés

- Efectos de tratamientos: $\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$, que representa el efecto del tratamiento A_iB_j sobre la respuesta promedio global. Note que $\sum_{i=1}^a \sum_{j=1}^b \tau_{ij} = 0$.
- Medias de tratamientos: $\mu_{ij} = \mu + \tau_{ij}$, de donde $\tau_{ij} = \mu_{ij} - \mu$.
- Media del nivel i del factor A: $\mu_{i.} = \mu + \alpha_i$, de donde $\alpha_i = \mu_{i.} - \mu$. Note que

$$\mu_{i.} = \frac{1}{b} \sum_{j=1}^b \mu_{ij}$$
- Media del nivel j del factor B: $\mu_{.j} = \mu + \beta_j$, de donde $\beta_j = \mu_{.j} - \mu$. Note que

$$\mu_{.j} = \frac{1}{a} \sum_{i=1}^a \mu_{ij}$$
- Media global: Puede ser expresada por cualquiera de las siguientes formas: $\mu = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \mu_{ij}$, $\mu = \frac{1}{a} \sum_{i=1}^a \mu_{i.}$, $\mu = \frac{1}{b} \sum_{j=1}^b \mu_{.j}$
- Efectos de interacción: $(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$

8.3.3. Significado de la interacción

Interacción entre factores A y B significa que las diferencias entre las medias de tratamientos según los niveles de un factor varían (bien sea en dimensión y/o signo) de nivel a nivel del otro factor. Considere por ejemplo un estudio simple de dos factores en el cual los efectos de sexo (factor A) y edad (factor B) sobre el aprendizaje de

una tarea son de interés, siendo la variable respuesta el tiempo de aprendizaje en minutos. Los datos en las siguientes tablas corresponden a medias poblacionales por tratamientos y niveles de cada factor (μ_{ij} en las celdas, $\mu_{i\bullet}$ en las filas y $\mu_{\bullet j}$ en las columnas):

Tabla 1	Edad			$\mu_{i\bullet}$
género	joven	medio	viejo	
hombre	11	13	18	14.00
mujer	7	9	14	10.00
$\mu_{\bullet j}$	9	11	16	12.00

Tabla 3	edad			$\mu_{i\bullet}$
género	joven	medio	viejo	
hombre	9	12	14	11.67
mujer	7	10	16	11.00
$\mu_{\bullet j}$	8	11	15	11.33

Tabla 2	edad			$\mu_{i\bullet}$
género	joven	medio	viejo	
hombre	9	12	18.00	13.00
mujer	9	10	14.00	11.00
$\mu_{\bullet j}$	9	11	16.00	12.00

Tabla 4	edad			$\mu_{i\bullet}$
género	joven	medio	viejo	
hombre	9.75	12	17.25	13.00
mujer	8.25	10	14.75	11.00
$\mu_{\bullet j}$	9.00	11	16.00	12.00

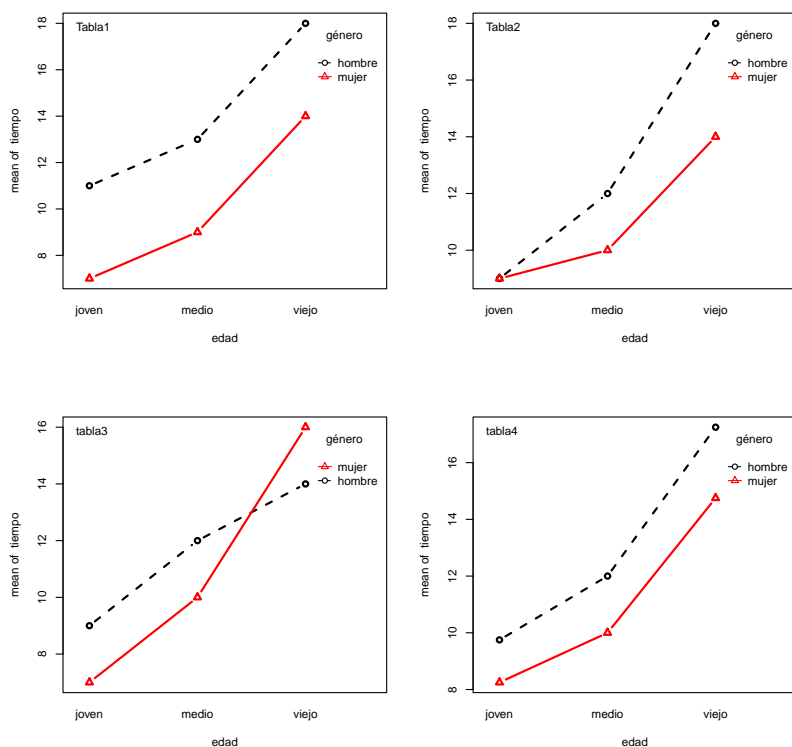


Figura 8.1: Gráficos de perfiles de medias.

Los gráficos anteriores son conocidos como **gráficos de perfiles de medias** en los cuales se presentan las medias de tratamiento en cada uno de los niveles de uno de los factores fijando los niveles del otro factor ubicado en el eje horizontal. En el ejemplo se han graficado las medias de tratamiento según Género en cada nivel de edad, así las medias μ_{1j} (con Género=hombre) son unidas por la línea discontinua en negro, en tanto que las medias μ_{2j} (con Género=mujer) son unidas por la línea continua en rojo.

Sea $i = 1$ para género hombre, $i = 2$ para Género=mujer, $j = 1$ para edad =joven, $j = 2$ para edad=medio y $j = 3$ para edad=viejo. La **Tabla 1** y su gráfica correspondiente, presentan el caso en el cual los factores no interactúan, pues note que las diferencias en las medias μ_{ij} de género (hombre - mujer, es decir $\mu_{1j} - \mu_{2j}$) en cada uno de los niveles de edad, es igual a 4 (también podría analizarse las diferencias en las medias μ_{ij} de edad en cada nivel de género y comprobarse que $\mu_{i1} - \mu_{i2} = -2$, $\mu_{i1} - \mu_{i3} = -7$, y $\mu_{i2} - \mu_{i3} = -5$ para $i = 1, 2$):

dif. en Tabla 1	Edad		
Género	joven	medio	viejo
Hombre - mujer	4	4	4

Note además que la edad tiene el efecto de incrementar los tiempos promedios de aprendizaje de la tarea para cada género. Lo anterior se refleja en el gráfico de medias a través de *perfiles de medias paralelos*.

En las Tablas 2 a 4 se presentan casos de interacción, y en sus gráficas respectivas los *perfiles de medias no son paralelos*. Podemos ver que en estos casos, las diferencias en las medias μ_{ij} de género (hombre - mujer, es decir $\mu_{1j} - \mu_{2j}$) en cada uno de los niveles de edad no son iguales:

Dif. en Tabla 2	Edad		
Género	joven	medio	viejo
Hombre - mujer	0	2	4

En este caso las diferencias hombre-mujer entre las medias del tiempo de aprendizaje de la tarea aumentan de manera positiva a medida que el rango de edad es

mayor: De acuerdo al ejemplo, cuando los hombres y mujeres son jóvenes tienen el mismo tiempo promedio de aprendizaje en la tarea de interés, pero a medida que envejecen, los tiempos promedios aumentan y a los hombres les toma mayor tiempo promedio que a las mujeres para aprender y dicha diferencia se amplía con la edad.

Dif. en Tabla 3	Edad		
Género	joven	medio	viejo
Hombre - mujer	2	2	-2

En este caso, hasta una edad media, los hombres en promedio tardan 2 minutos más que las mujeres en aprender la tarea de interés, pero a las mujeres viejas les toma en promedio 2 minutos más que a los hombres viejos en aprender tal tarea; los tiempos promedios para cada género aumentan con la edad. Observe que la interacción en este caso tuvo el efecto de invertir la dirección de las diferencias entre las medias de tratamiento.

Dif. en Tabla 4	Edad		
Género	joven	medio	viejo
Hombre - mujer	1.5	2	2.5

En este caso, siempre existe una diferencia positiva en las medias de aprendizaje de los hombres respecto al de las mujeres en cada nivel de edad, diferencias que aumentan levemente con la edad, a la vez que en cada género, los tiempos medios también aumentan con la edad. Note que la interacción se puede catalogar como **leve** dado que los perfiles de media no son exageradamente no paralelos.

Diagnóstico de la presencia de interacción

Muestralmente podemos establecer la presencia de interacción mediante una de las siguientes formas equivalentes:

1. Si se observa que los gráficos de perfiles de medias muestrales de tratamientos ($\bar{Y}_{ij\bullet}$) no son paralelos.
2. Si la diferencia entre las respuestas medias para cualesquiera dos niveles del factor A no es la misma para todos los niveles de B.

3. Si la diferencia entre las respuestas medias para cualesquiera dos niveles del factor B no es la misma para todos los niveles de A.
4. Si las medias de tratamientos no son todas iguales a la suma de la media global más los respectivos efectos principales de los factores A y B.

8.3.4. Encubrimiento o enmascaramiento de efectos principales

Los efectos α_i y β_j constituyen los efectos principales del modelo. Cuando los efectos de interacción son significativos, estos pueden **encubrir o enmascarar** a los efectos principales de uno o de ambos factores, esto significa que en las pruebas F asociadas a la significancia simultánea de los efectos principales del respectivo factor, no se rechaza la hipótesis nula, es decir, se concluye que el factor A, o el B o ambos, no son significativos, cuando ya se ha probado que estos interactúan significativamente.

Pero tenga en cuenta que, **en presencia de interacción significativa**, no se puede concluir que alguno o ambos factores no son significativos a pesar de que el respectivo test de significancia individual para cada factor arroje como resultado la no significancia, porque tal evento es consecuencia de un **encubrimiento** debido a la interacción. Este fenómeno suele suceder en casos donde la interacción se manifiesta invirtiendo en signo las diferencias en las medias de tratamientos de un factor en cada nivel del otro factor (los perfiles de medias se cruzan).

Como previamente se dijo, para poder aislar o separar correctamente el error experimental del efecto de interacción es necesario una estimación independiente del primero, lo cual se logra mediante el uso de réplicas en cada tratamiento. Para mayor eficiencia se sugieren diseños balanceados.

8.3.5. Estimaciones

Interesa estimar medias, efectos principales y efectos de interacción. Previamente se dieron los estimadores para las medias de interés, a continuación se dan sus errores estándar e intervalos de confianza junto y para los efectos estimados.

Tabla 8.3: Parámetros y sus estimadores de MCO en un diseño factorial de efectos fijos con interacción, en un DCA.

Medias	Estimador	Error estándar	I.C del $(1 - \gamma)100\%$
μ	$\bar{Y}_{\bullet\bullet\bullet}$	$S_{\bar{Y}_{\bullet\bullet\bullet}} = \sqrt{\frac{MSE}{abn}}$	$\mu \in \bar{Y}_{\bullet\bullet\bullet} \pm t_{\gamma/2, ab(n-1)} \times S_{\bar{Y}_{\bullet\bullet\bullet}}$
$\mu_{i\bullet}$	$\bar{Y}_{i\bullet\bullet}$	$S_{\bar{Y}_{i\bullet\bullet}} = \sqrt{\frac{MSE}{bn}}$	$\mu_{i\bullet} \in \bar{Y}_{i\bullet\bullet} \pm t_{\gamma/2, ab(n-1)} \times S_{\bar{Y}_{i\bullet\bullet}}$
$\mu_{\bullet j}$	$\bar{Y}_{\bullet j\bullet}$	$S_{\bar{Y}_{\bullet j\bullet}} = \sqrt{\frac{MSE}{an}}$	$\mu_{\bullet j} \in \bar{Y}_{\bullet j\bullet} \pm t_{\gamma/2, ab(n-1)} \times S_{\bar{Y}_{\bullet j\bullet}}$
μ_{ij}	$\bar{Y}_{ij\bullet}$	$S_{\bar{Y}_{ij\bullet}} = \sqrt{\frac{MSE}{n}}$	$\mu_{ij} \in \bar{Y}_{ij\bullet} \pm t_{\gamma/2, ab(n-1)} \times S_{\bar{Y}_{ij\bullet}}$
Efectos	Estimador	Error estándar	I.C del $(1 - \gamma)100\%$
α_i	$\hat{\alpha}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$	$S_{\hat{\alpha}_i} = \sqrt{MSE \left(\frac{1}{bn} - \frac{1}{abn} \right)}$	$\alpha_i \in \hat{\alpha}_i \pm t_{\gamma/2, ab(n-1)} \times S_{\hat{\alpha}_i}$
β_j	$\hat{\beta}_j = \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$	$S_{\hat{\beta}_j} = \sqrt{MSE \left(\frac{1}{an} - \frac{1}{abn} \right)}$	$\beta_j \in \hat{\beta}_j \pm t_{\gamma/2, ab(n-1)} \times S_{\hat{\beta}_j}$
$(\alpha\beta)_{ij}$	$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}$	$S_{\widehat{(\alpha\beta)}_{ij}} = \sqrt{\frac{(a-1)(b-1)}{abn} MSE}$	$(\alpha\beta)_{ij} \in \widehat{(\alpha\beta)}_{ij} \pm t_{\gamma/2, ab(n-1)} \times S_{\widehat{(\alpha\beta)}_{ij}}$

8.3.6. Análisis de varianza

El ANOVA descompone la variabilidad total de la siguiente manera

$$SST = SSA + SSB + SS(AB) + SSE,$$

en donde:

- SST es la variabilidad total observada en la respuesta, tiene $abn - 1$ grados de libertad
- SSA mide la variación entre niveles del factor A, tiene $a - 1$ grados de libertad
- SSB mide la variación entre niveles del factor B, tiene $b - 1$ grados de libertad
- $SS(AB)$ mide la variación entre niveles del factor A mide la variación entre tratamientos (las celdas en la tabla del diseño), tiene $(a - 1)(b - 1)$ grados de libertad
- SSE mide la variación dentro de celdas, tiene $ab(n - 1)$ grados de libertad.

Nota 8.2. También puede hacerse la siguiente descomposición:

$$SST = SS(trat) + SSE,$$

donde $SS(trat)$ es la suma de cuadrados debida a los tratamientos,

$$SS(trat) = SSA + SSB + SS(AB) = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2,$$

con $g.l = ab - 1$, sin embargo no tiene relevancia para el análisis que se debe hacer.

En la Tabla 8.4, se presenta la tabla ANOVA del diseño así como las fórmulas de las distintas sumas de cuadrados y de las medias muestrales.

8.3.7. Hipótesis a probar

En su orden, deben realizarse los siguientes tests, con γ como nivel de significancia:

Significancia de la interacción

$$\begin{aligned} H_0 : (\alpha\beta)_{11} &= (\alpha\beta)_{12} = \dots = (\alpha\beta)_{ab} = 0 \\ H_1 : \text{algún } (\alpha\beta)_{ij} &\neq 0, \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \end{aligned} \quad (8.3)$$

Para este test tenemos como estadístico de prueba $F_3 = \frac{MS(AB)}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{(a-1)(b-1), ab(n-1)}$.

Rechazar a un nivel de significancia γ , si $F_3 > f_{\gamma, (a-1)(b-1), ab(n-1)}$, o si $P(f_{(a-1)(b-1), ab(n-1)} > F_3)$ es pequeño.

Si la interacción es significativa (es decir, se rechaza H_0), proceder con precaución en la interpretación de los resultados de los test de significancia sobre los efectos principales:

Significancia de los efectos principales del factor A

$$\begin{aligned} H_0 : \alpha_1 &= \alpha_2 = \dots = \alpha_a = 0 \\ H_1 : \text{algún } \alpha_i &\neq 0 \quad i = 1, 2, \dots, a \end{aligned} \quad (8.4)$$

El estadístico de prueba es $F_1 = \frac{MSA}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{a-1, ab(n-1)}$. Rechazar a un nivel de significancia γ , si $F_1 > f_{\gamma, a-1, ab(n-1)}$ o si $P(f_{a-1, ab(n-1)} > F_1)$ es pequeño.

Tabla 8.4: ANOVA en un DCA balanceado, con dos factores de efectos fijos con interacción

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F	Valor P
A	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + \frac{bn \sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
B	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b - 1}$	$\frac{MSB}{MSE}$	$P(f_{b-1, dfe} > F_0)$
AB	dfi	$SS(AB)$	$MS(AB) = \frac{SS(AB)}{dfi}$	$E[MS(AB)] = \sigma^2 + \frac{n \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta)_{ij}^2}{dfi}$	$\frac{MS(AB)}{MSE}$	$P(f_{dfi, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abn - 1$	SST				
Fórmulas						
<p>dfi = $(a - 1)(b - 1)$, los grados de libertad de la interacción AB</p> <p>dfe = $(abn - 1) - (a - 1) - (b - 1) - (a - 1)(b - 1) = ab(n - 1)$, los grados de libertad del error.</p> <p>$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$</p> <p>$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSE = SST - SSA - SSB - SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij\bullet})^2$</p> <p>donde,</p> <p>$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, es el estimador de la media μ_{ij} del tratamiento $A_i B_j$</p> <p>$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A, es el estimador de la media $\mu_{i\bullet}$</p> <p>$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B, es el estimador de la media $\mu_{\bullet j}$</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$. Media muestral de todas las abn observaciones, es el estimador de la media μ, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$. es decir el promedio de las medias muestrales de los ab tratamientos, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$. es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$. es decir es el promedio de las medias muestrales de los b niveles del factor B.</p>						

Nota 8.3. Si no se rechaza H_0 bajo interacción significativa, es posible que haya encubrimiento de los efectos de este factor por parte de la interacción.

Significancia de los efectos principales del factor B

$$\begin{aligned}
 H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0 \\
 H_1 : \text{algún } \beta_j \neq 0 \quad j = 1, 2, \dots, b
 \end{aligned}
 \tag{8.5}$$

El estadístico de prueba $F_2 = \frac{MSB}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{b-1, ab(n-1)}$. Rechazar a un nivel de significancia γ , si $F_2 > f_{\gamma, b-1, ab(n-1)}$ o si $P(f_{b-1, ab(n-1)} > F_2)$ es pequeño.

Nota 8.4. Si no se rechaza H_0 bajo interacción significativa, es posible que haya encubrimiento de los efectos de este factor por parte de la interacción.

Tener presente que:

- La estimación de un efecto principal es una diferencia promediada o contraste entre todas las medias de los niveles del factor correspondiente, por ello al hacer inferencias sobre tales medias con base en las estimaciones de los efectos principales, en presencia de interacción, se están haciendo generalizaciones sobre los niveles del factor, y por tanto dichas inferencias estarán equivocadas.
- Una comparación generalizada entre las medias de un factor tiene sentido si las diferencias existentes entre dichas medias se sostienen de nivel a nivel del otro factor, es decir, cuando no hay interacción. De lo contrario, lo mejor es realizar comparaciones y contrastes de medias de tratamientos μ_{ij} que ayuden a interpretar los efectos significativos de la interacción, Así ante el fenómeno de encubrimiento o enmascaramiento de los efectos principales de un factor, sería más conveniente comparar las medias de dicho factor pero en cada uno de los niveles del otro factor (es decir, analizar **los efectos simples de un factor** que corresponden a las comparaciones entre los niveles de tal factor dentro de un solo nivel dado del otro factor).

8.3.8. Evaluación de los efectos simples

En la Tabla 8.5 vemos la definición de los efectos simples de A en cada nivel de B. En la Tabla 8.6 se definen los efectos simples de B en cada nivel de A.

Tabla 8.5: Definición de efectos simples de A en cada nivel de B

Niveles de A	Medias de tratamiento con nivel B_1 de B	Efectos simples de A en nivel B_1
A_1	μ_{11}	$\mu_{11} - \mu_{\bullet 1} = (\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}) - (\mu + \beta_1) = \alpha_1 + (\alpha\beta)_{11}$
A_2	μ_{21}	$\mu_{21} - \mu_{\bullet 1} = (\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}) - (\mu + \beta_1) = \alpha_2 + (\alpha\beta)_{21}$
\vdots	\vdots	\vdots
A_a	μ_{a1}	$\mu_{a1} - \mu_{\bullet 1} = (\mu + \alpha_a + \beta_1 + (\alpha\beta)_{a1}) - (\mu + \beta_1) = \alpha_a + (\alpha\beta)_{a1}$
Promedio	$\mu_{\bullet 1} = \mu + \beta_1$	Suma de efectos simples de A en nivel B_1 de B: $\sum_{i=1}^a (\alpha_i + (\alpha\beta)_{i1}) = 0$
Niveles de A	Medias de tratamiento con nivel B_2 de B	Efectos simples de A en nivel B_2
A_1	μ_{12}	$\mu_{12} - \mu_{\bullet 2} = (\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}) - (\mu + \beta_2) = \alpha_1 + (\alpha\beta)_{12}$
A_2	μ_{22}	$\mu_{22} - \mu_{\bullet 2} = (\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) - (\mu + \beta_2) = \alpha_2 + (\alpha\beta)_{22}$
\vdots	\vdots	\vdots
A_a	μ_{a2}	$\mu_{a2} - \mu_{\bullet 2} = (\mu + \alpha_a + \beta_2 + (\alpha\beta)_{a2}) - (\mu + \beta_2) = \alpha_a + (\alpha\beta)_{a2}$
Promedio	$\mu_{\bullet 2} = \mu + \beta_2$	Suma de efectos simples de A en nivel B_2 de B: $\sum_{i=1}^a (\alpha_i + (\alpha\beta)_{i2}) = 0$
\vdots	\vdots	\vdots
Niveles de A	Medias de tratamiento con nivel B_b de B	Efectos simples de A en nivel B_b
A_1	μ_{1b}	$\mu_{1b} - \mu_{\bullet b} = (\mu + \alpha_1 + \beta_b + (\alpha\beta)_{1b}) - (\mu + \beta_b) = \alpha_1 + (\alpha\beta)_{1b}$
A_2	μ_{2b}	$\mu_{2b} - \mu_{\bullet b} = (\mu + \alpha_2 + \beta_b + (\alpha\beta)_{2b}) - (\mu + \beta_b) = \alpha_2 + (\alpha\beta)_{2b}$
\vdots	\vdots	\vdots
A_a	μ_{ab}	$\mu_{ab} - \mu_{\bullet b} = (\mu + \alpha_a + \beta_b + (\alpha\beta)_{ab}) - (\mu + \beta_b) = \alpha_a + (\alpha\beta)_{ab}$
Promedio	$\mu_{\bullet b} = \mu + \beta_b$	Suma de efectos simples de A en nivel B_b de B: $\sum_{i=1}^a (\alpha_i + (\alpha\beta)_{ib}) = 0$

Podemos estimar los efectos simples de cada factor en cada nivel del otro factor. Estos estimadores son definidos en la Tabla 8.7. *Podemos además probar la significancia de los efectos simples de un factor en cada nivel del otro, o equivalentemente, probar la igualdad de medias de un factor en cada nivel del otro.* Por ejemplo, suponga un experimento con $a = 3$ y $b = 2$ en el cual los datos evidencian interacción significativa pero aparentemente el factor A no es significativo; en este caso sería conveniente realizar las siguientes pruebas:

Tabla 8.6: Definición de efectos simples de B en cada nivel de A

Niveles de B	Medias de tratamiento con nivel A_1 de A	Efectos simples de B en nivel A_1
B_1	μ_{11}	$\mu_{11} - \mu_{1\bullet} = (\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}) - (\mu + \alpha_1) = \beta_1 + (\alpha\beta)_{11}$
B_2	μ_{12}	$\mu_{12} - \mu_{1\bullet} = (\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}) - (\mu + \alpha_1) = \beta_2 + (\alpha\beta)_{12}$
\vdots	\vdots	\vdots
B_b	μ_{1b}	$\mu_{1b} - \mu_{1\bullet} = (\mu + \alpha_1 + \beta_b + (\alpha\beta)_{1b}) - (\mu + \alpha_1) = \beta_b + (\alpha\beta)_{1b}$
Promedio	$\mu_{1\bullet} = \mu + \alpha_1$	Suma de Efectos simples de B en nivel A_1 de A: $\sum_{j=1}^b (\beta_j + (\alpha\beta)_{1j}) = 0$
Niveles de B	Medias de tratamiento con nivel A_2 de A	Efectos simples de B en nivel A_2
B_1	μ_{21}	$\mu_{21} - \mu_{2\bullet} = (\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}) - (\mu + \alpha_2) = \beta_1 + (\alpha\beta)_{21}$
B_2	μ_{22}	$\mu_{22} - \mu_{2\bullet} = (\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) - (\mu + \alpha_2) = \beta_2 + (\alpha\beta)_{22}$
\vdots	\vdots	\vdots
B_b	μ_{2b}	$\mu_{2b} - \mu_{2\bullet} = (\mu + \alpha_2 + \beta_b + (\alpha\beta)_{2b}) - (\mu + \alpha_2) = \beta_b + (\alpha\beta)_{2b}$
Promedio	$\mu_{2\bullet} = \mu + \alpha_2$	Suma de Efectos simples de B en nivel A_2 de A: $\sum_{j=1}^b (\beta_j + (\alpha\beta)_{2j}) = 0$
\vdots	\vdots	\vdots
Niveles de B	Medias de tratamiento con nivel A_a de A	Efectos simples de B en nivel A_a
B_1	μ_{a1}	$\mu_{a1} - \mu_{a\bullet} = (\mu + \alpha_a + \beta_1 + (\alpha\beta)_{a1}) - (\mu + \alpha_a) = \beta_1 + (\alpha\beta)_{a1}$
B_2	μ_{a2}	$\mu_{a2} - \mu_{a\bullet} = (\mu + \alpha_a + \beta_2 + (\alpha\beta)_{a2}) - (\mu + \alpha_a) = \beta_2 + (\alpha\beta)_{a2}$
\vdots	\vdots	\vdots
B_b	μ_{ab}	$\mu_{ab} - \mu_{a\bullet} = (\mu + \alpha_a + \beta_b + (\alpha\beta)_{ab}) - (\mu + \alpha_a) = \beta_b + (\alpha\beta)_{ab}$
Promedio	$\mu_{a\bullet} = \mu + \alpha_a$	Suma de Efectos simples de B en nivel A_a de A: $\sum_{j=1}^b (\beta_j + (\alpha\beta)_{aj}) = 0$

a) Comparar medias de A en el nivel 1 de B:

$$H_0 : \mu_{11} = \mu_{21} = \mu_{31} \text{ vs. } H_1 : \text{ algún par } \mu_{i1} \neq \mu_{i'1} \text{ con } i \neq i'.$$

Para la prueba se debe calcular la suma de cuadrados debidas a las celdas A_1B_1 , A_2B_1 y A_3B_1 , que denominaremos para este caso como $SS(A|B_1)$, donde

$$SS(A|B_1) = \sum_{i=1}^a \sum_{k=1}^n (Y_{i1k} - \bar{Y}_{\bullet 1\bullet})^2 - \sum_{i=1}^a \sum_{k=1}^n (Y_{i1k} - \bar{Y}_{i1\bullet})^2 = \sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i1\bullet} - \bar{Y}_{\bullet 1\bullet})^2$$

y el estadístico de la prueba y su distribución es:

$$F_{A|B_1} = \frac{SS(A|B_1) / (a - 1)}{MSE} \underset{\text{bajo } H_0}{\sim} f_{a-1, ab(n-1)},$$

donde $a = 3$, $b = 2$ en este ejemplo.

Tabla 8.7: Estimadores de efectos simples

Efectos simples de A			Efectos simples de B		
En el nivel B_1 de B			En el nivel A_1 de A		
Niveles de A	Efectos simples	Estimador	Niveles de B	Efectos simples	Estimador
A_1	$\mu_{11} - \mu_{\bullet 1}$	$\bar{Y}_{11\bullet} - \bar{Y}_{\bullet 1\bullet}$	B_1	$\mu_{11} - \mu_{1\bullet}$	$\bar{Y}_{11\bullet} - \bar{Y}_{1\bullet\bullet}$
A_2	$\mu_{21} - \mu_{\bullet 1}$	$\bar{Y}_{21\bullet} - \bar{Y}_{\bullet 1\bullet}$	B_2	$\mu_{12} - \mu_{1\bullet}$	$\bar{Y}_{12\bullet} - \bar{Y}_{1\bullet\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_a	$\mu_{a1} - \mu_{\bullet 1}$	$\bar{Y}_{a1\bullet} - \bar{Y}_{\bullet 1\bullet}$	B_b	$\mu_{1b} - \mu_{1\bullet}$	$\bar{Y}_{1b\bullet} - \bar{Y}_{1\bullet\bullet}$
$SS(A B_1) = \sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i1\bullet} - \bar{Y}_{\bullet 1\bullet})^2$			$SS(B A_1) = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{1j\bullet} - \bar{Y}_{1\bullet\bullet})^2$		
En el nivel B_2 de B			En el nivel A_2 de A		
Niveles de A	Efectos simples	Estimador	Niveles de B	Efectos simples	Estimador
A_1	$\mu_{12} - \mu_{\bullet 2}$	$\bar{Y}_{12\bullet} - \bar{Y}_{\bullet 2\bullet}$	B_1	$\mu_{21} - \mu_{2\bullet}$	$\bar{Y}_{21\bullet} - \bar{Y}_{2\bullet\bullet}$
A_2	$\mu_{22} - \mu_{\bullet 2}$	$\bar{Y}_{22\bullet} - \bar{Y}_{\bullet 2\bullet}$	B_2	$\mu_{22} - \mu_{2\bullet}$	$\bar{Y}_{22\bullet} - \bar{Y}_{2\bullet\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_a	$\mu_{a2} - \mu_{\bullet 2}$	$\bar{Y}_{a2\bullet} - \bar{Y}_{\bullet 2\bullet}$	B_b	$\mu_{2b} - \mu_{2\bullet}$	$\bar{Y}_{2b\bullet} - \bar{Y}_{2\bullet\bullet}$
$SS(A B_2) = \sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i2\bullet} - \bar{Y}_{\bullet 2\bullet})^2$			$SS(B A_2) = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{2j\bullet} - \bar{Y}_{2\bullet\bullet})^2$		
\vdots			\vdots		
En el nivel B_b de B			En el nivel A_a de A		
Niveles de A	Efectos simples	Estimador	Niveles de B	Efectos simples	Estimador
A_1	$\mu_{1b} - \mu_{\bullet b}$	$\bar{Y}_{1b\bullet} - \bar{Y}_{\bullet b\bullet}$	B_1	$\mu_{a1} - \mu_{a\bullet}$	$\bar{Y}_{a1\bullet} - \bar{Y}_{a\bullet\bullet}$
A_2	$\mu_{2b} - \mu_{\bullet b}$	$\bar{Y}_{2b\bullet} - \bar{Y}_{\bullet b\bullet}$	B_2	$\mu_{a2} - \mu_{a\bullet}$	$\bar{Y}_{a2\bullet} - \bar{Y}_{a\bullet\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_a	$\mu_{ab} - \mu_{\bullet b}$	$\bar{Y}_{ab\bullet} - \bar{Y}_{\bullet b\bullet}$	B_b	$\mu_{ab} - \mu_{a\bullet}$	$\bar{Y}_{ab\bullet} - \bar{Y}_{a\bullet\bullet}$
$SS(A B_b) = \sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{ib\bullet} - \bar{Y}_{\bullet b\bullet})^2$			$SS(B A_a) = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{aj\bullet} - \bar{Y}_{a\bullet\bullet})^2$		

b) Comparar medias de A en el nivel 2 de B:

$$H_0 : \mu_{12} = \mu_{22} = \mu_{32} \text{ vs. } H_1 : \text{algún par } \mu_{i2} \neq \mu_{i'2} \text{ con } i \neq i'.$$

Para la prueba se debe calcular las sumas de cuadrados debidas a las celdas A_1B_2 , A_2B_2 y A_3B_2 , que denominaremos para este caso como $SS(A|B_2)$, donde

$$SS(A|B_2) = \sum_{i=1}^a \sum_{k=1}^n (Y_{i2k} - \bar{Y}_{\bullet 2\bullet})^2 = \sum_{i=1}^a \sum_{k=1}^n (Y_{i2k} - \bar{Y}_{i2\bullet})^2 = \sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i2\bullet} - \bar{Y}_{\bullet 2\bullet})^2$$

y el estadístico de la prueba y su distribución es:

$$F_{A|B_2} = \frac{SS(A|B_2) / (a - 1)}{MSE} \underset{\sim}{\underset{H_0}{\text{bajo}}} f_{a-1, ab(n-1)},$$

donde $a = 3$, $b = 2$ en este ejemplo.

También, se podrían realizar las comparaciones de las medias del factor B en cada nivel de A. Siguiendo el ejemplo con $a = 3$ y $b = 2$, los tests a realizar serían:

c) Comparar medias de B en el nivel 1 de A,

$$H_0 : \mu_{11} = \mu_{12} \text{ vs. } H_1 : \mu_{11} \neq \mu_{12},$$

Para la prueba se debe calcular las sumas de cuadrados debidas a las celdas A_1B_1 , A_1B_2 , es decir $SS(B|A_1)$, donde

$$SS(B|A_1) = \sum_{j=1}^b \sum_{k=1}^n (Y_{1jk} - \bar{Y}_{1\bullet\bullet})^2 - \sum_{j=1}^b \sum_{k=1}^n (Y_{1jk} - \bar{Y}_{1j\bullet})^2 = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{1j\bullet} - \bar{Y}_{1\bullet\bullet})^2$$

con estadístico de prueba dado por

$$F_{B|A_1} = \frac{SS(B|A_1) / (b-1)}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{b-1, ab(n-1)}$$

con $a = 3$ y $b = 2$ en este ejemplo.

d) Comparar medias de B en el nivel 2 de A,

$$H_0 : \mu_{21} = \mu_{22} \text{ vs. } H_1 : \mu_{21} \neq \mu_{22}$$

En este caso se debe calcular las sumas de cuadrados debidas a las celdas A_2B_1 , A_2B_2 , dada por $SS(B|A_2)$,

$$SS(B|A_2) = \sum_{j=1}^b \sum_{k=1}^n (Y_{2jk} - \bar{Y}_{2\bullet\bullet})^2 - \sum_{j=1}^b \sum_{k=1}^n (Y_{2jk} - \bar{Y}_{2j\bullet})^2 = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{2j\bullet} - \bar{Y}_{2\bullet\bullet})^2,$$

con estadístico de prueba

$$F_{B|A_2} = \frac{SS(B|A_2) / (b-1)}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{b-1, ab(n-1)},$$

con $a = 3$ y $b = 2$ en este ejemplo.

e) comparar medias de B en el nivel 3 de A,

$$H_0 : \mu_{31} = \mu_{32} \text{ vs. } H_1 : \mu_{31} \neq \mu_{32}.$$

Para la prueba se debe calcular las sumas de cuadrados debidas a las celdas A_3B_1 , A_3B_2 , es decir, $SS(B|A_3)$,

$$SS(B|A_3) = \sum_{j=1}^b \sum_{k=1}^n (Y_{3jk} - \bar{Y}_{3\bullet\bullet})^2 - \sum_{j=1}^b \sum_{k=1}^n (Y_{3jk} - \bar{Y}_{3j\bullet})^2 = \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{3j\bullet} - \bar{Y}_{3\bullet\bullet})^2,$$

donde el estadístico de prueba corresponde a

$$F_{B|A_3} = \frac{SS(B|A_3) / (b-1)}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{b-1, ab(n-1)}$$

En general las pruebas se pueden resumir como indica la Tabla 8.8.

8.3.9. Comparaciones múltiples

Si no hay encubrimiento o enmascaramiento de efectos principales por parte de las interacciones

Pueden aplicarse los procedimientos de Bonferroni, Tukey, Scheffé, y Dunett sobre las medias de cada factor. Por ejemplo con Tukey, para el caso del factor A, la diferencia honestamente significativa para comparar las medias $\mu_{i\bullet}$, está dada por

$$HSD_A = q_\gamma(a, ab(n-1)) \sqrt{\frac{MSE}{nb}} \quad (8.6)$$

donde $q_\gamma(a, ab(n-1))$ es el valor crítico del rango estudentizado a usar.

Para el caso del factor B la diferencia honestamente significativa para comparar las medias $\mu_{\bullet j}$, está dada por

$$HSD_B = q_\gamma(b, ab(n-1)) \sqrt{\frac{MSE}{na}} \quad (8.7)$$

donde $q_\gamma(b, ab(n-1))$ es el valor crítico del rango estudentizado a usar.

Si interesa realizar inferencias sobre medias de tratamientos μ_{ij} , en este caso

$$HSD_{AB} = q_\gamma(ab, ab(n-1)) \sqrt{\frac{MSE}{n}} \quad (8.8)$$

También se pueden evaluar contrastes entre medias de nivel de un factor y contrastes entre medias de tratamientos. Por ejemplo, un intervalo de confianza del $(1-\gamma)100\%$ para un contraste de las medias del factor A está dado por

$$\sum_{i=1}^a c_i \mu_{i\bullet} \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet\bullet} \pm t_{\gamma/2, ab(n-1)} \sqrt{MSE \frac{\sum_{i=1}^a c_i^2}{bn}} \right), \text{ con } \sum_{i=1}^a c_i = 0 \quad (8.9)$$

Tabla 8.8: significancia de los efectos simples

De A en cada nivel de B			
En el nivel B_1			
Test	$SS(A B_1)$	$F_{A B_1}$	VP
$H_0 : \mu_{11} = \mu_{21} = \dots = \mu_{a1}$ vs. $H_1 : \text{algún par } \mu_{i1} \neq \mu_{i'1} \text{ con } i \neq i'$	$\sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i1\bullet} - \bar{Y}_{\bullet 1\bullet})^2$	$\frac{SS(A B_1)}{\frac{(a-1)}{MSE}}$	$P(f_{a-1,ab(n-1)} > F_{A B_1})$
En el nivel B_2			
Test	$SS(A B_2)$	$F_{A B_2}$	VP
$H_0 : \mu_{12} = \mu_{22} = \dots = \mu_{a2}$ vs. $H_1 : \text{algún par } \mu_{i2} \neq \mu_{i'2} \text{ con } i \neq i'$	$\sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{i2\bullet} - \bar{Y}_{\bullet 2\bullet})^2$	$\frac{SS(A B_2)}{\frac{(a-1)}{MSE}}$	$P(f_{a-1,ab(n-1)} > F_{A B_2})$
\vdots			
En el nivel B_b			
Test	$SS(A B_b)$	$F_{A B_b}$	VP
$H_0 : \mu_{1b} = \mu_{2b} = \dots = \mu_{ab}$ vs. $H_1 : \text{algún par } \mu_{ib} \neq \mu_{i'b} \text{ con } i \neq i'$	$\sum_{i=1}^a \sum_{k=1}^n (\bar{Y}_{ib\bullet} - \bar{Y}_{\bullet b\bullet})^2$	$\frac{SS(A B_b)}{\frac{(a-1)}{MSE}}$	$P(f_{a-1,ab(n-1)} > F_{A B_b})$
De B en cada nivel de A			
En el nivel A_1			
Test	$SS(B A_1)$	$F_{B A_1}$	VP
$H_0 : \mu_{11} = \mu_{12} = \dots = \mu_{1b}$ vs. $H_1 : \text{algún par } \mu_{1j} \neq \mu_{1j'} \text{ con } j \neq j'$	$\sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{1j\bullet} - \bar{Y}_{1\bullet\bullet})^2$	$\frac{SS(B A_1)}{\frac{(b-1)}{MSE}}$	$P(f_{b-1,ab(n-1)} > F_{B A_1})$
En el nivel A_2			
Test	$SS(B A_2)$	$F_{B A_2}$	VP
$H_0 : \mu_{21} = \mu_{22} = \dots = \mu_{2b}$ vs. $H_1 : \text{algún par } \mu_{2j} \neq \mu_{2j'} \text{ con } j \neq j'$	$\sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{2j\bullet} - \bar{Y}_{2\bullet\bullet})^2$	$\frac{SS(B A_2)}{\frac{(b-1)}{MSE}}$	$P(f_{b-1,ab(n-1)} > F_{B A_2})$
\vdots			
En el nivel A_a			
Test	$SS(B A_a)$	$F_{B A_a}$	VP
$H_0 : \mu_{a1} = \mu_{a2} = \dots = \mu_{ab}$ vs. $H_1 : \text{algún par } \mu_{aj} \neq \mu_{aj'} \text{ con } j \neq j'$	$\sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{aj\bullet} - \bar{Y}_{a\bullet\bullet})^2$	$\frac{SS(B A_a)}{\frac{(b-1)}{MSE}}$	$P(f_{b-1,ab(n-1)} > F_{B A_a})$

Un intervalo de confianza del $(1-\gamma)100\%$ para un contraste entre medias de tratamientos μ_{ij} estaría dado por

$$\sum_{i=1}^a \sum_{j=1}^b d_{ij} \mu_{ij} \in \left(\sum_{i=1}^a \sum_{j=1}^b d_{ij} \bar{Y}_{ij\bullet} \pm t_{\gamma/2, ab(n-1)} \sqrt{MSE \frac{\sum_{i=1}^a \sum_{j=1}^b d_{ij}^2}{n}} \right) \quad (8.10)$$

con $\sum_{i=1}^a d_{ij} = 0, \forall j, \sum_{j=1}^b d_{ij} = 0 \forall i$.

Si hay encubrimiento de los efectos principales

Las comparaciones podrían aplicarse a las medias de un factor en cada nivel del otro. Por ejemplo con el método de Tukey para comparar todas las medias de A en el nivel B_j , es decir las medias $\mu_{1j}, \mu_{2j}, \dots, \mu_{aj}$, la diferencia honestamente significativa a usar está dada por

$$HSD_{A|B_j} = q_\gamma(a, ab(n-1)) \sqrt{\frac{MSE}{n}}, \quad (8.11)$$

en tanto que para comparar todas las medias de B en el nivel A_i , o sea las medias $\mu_{i1}, \mu_{i2}, \dots, \mu_{ib}$, la diferencia honestamente significativa a usar está dada por

$$HSD_{B|A_i} = q_\gamma(b, ab(n-1)) \sqrt{\frac{MSE}{n}} \quad (8.12)$$

Un intervalo de confianza para un contraste de las medias del factor A en el nivel j de B estaría dado por:

$$\sum_{i=1}^a d_i \mu_{ij} \in \left(\sum_{i=1}^a d_i \bar{Y}_{ij\bullet} \pm t_{\gamma/2, ab(n-1)} \sqrt{MSE \frac{\sum_{i=1}^a d_i^2}{n}} \right) \text{ con } \sum_{i=1}^a d_i = 0. \quad (8.13)$$

8.3.10. Validación de supuestos

Debemos calcular los residuales de este modelo, los cuales corresponden a

$$\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{ij\bullet}.$$

Tenga en cuenta que los valores ajustados para cada respuesta corresponden a

$$\hat{Y}_{ijk} = \bar{Y}_{ij\bullet}.$$

Con estos valores procedemos a realizar lo siguiente:

Verificación del supuesto de independencia

Mediante el gráfico de los residuales vs. tiempo u orden de corrida experimental, Ljung-Box, Durbin-Watson, etc.

Verificación del supuesto de varianza constante

Mediante los gráficos de residuales (estandarizados) vs. valores ajustados, residuales (estandarizados) vs. factor A, residuales (estandarizados) vs. factor B.

Nota 8.5. Si hay alguna duda respecto a este supuesto después de evaluarlo gráficamente, se puede aplicar la regla empírica, donde se calculan las varianzas muestrales de cada celda (tratamiento), y entre estas varianzas se establecen los valores $S_{\text{máx}}^2$ y $S_{\text{mín}}^2$, y se evalúa si $S_{\text{máx}}^2/S_{\text{mín}}^2 < 3$, en cuyo caso se considera que el supuesto de varianza constante se cumple.

Test de normalidad

Aplicamos la prueba Shapiro Wilk sobre los residuales (estandarizados) para determinar si los errores son o no normales.

8.4. Modelo factorial dos factores de efectos fijos sin interacción, en un DCA

En caso de que se tenga evidencia previa respecto a la no significancia del término de interacción, el modelo factorial debe plantearse de la siguiente manera:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \text{ con } \varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2) \quad \forall i, j, k. \quad \sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = 0, \quad (8.14)$$

$i = 1, \dots, a, j = 1, \dots, b$ y $k = 1, \dots, n$.

8.4.1. Estimaciones

Interesa estimar medias y efectos principales. Los estimadores de MCO son presentados en la Tabla 8.9. Note que la estimación de μ_{ij} ya no es la media muestral $\bar{Y}_{ij\bullet}$, y en consecuencia, tampoco lo es la estimación de las respuestas Y_{ijk} .

8.4.2. ANOVA

Obviamente en el ANOVA la suma de cuadrados debida al término de error, es diferente en un modelo sin interacción respecto a un modelo con interacción. La descomposición de la variabilidad total queda de la siguiente manera

$$SST = SSA + SSB + SSE \quad (8.15)$$

Vea en la Tabla 8.10, el análisis de varianza y observe que la expresión del SSE y sus grados de libertad son diferentes a lo que se especificó en el caso con interacción.

Tabla 8.9: Parámetros y sus estimadores de MCO en un diseño factorial de efectos fijos sin interacción, en un DCA.

Medias	Estimador	Error estándar	I.C del $(1 - \gamma)100\%$
μ	$\bar{Y}_{\bullet\bullet}$	$S_{\bar{Y}_{\bullet\bullet}} = \sqrt{\frac{MSE}{abn}}$	$\mu \in \bar{Y}_{\bullet\bullet} \pm t_{\gamma/2, abn-a-b+1} \times S_{\bar{Y}_{\bullet\bullet}}$
$\mu_{i\bullet}$	$\bar{Y}_{i\bullet}$	$S_{\bar{Y}_{i\bullet}} = \sqrt{\frac{MSE}{bn}}$	$\mu_{i\bullet} \in \bar{Y}_{i\bullet} \pm t_{\gamma/2, abn-a-b+1} \times S_{\bar{Y}_{i\bullet}}$
$\mu_{\bullet j}$	$\bar{Y}_{\bullet j}$	$S_{\bar{Y}_{\bullet j}} = \sqrt{\frac{MSE}{an}}$	$\mu_{\bullet j} \in \bar{Y}_{\bullet j} \pm t_{\gamma/2, abn-a-b+1} \times S_{\bar{Y}_{\bullet j}}$
μ_{ij}	$\bar{Y}_{i\bullet} + \bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$	$S_{\hat{\mu}_{ij}} = \sqrt{\frac{MSE(a+b-1)}{abn}}$	$\mu_{ij} \in \hat{\mu}_{ij} \pm t_{\gamma/2, abn-a-b+1} \times S_{\hat{\mu}_{ij}}$
Efectos	Estimador	Error estándar	I.C del $(1 - \gamma)100\%$
α_i	$\hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$	$S_{\hat{\alpha}_i} = \sqrt{MSE \left(\frac{1}{bn} - \frac{1}{abn} \right)}$	$\alpha_i \in \hat{\alpha}_i \pm t_{\gamma/2, abn-a-b+1} \times S_{\hat{\alpha}_i}$
β_j	$\hat{\beta}_j = \bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$	$S_{\hat{\beta}_j} = \sqrt{MSE \left(\frac{1}{an} - \frac{1}{abn} \right)}$	$\beta_j \in \hat{\beta}_j \pm t_{\gamma/2, abn-a-b+1} \times S_{\hat{\beta}_j}$

8.4.3. Hipótesis a probar

Con γ como nivel de significancia:

Significancia del factor A

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

$$H_1 : \text{algún } \alpha_i \neq 0 \quad i = 1, 2, \dots, a$$

Estadístico de prueba $F_1 = \frac{MSA}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{a-1, abn-a-b+1}$

Rechazar a un nivel de significancia γ , si $F_1 > f_{\gamma, a-1, abn-a-b+1}$ o si $P(f_{a-1, abn-a-b+1} > F_1)$ pequeño

Significancia del factor B

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0$$

$$H_1 : \text{algún } \beta_j \neq 0 \quad j = 1, 2, \dots, b$$

Estadístico de prueba $F_2 = \frac{MSB}{MSE} \underset{\sim}{\sim}^{bajo H_0} f_{b-1, abn-a-b+1}$

Tabla 8.10: ANOVA en un DCA balanceado, con dos factores de efectos fijos sin interacción

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F	Valor P
A	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + \frac{bn \sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
B	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b - 1}$	$\frac{MSB}{MSE}$	$P(f_{b-1, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abn - 1$	SST				
Fórmulas						
<p>dfe = $(abn - 1) - (a - 1) - (b - 1) = abn - a - b + 1$, los grados de libertad del error.</p> <p>$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSE = SST - SSA - SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$</p> <p>donde,</p> <p>$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$ (ya no corresponde al estimador de MCO de μ_{ij})</p> <p>$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$ Media muestral en el nivel i del factor A, es el estimador de la media $\mu_{i\bullet}$</p> <p>$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$ Media muestral en el nivel j del factor B, es el estimador de la media $\mu_{\bullet j}$</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.</p>						

Rechazar a un nivel de significancia γ , si $F_2 > f_{\gamma, b-1, abn-a-b+1}$ o si $P(f_{b-1, abn-a-b+1} > F_2)$ es pequeño.

8.4.4. Comparaciones múltiples y contrastes de medias

Pueden realizarse comparaciones múltiples sobre medias de los factores. Por ejemplo, los tests de Tukey sobre las medias de cada factor, teniendo en cuenta que para

el caso del factor A la diferencia honestamente significativa está dada por

$$HSD_A = q_\gamma(a, abn - a - b + 1) \sqrt{\frac{MSE}{nb}}, \quad (8.16)$$

donde $q_\gamma(a, abn - a - b + 1)$ es el valor crítico del rango estudentizado a usar.

Para el caso del factor B la diferencia honestamente significativa está dada por

$$HSD_B = q_\gamma(b, abn - a - b + 1) \sqrt{\frac{MSE}{na}}, \quad (8.17)$$

donde $q_\gamma(b, abn - a - b + 1)$ es el valor crítico del rango estudentizado a usar.

Se pueden evaluar contrastes entre medias de nivel de un factor, y contrastes entre medias de tratamientos. Por ejemplo, un intervalo de confianza del $(1 - \gamma)100\%$ para un contraste de las medias del factor A está dado por

$$\sum_{i=1}^a c_i \mu_{i\bullet} \in \left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet\bullet} \pm t_{\gamma/2, abn-a-b+1} \sqrt{MSE \frac{\sum_{i=1}^a c_i^2}{bn}} \right) \text{ con } \sum_{i=1}^a c_i = 0. \quad (8.18)$$

Un intervalo de confianza del $(1 - \gamma)100\%$ para un contraste entre medias de tratamientos, $\omega = \sum_{i=1}^a \sum_{j=1}^b d_{ij} \mu_{ij}$, estaría dado por

$$\omega \in \left(\sum_{i=1}^a \sum_{j=1}^b d_{ij} \hat{\mu}_{ij} \pm t_{\gamma/2, abn-a-b+1} \sqrt{\widehat{\text{Var}}(\hat{\omega})} \right), \text{ con } \sum_{i=1}^a d_{ij} = 0 \forall j, \sum_{j=1}^b d_{ij} = 0 \forall i, \quad (8.19)$$

con $\hat{\mu}_{ij} = \bar{Y}_{i\bullet\bullet} + \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$.

8.4.5. Validación de supuestos

De la misma forma que en el modelo con interacción, sin embargo los residuos y valores ajustados son

$$\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}$$

$$\hat{Y}_{ijk} = \bar{Y}_{i\bullet\bullet} + \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$$

8.5. Ejemplo experimento factorial dos factores de efectos fijos en un DCA (adaptado de [7]: Página 868, Problema 19.16)

El personal de un centro de servicio electrónico incluye tres técnicos especializados en reparar las tres marcas más usadas de procesadores para microcomputadores. Se desea estudiar los efectos del factor TECNICO y del factor MARCA de procesador, sobre el tiempo de servicio dado en minutos. A cada técnico se le asignaron aleatoriamente cinco trabajos de cada una de las marcas y se registró el número de minutos requerido para completar cada trabajo.

Técnico	MARCA		
	1	2	3
1	62	57	59
	48	45	53
	63	39	67
	57	54	66
	69	44	47
2	51	61	55
	57	58	58
	45	70	50
	50	66	69
	39	51	49
3	59	58	47
	65	63	56
	55	70	51
	52	53	44
	70	60	50

1. Defina claramente el modelo de análisis de varianza requerido para estudiar este problema.
2. Haga un análisis descriptivo de los datos.
3. Determine si la interacción es significativa. Escriba la expresión empleada para calcular el valor p. Emplee un nivel de significancia de 0.05 en todas las pruebas.
4. Discuta la significancia de los efectos principales factores TECNICO y MARCA. En el caso de indicios de encubrimiento realice pruebas de hipótesis para confirmarlo y haga las recomendaciones.
5. Construya la tabla de medias por tratamiento.

6 Plantee un contraste para probar si el tiempo promedio empleado para reparar un micro de la marca 2 y uno de la marca 3 es el mismo para el técnico 2 y 3.

7 Compare las medias del factor TÉCNICO en cada nivel del factor MARCA, y viceversa ¿Permiten estos resultados tomar decisiones acerca del problema en cuestión? Cuáles son las recomendaciones que usted haría si quiere disminuir el tiempo de servicio del centro de servicios electrónicos.

8 Valide supuestos sobre los errores del modelo (varianza cte. y normalidad)

8.5.1. Modelo

El modelo para el análisis de varianza corresponde entonces a:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

donde tomando como factor A: TECNICO y B: MARCA, con niveles $a = 3$ y $b = 3$ respectivamente y $n = 5$,

Y_{ijk} es el tiempo de reparación en la k-ésima unidad de la marca j con técnico i

μ es el promedio global del tiempo de reparación de una unidad

α_i es el efecto del técnico i sobre el promedio global del tiempo de reparación de una unidad

β_j es el efecto de la marca j sobre el promedio global del tiempo de reparación de una unidad

$(\alpha\beta)_{ij}$ es el efecto de la interacción del técnico i con la marca j sobre el promedio global del tiempo de reparación de una unidad

ε_{ijk} es el error aleatorio en la k-ésima réplica del tratamiento $A_i B_j$. Supuestos $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$.

Restricciones para estimación por mínimos cuadrados: $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.

8.5.2. Análisis descriptivo

En primera instancia debemos analizar los gráficos de interacción presentados en en las Figuras 8.2(a) y 8.2(b). En estas figuras se observa que los perfiles de medias de las marcas y de los técnicos, respectivamente, no son paralelos, lo cual implica que posiblemente la interacción entre técnicos y marcas es significativa. A nivel del técnico 1 se observa que su menor tiempo promedio corresponde a la reparación de unidades marca 2 y tarda más con las otras dos marcas; el tiempo promedio de este técnico con las marcas 2 y 3 es aproximadamente el mismo. A nivel del técnico 2 los tiempos promedios de reparación son distintos para las tres marcas, y el menor tiempo promedio lo alcanza con la marca 1, en tanto que su mayor tiempo promedio

fue con la marca 2. El técnico 3 presentó el menor tiempo promedio con la marca 3, en tanto que con las marcas 1 y 2 tardó más y para estos dos casos los tiempos promedios observados resultaron aproximadamente iguales. También se observa que los tiempos promedio para la marca 2 con los técnicos 2 y 3 son muy similares; y los tiempos promedio para la marca 1 con los técnicos 1 y 3 son también similares entre sí. Pero el hecho más importante a destacar es que cada técnico es especialista en una marca y que para cada marca sólo un técnico alcanza el mejor (es decir el menor) tiempo promedio. Si el objetivo del estudio es buscar mejorar el tiempo promedio de reparación, uno estaría tentado a recomendar que los trabajos de reparación de unidades de la marca 1 sean asignados al técnico 2, los de la marca 2 al técnico 1 y los de la marca 3 al técnico 3. Sin embargo, implementar tal medida haría al sistema ineficiente, es como tener “una cola en espera de servicio para cada servidor”, lo cual en la teoría de colas, no es la mejor disposición. La administración del centro de servicio debe ante esta situación, implementar medidas correctivas para lograr que todos los tres técnicos sean igualmente efectivos con cualquiera de las marcas.

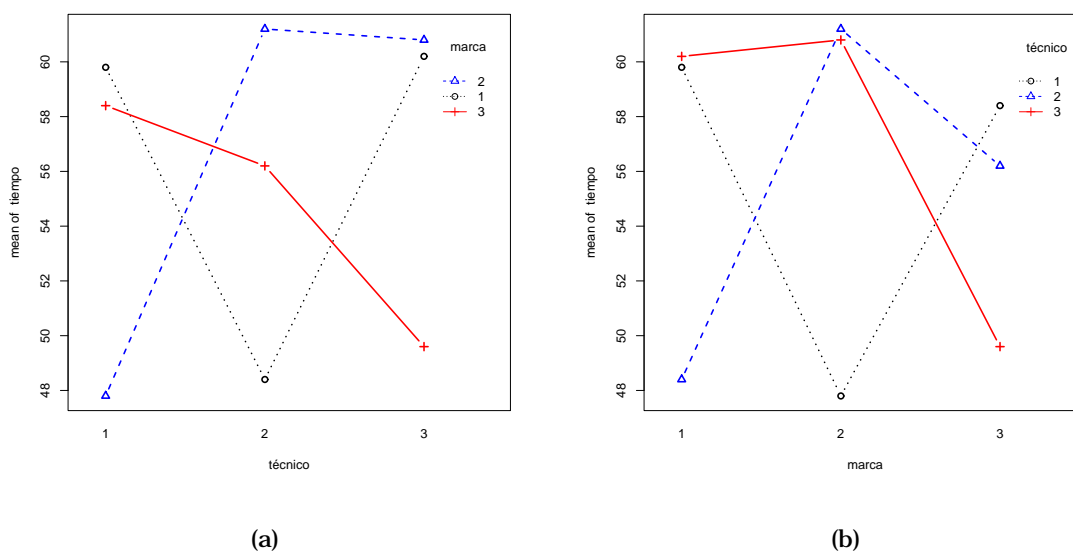


Figura 8.2: (a) Perfiles de medias de tratamientos según marca. (b) perfiles de medias de tratamientos según técnico.

Sin embargo, cuando se examinan los boxplots en las Figuras 8.3(a) y 8.3(b) muestran que prácticamente los tiempos promedios de reparación no difieren ni por técnico ni por marca, es decir, mirar aisladamente estas dos figuras nos llevaría a concluir erróneamente que ninguno de los dos factores es significativo!. Si en realidad hay interacción significativa, entonces lo que vemos en las Figuras 8.3(a) y 8.3(b), es un enmascaramiento de los efectos principales de cada factor por parte de la interacción. Adicionalmente en las Figuras 8.3(a) y 8.3(b) vemos que según los altos de las cajas, la dispersión de los tiempos de reparación puede concebirse aproximadamente constante de nivel a nivel de cada factor. Los boxplots que deben presentarse y analizarse

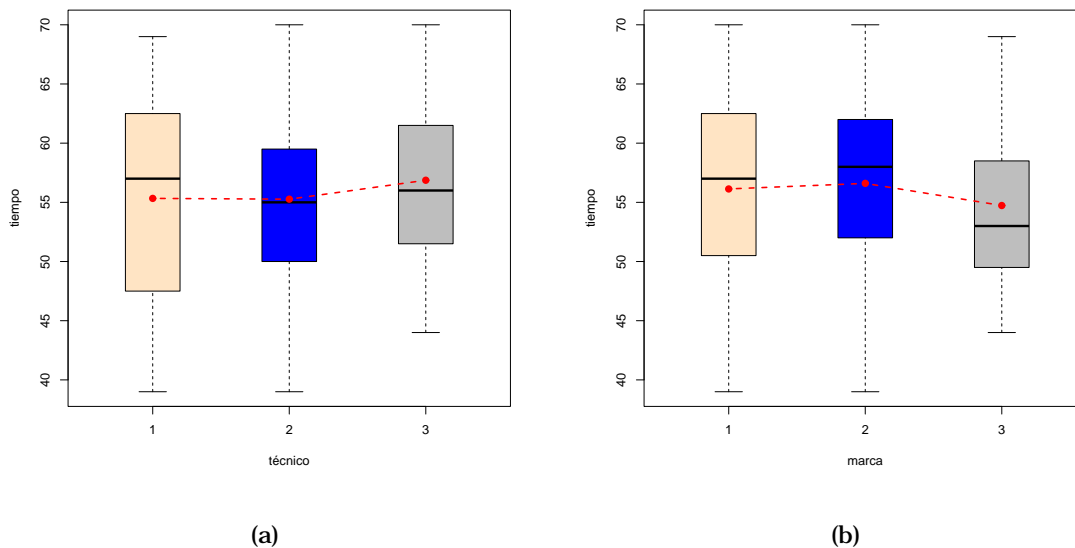


Figura 8.3: (a) Distribuciones según técnico. (b) Distribuciones según marca.

cuando hay interacción significativa, son como los que aparecen en las Figuras 8.4(a) y 8.4(b), en donde en términos de las medias de tratamientos, llegamos a las mismas conclusiones previamente extraídas de las gráficas de interacción.

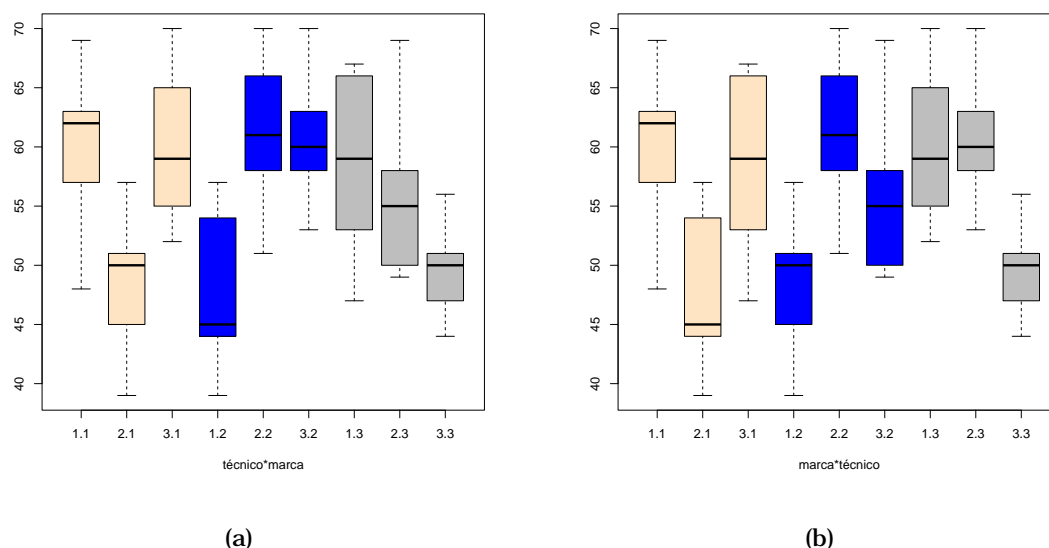


Figura 8.4: (a) Distribuciones de tiempos de técnicos en cada marca. (b) Distribuciones de tiempos por marcas en cada técnico.

8.5.3. ANOVA y tests de hipótesis

Salida R 8.5.1. Tabla ANOVA

```
Analysis of Variance Table
Response: tiempo
              Df Sum Sq Mean Sq F value    Pr(>F)
técnico         2   24.58   12.289    0.2363  0.7907788
marca           2   28.31   14.156    0.2722  0.7632826
técnico:marca    4 1215.29  303.822   5.8415 0.0009941 ***
Residuals      36 1872.40   52.011
---
```

Significancia de la interacción

Con base en la anterior salida R, se procede a realizar el test de hipótesis siguiente:

$H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \dots = (\alpha\beta)_{33} = 0$ (interacción técnico \times marca no es significativa)

$H_1 : \text{algún } (\alpha\beta)_{ij} \neq 0, i = 1, 2, 3 \text{ y } j = 1, 2, 3$ (interacción técnico \times marca es significativa)

El estadístico de prueba es

$$F_3 = MS(AB) / MSE \stackrel{H_0}{\sim} f_{4,36}, \quad \text{con } VP = P(f_{4,36} > 5.84) = 0.0010$$

Se concluye con una significancia de 0.05 que sí hay interacción significativa entre los factores TÉCNICO y MARCA, por lo tanto los efectos de cada factor dependen del otro. Como consecuencia de esto es necesario interpretar con precaución los resultados de los tests sobre la significancia de los efectos principales de cada factor: De hecho, *si la interacción es significativa, entonces los dos factores son significativos a través de la interacción!*.

Tests sobre los efectos principales

Con base en la salida R, se procede a realizar los siguientes test de hipótesis:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ (efectos principales del factor técnico no son significativos)

$H_1 : \text{algún } \alpha_i \neq 0, i = 1, 2, 3$ (algunos efectos principales del técnico son significativos)

o bien, en términos de las medias de este factor:

$$H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \mu_{3\bullet} \left(\begin{array}{l} \text{Las medias de los tiempos de reparación según el factor técnico} \\ \text{no difieren significativamente} \end{array} \right)$$

$$H_1 : \mu_{i\bullet} \neq \mu_{j\bullet} \left(\begin{array}{l} \text{Al menos dos de las medias de los tiempos de reparación según el factor} \\ \text{técnico difieren significativamente} \end{array} \right)$$

En cualquier caso el estadístico de prueba corresponde a

$$F_1 = MSA/MSE \stackrel{H_0}{\sim} f_{2,36} \text{ con } VP = P(f_{2,36} > 0.24) = 0.7908$$

Se concluye con una significancia de 0.05, *que ante interacción significativa, los efectos principales del factor TÉCNICO son enmascarados por la interacción y por lo tanto no pueden ser medidos sus efectos principales!* Esto quiere decir que no tiene sentido estimar ni efectos, ni medias marginalmente para los niveles de TÉCNICO y tampoco tiene sentido construir I.C, o hacer comparaciones de Tukey o contrastes con las medias de este factor.

Con relación al factor marca,

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (efectos principales del factor marca no son significativos)

$H_1 : \text{algún } \beta_j \neq 0, j = 1, 2, 3$ (algunos efectos principales del factor marca son significativos)

o bien, en términos de las medias de este factor:

$$H_0 : \mu_{\bullet 1} = \mu_{\bullet 2} = \mu_{\bullet 3} \left(\begin{array}{l} \text{Las medias de los tiempos de reparación según el factor marca} \\ \text{no difieren significativamente} \end{array} \right)$$

$$H_1 : \text{algún par } \mu_{\bullet i} \neq \mu_{\bullet j} \left(\begin{array}{l} \text{Al menos dos de las medias de los tiempos de reparación según} \\ \text{el factor marca difieren significativamente} \end{array} \right)$$

En cualquier caso el estadístico de prueba corresponde a

$$F_2 = MSB/MSE \stackrel{H_0}{\sim} f_{2,36}, \text{ con } VP = P(f_{2,36} > 0.27) = 0.7633$$

Se concluye con una significancia de 0.05, *que ante interacción significativa, los efectos principales del factor MARCA son enmascarados por la interacción y por lo tanto no pueden ser medidos sus efectos principales!* Esto quiere decir que no tiene sentido estimar ni efectos, ni medias para los niveles de MARCA y tampoco tiene sentido construir I.C, o hacer comparaciones de Tukey o contrastes con las medias de este factor.

Lo más recomendable es realizar contrastes o comparaciones de medias de tratamientos μ_{ij} , o bien, comparar medias de un factor en cada nivel del otro factor. Sin embargo, el problema con Tukey es que se tienen nueve medias de tratamientos, un número demasiado grande como para que los procedimientos de comparaciones de pares de medias tengan potencia para detectar diferencias.

8.5.4. Inferencias sobre medias de tratamientos técnico*marca

Estimaciones e intervalos de confianza del 95 %

La siguiente tabla presenta las medias estimadas $\bar{Y}_{ij\bullet}$ de cada tratamiento $A_i B_j$, así como los intervalos de confianza que corresponden a

$$\mu_{ij} \in \bar{Y}_{ij\bullet} \pm t_{0.025, ab(n-1)} \sqrt{\frac{MSE}{n}}$$

Salida R 8.5.2. *Estimaciones de medias de tratamientos y los I.C respectivos, calculados en R con la función `lsmeans()` de la librería del mismo nombre.*

```
> summary(lsmeans(modelo, ~técnico*marca))
técnico marca lsmean      SE df lower.CL upper.CL
1         1      59.8 3.225248 36 53.25889 66.34111
2         1      48.4 3.225248 36 41.85889 54.94111
3         1      60.2 3.225248 36 53.65889 66.74111
1         2      47.8 3.225248 36 41.25889 54.34111
2         2      61.2 3.225248 36 54.65889 67.74111
3         2      60.8 3.225248 36 54.25889 67.34111
1         3      58.4 3.225248 36 51.85889 64.94111
2         3      56.2 3.225248 36 49.65889 62.74111
3         3      49.6 3.225248 36 43.05889 56.14111
Confidence level used: 0.95
```

En esta salida R puede observarse la similaridad en las estimaciones e I.C para las medias de los tratamientos: $\{(1,1), (1,3), (2,2), (3,1), (3,2)\}$ y $\{(1,2), (2,1), (3,3)\}$, lo cual puede verificarse en las Figuras 8.2(a), 8.2(b), 8.4(a) y 8.4(b). Podría entonces realizarse contrastes de medias para verificar por ejemplo si

$$\mu_{11} = \mu_{13} = \mu_{22} = \mu_{31} = \mu_{32}$$

y otro para determinar si

$$\mu_{12} = \mu_{21} = \mu_{33},$$

o un procedimiento de Tukey sobre las $ab = 9$ medias de tratamientos, para lo cual se puede construir los I.C de Tukey del 95 %, así:

$$(\mu_{ij} - \mu_{kl}) \in (\bar{Y}_{ij\bullet} - \bar{Y}_{kl\bullet}) \pm q_{0.05}(ab, ab(n-1)) \sqrt{\frac{MSE}{n}}$$

y verificar si las medias de tratamientos indicadas resultan iguales, como se muestra en la siguiente salida R. Sin embargo, note que en este caso no se obtienen diferencias significativas entre tratamientos, no porque no existan tales diferencias, sino porque como ya se ha dicho, se pierde potencia en este procedimiento con un número tan grande de comparaciones. Aquí las medias de tratamiento μ_{ij} , son etiquetadas, respectivamente, como $i : j$ así, por ejemplo, en la comparación 2:1 - 1:1 se hace referencia a la diferencia de las medias $\mu_{21} - \mu_{11}$.

Salida R 8.5.3. Intervalos de confianza de Tukey del 95 % para las medias de tratamientos técnico*marca.

```
> TukeyHSD(modelo, "técnico:marca", conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = tiempo ~ técnico * marca)
$'técnico:marca'

      diff      lwr      upr    p adj
2:1-1:1 -11.4 -26.43865  3.63865 0.2658953
3:1-1:1   0.4 -14.63865 15.43865 1.0000000
1:2-1:1 -12.0 -27.03865  3.03865 0.2095773
2:2-1:1   1.4 -13.63865 16.43865 0.9999970
3:2-1:1   1.0 -14.03865 16.03865 0.9999998
1:3-1:1  -1.4 -16.43865 13.63865 0.9999970
2:3-1:1  -3.6 -18.63865 11.43865 0.9965056
3:3-1:1 -10.2 -25.23865  4.83865 0.4054029
3:1-2:1  11.8  -3.23865 26.83865 0.2273106
1:2-2:1  -0.6 -15.63865 14.43865 1.0000000
2:2-2:1  12.8  -2.23865 27.83865 0.1487668
3:2-2:1  12.4  -2.63865 27.43865 0.1771874
1:3-2:1  10.0  -5.03865 25.03865 0.4316582
2:3-2:1   7.8  -7.23865 22.83865 0.7366003
3:3-2:1   1.2 -13.83865 16.23865 0.9999991
1:2-3:1 -12.4 -27.43865  2.63865 0.1771874
2:2-3:1   1.0 -14.03865 16.03865 0.9999998
3:2-3:1   0.6 -14.43865 15.63865 1.0000000
1:3-3:1  -1.8 -16.83865 13.23865 0.9999787
2:3-3:1  -4.0 -19.03865 11.03865 0.9928966
3:3-3:1 -10.6 -25.63865  4.43865 0.3552211
2:2-1:2  13.4  -1.63865 28.43865 0.1130579
3:2-1:2  13.0  -2.03865 28.03865 0.1359755
1:3-1:2  10.6  -4.43865 25.63865 0.3552211
2:3-1:2   8.4  -6.63865 23.43865 0.6555252
3:3-1:2   1.8 -13.23865 16.83865 0.9999787
3:2-2:2  -0.4 -15.43865 14.63865 1.0000000
1:3-2:2  -2.8 -17.83865 12.23865 0.9994109
2:3-2:2  -5.0 -20.03865 10.03865 0.9710966
3:3-2:2 -11.6 -26.63865  3.43865 0.2460831
1:3-3:2  -2.4 -17.43865 12.63865 0.9998112
2:3-3:2  -4.6 -19.63865 10.43865 0.9825836
```

3:3-3:2	-11.2	-26.23865	3.83865	0.2867388
2:3-1:3	-2.2	-17.23865	12.83865	0.9999018
3:3-1:3	-8.8	-23.83865	6.23865	0.5990906
3:3-2:3	-6.6	-21.63865	8.43865	0.8715885

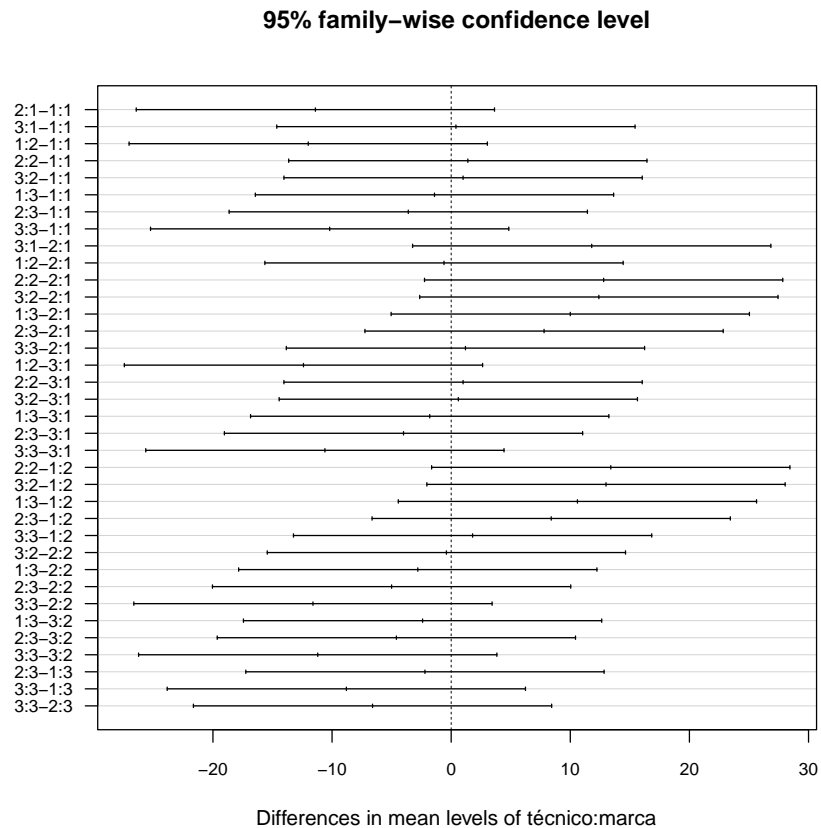


Figura 8.5: Gráficos I.C de Tukey del 95 % de confianza, para las diferencias de medias de tratamientos. Observe que no se detectan diferencias pues el cero está contenido en todos los intervalos, sin embargo, esto es debido a una gran reducción de la potencia ante un número grande de comparaciones. Observe además que en muchos casos el cero está muy cerca a uno de los límites, y en esas circunstancias es muy probable que sí existan diferencias entre las medias de tratamientos para las cuales se observa esta situación.

Contraste para probar si el tiempo promedio empleado para reparar un micro de la marca 2 y uno de la marca 3 es el mismo para el técnico 2 y 3

Se pide comparar el promedio de las medias μ_{22} y μ_{23} es igual al promedio de las medias μ_{32} y μ_{33} . Por tanto el contraste indicado es

$$W = \frac{1}{2}(\mu_{22} + \mu_{23}) - \frac{1}{2}(\mu_{32} + \mu_{33})$$

El test de hipótesis corresponde a:

$$\begin{array}{ll} H_0 : W = 0 & H_0 : \mu_{22} + \mu_{23} = \mu_{32} + \mu_{33} \\ \text{o equivalentemente,} & \\ H_1 : W \neq 0, & H_1 : \mu_{22} + \mu_{23} \neq \mu_{32} + \mu_{33} \end{array}$$

Como cada media muestral de tratamientos satisface que,

$$\bar{Y}_{ij\bullet} \sim N\left(\mu_{ij}, \frac{\sigma^2}{n}\right)$$

y todas las medias muestrales de tratamiento son mutuamente independientes, entonces el estimador del contraste W es

$$\widehat{W} = \frac{1}{2}(\bar{Y}_{22\bullet} + \bar{Y}_{23\bullet}) - \frac{1}{2}(\bar{Y}_{32\bullet} + \bar{Y}_{33\bullet}) \sim N\left(W, \sigma^2 \left[\frac{1}{4n} + \frac{1}{4n} + \frac{1}{4n} + \frac{1}{4n}\right]\right) = N\left(W, \frac{\sigma^2}{n}\right).$$

Para esta prueba se puede construir un estadístico t dado por

$$T_0 = \frac{0.5 \times (\bar{Y}_{22\bullet} + \bar{Y}_{23\bullet}) - 0.5 \times (\bar{Y}_{32\bullet} + \bar{Y}_{33\bullet})}{\sqrt{\frac{MSE}{n}}} \stackrel{H_0}{\sim} t_{ab(n-1)},$$

y se rechazaría H_0 a un nivel de significancia de 0.05 si $|T_0| > t_{0.025, ab(n-1)} = t_{0.025, 36} = 2.028$. Para la prueba en cuestión se tiene que

$$\bar{Y}_{22\bullet} = 61.20, \bar{Y}_{23\bullet} = 56.20, \bar{Y}_{32\bullet} = 60.80, \bar{Y}_{33\bullet} = 49.60, n = 5, MSE = 52.011111,$$

luego $T_0 = 1.0852$, por tanto se concluye que el tiempo promedio empleado para reparar un micro de la marca 2 y uno de la marca 3 es el mismo para los técnicos 2 y 3.

Probando la igualdad de las medias de un factor en cada nivel del otro factor

a) Resultados para comparar las medias del factor marca en cada nivel de técnico Los siguientes resultados fueron obtenidos en R con la función usuario

`Comparmediasslices2()`. En esa función el argumento `modelo` es para especificar el objeto R en que previamente se guardó el ajuste del modelo anova con función `aov` y `nreplicas` es para especificar el número de réplicas de tratamientos n . Ver Código R 8.8.1.

Salida R 8.5.4. Resultados para probar la igualdad de las medias del factor marca en cada nivel del factor técnico.

Significancia Efectos técnico * marca sobre respuesta tiempo en cada nivel de técnico					
técnico	Df	Sum.Sq	Mean.Sq	F0	P.value
1	1	2 430.5333	215.2667	4.138859	0.02410520
2	2	2 416.1333	208.0667	4.000427	0.02698640
3	3	2 396.9333	198.4667	3.815851	0.03140553

La anterior salida R presenta en cada renglón, los resultados para probar, respectivamente, los siguientes tests:

Renglón 1 Comparación de las medias según las marcas, a nivel del técnico 1:

$$H_0 : \mu_{11} = \mu_{12} = \mu_{13}$$

El estadístico para probar esta hipótesis corresponde a $F_0 = MS(B|A_1)/MSE = 4.14$, $F_0 \stackrel{H_0}{\sim} f_{b-1, ab(n-1)}$, donde

$$MS(B|A_1) = n \sum_{j=1}^b (\bar{Y}_{1j\bullet} - \bar{Y}_{1\bullet\bullet})^2 / (b-1),$$

es la suma de cuadrados medios debida al factor MARCA pero correspondiente a las observaciones con el técnico 1. El $VP = P(f_{2,36} > 4.14) = 0.0241$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con el técnico 1 los tiempos promedios de reparación por unidad de al menos dos de las tres marcas son diferentes.

Renglón 2 Comparación de las medias de las marcas a nivel del técnico 2:

$$H_0 : \mu_{21} = \mu_{22} = \mu_{23}$$

El estadístico para probar esta hipótesis corresponde a: $F_0 = MS(B|A_2)/MSE = 4.00$, $F_0 \stackrel{H_0}{\sim} f_{b-1, ab(n-1)}$, donde

$$MS(B|A_2) = n \sum_{j=1}^b (\bar{Y}_{2j\bullet} - \bar{Y}_{2\bullet\bullet})^2 / (b-1)$$

es la suma de cuadrados medios debida al factor MARCA pero correspondiente a las observaciones con el técnico 2. El $VP = P(f_{2,36} > 4.00) = 0.0270$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con el técnico 2 los tiempos promedios de reparación por unidad de al menos dos de las tres marcas son diferentes.

Renglón 3 Comparación de las medias de las marcas a nivel del técnico 3:

$$H_0 : \mu_{31} = \mu_{32} = \mu_{33}$$

El estadístico para probar esta hipótesis corresponde a: $F_0 = MS(B|A_3)/MSE = 3.82$, $F_0 \stackrel{H_0}{\sim} f_{b-1, ab(n-1)}$, donde

$$MS(B|A_3) = n \sum_{j=1}^b (\bar{Y}_{3j\bullet} - \bar{Y}_{3\bullet\bullet})^2 / (b-1)$$

es la suma de cuadrados medios debida al factor MARCA pero correspondiente a las observaciones con el técnico 3. El $VP = P(f_{2,36} > 3.82) = 0.0314$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con el técnico 3 los tiempos promedios de reparación por unidad de al menos dos de las tres marcas son diferentes.

b) Resultados para comparar las medias del factor técnico en cada nivel de marca

Salida R 8.5.5. Resultados para probar la igualdad de las medias del factor técnico en cada nivel del factor marca.

Significancia Efectos técnico * marca sobre respuesta tiempo en cada nivel de marca						
	marca	Df	Sum.Sq	Mean.Sq	F0	P.value
1	1	2	448.9333	224.4667	4.315744	0.020888163
2	2	2	581.2000	290.6000	5.587268	0.007703847
3	3	2	209.7333	104.8667	2.016236	0.147918236

En la anterior salida R se aprecian los resultados de los testes sobre la igualdad de las medias del factor TÉCNICO en cada nivel del factor MARCA, veamos cada una:

Renglón 1 Comparación de las medias de los técnicos a nivel de la marca 1:

$$H_0 : \mu_{11} = \mu_{21} = \mu_{31}$$

El estadístico para probar esta hipótesis corresponde a $F_0 = MS(A|B_1)/MSE = 4.32$, $F_0 \stackrel{H_0}{\sim} f_{a-1, ab(n-1)}$, donde

$$MS(A|B_1) = n \sum_{i=1}^a (\bar{Y}_{i1\bullet} - \bar{Y}_{\bullet 1\bullet})^2 / (a - 1)$$

es la suma de cuadrados medios debida al factor TÉCNICO pero correspondiente a las observaciones con la marca 1. El $VP = P(f_{2,36} > 4.32) = 0.0209$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con la marca 1 los técnicos tardan en promedio diferente tiempo para reparar una unidad (al menos dos de ellos).

Renglón 2 Comparación de las medias de los técnicos a nivel de la marca 2:

$$H_0 : \mu_{12} = \mu_{22} = \mu_{32}$$

El estadístico para probar esta hipótesis corresponde a: $F_0 = MS(A|B_2)/MSE = 5.59$, $F_0 \stackrel{H_0}{\sim} f_{a-1, ab(n-1)}$, donde

$$MS(A|B_2) = n \sum_{i=1}^a (\bar{Y}_{i2\bullet} - \bar{Y}_{\bullet 2\bullet})^2 / (a - 1)$$

es la suma de cuadrados medios debida al factor TÉCNICO pero correspondiente a las observaciones con la marca 2. El $VP = P(f_{2,36} > 5.59) = 0.0077$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con la marca 2 los técnicos tardan en promedio diferente tiempo para reparar una unidad (al menos dos de ellos).

Renglón 3 Comparación de las medias de los técnicos a nivel de la marca 3:

$$H_0 : \mu_{13} = \mu_{23} = \mu_{33}$$

El estadístico para probar esta hipótesis corresponde a: $F_0 = MS(A|B_3)/MSE = 2.02$, $F_0 \stackrel{H_0}{\sim} f_{a-1, ab(n-1)}$, donde

$$MS(A|B_3) = n \sum_{i=1}^a (\bar{Y}_{i3\bullet} - \bar{Y}_{\bullet 3\bullet})^2 / (a - 1)$$

es la suma de cuadrados medios debida al factor TÉCNICO pero correspondiente a las observaciones con la marca 3. El $VP = P(f_{2,36} > 2.02) = 0.1479$. Con estos resultados se concluye que a un nivel de significancia de 0.05, con la marca 3 los técnicos tardan en promedio el mismo tiempo para reparar una unidad. Sin embargo mire en las Figura 8.2(a), 8.2(b), 8.4(a) y 8.4(b), estas tres medias y en la salida R 8.5.2 los valores de las medias estimadas que corresponden a $\bar{Y}_{13\bullet} = 58.40$, $\bar{Y}_{23\bullet} = 56.20$, y $\bar{Y}_{33\bullet} = 49.60$, hay una diferencia de 8.8 y 6.6 minutos de las dos primeras medias respecto a la última; la pregunta es si en términos prácticos tales diferencias son o no significativas, a pesar que no se encuentra significancia estadística!

Respecto a la cuestión de si ¿permiten estos resultados tomar decisiones acerca del problema en cuestión? Cuáles son las recomendaciones que usted haría si quiere disminuir el tiempo de servicio del centro de servicios electrónicos: Como se mencionó antes, para mejorar los tiempos promedios no es conveniente asignar los trabajos de reparación a los técnicos según la marca con la cual mostraron mejor desempeño. Es preferible implementar a través de un reentrenamiento, una medida correctiva tendiente a minimizar los tiempos con todas las marcas y que los tres técnicos tengan igual desempeño.

8.5.5. Validación de supuestos

Con los residuos internamente estudentizados, se obtuvieron los resultados para validar la normalidad y los gráficos de residuos para evaluar la varianza y la no carencia de ajuste.

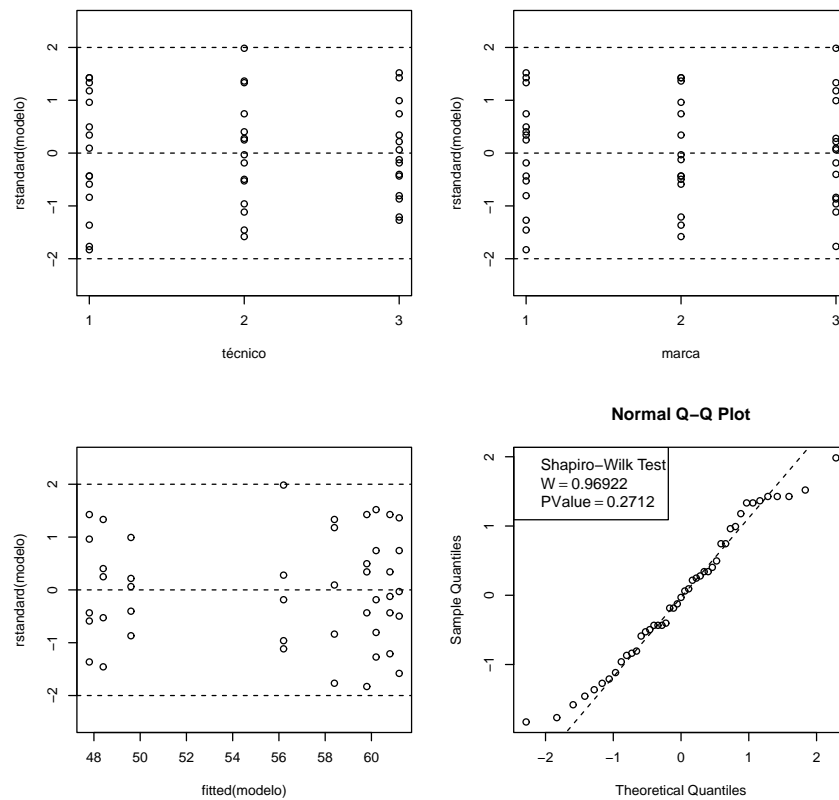


Figura 8.6: Gráficos de residuos internamente estudentizados y de probabilidad normal en ejemplo 8.5.

A partir del valor P de 0.2712 del test de Shapiro Wilk en la Figura 8.6 y del gráfico de probabilidad normal se puede concluir que hay una violación moderada del supuesto de normalidad para los errores del modelo. En la Figura 8.6 se presentan también las gráficas de residuales internamente estudentizados vs. técnico, vs. marca y vs. valores ajustados. En estas tres gráficas se observa un residual cortando la banda de +2, indicando un residual outlier. En términos generales puede considerarse que el supuesto de varianza constante se cumple aproximadamente.

Código R 8.5.1. Programa R usado e Ejemplo 8.5.

```
library(gmodels)
library(multcomp)
library(daewr)
library(car)
library(lsmeans)
library(agricolae)

reparación=data.frame(marca=factor(rep(1:3,times=15)),
técnico=factor(rep(c(1:3),each=15)),tiempo=scan())
62 57 59
48 45 53
63 39 67
57 54 66
69 44 47
51 61 55
57 58 58
45 70 50
50 66 69
39 51 49
59 58 47
65 63 56
55 70 51
52 53 44
70 60 50

attach(reparación)

#GRÁFICOS PARA ANALIZAR LOS DATOS EXPERIMENTALES
mediastécnico=sapply(split(tiempo,técnico),mean)
mediasmarca=sapply(split(tiempo,marca),mean)
#Gráficos de interacción
interaction.plot(técnico,marca,tiempo,type="b",pch=c(1,2,3),col=c("black","blue","red"),
lwd=2)
interaction.plot(marca,técnico,tiempo,type="b",pch=c(1,2,3),col=c("black","blue","red"),
lwd=2)
boxplot(tiempo~técnico,xlab="técnico",ylab="tiempo",boxwex=0.4,col=c("bisque","blue","gray"))
lines(1:3,mediastécnico,type="b",pch=19,col=2,lty=2,lwd=2)
boxplot(tiempo~marca,xlab="marca",ylab="tiempo",boxwex=0.4,col=c("bisque","blue","gray"))
lines(1:3,mediasmarca,type="b",pch=19,col=2,lty=2,lwd=2)
```


8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA247

```
cat("Significancia Efectos",names(modelo$model)[2],"*",names(modelo$model)[3],
    "sobre respuesta",names(modelo$model)[1],"en cada nivel de",names(modelo$xlevels)[1],"\\n")
print(slicebyfactor2.1)
cat("\\n")
cat("Significancia Efectos",names(modelo$model)[2],"*",names(modelo$model)[3],
    "sobre respuesta",names(modelo$model)[1],"en cada nivel de",names(modelo$xlevels)[2],"\\n")
print(slicebyfactor1.2)
}

Comparmediasslices2(modelo=modelo,nreplicas=5)

shapiro.test(rstandard(modelo))
#OBTENIENDO GRÁFICOS DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS,
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
stripchart(rstandard(modelo)~técnico,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,xlab="técnico")
abline(h=c(-2,0,2),lty=2)
stripchart(rstandard(modelo)~marca,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,xlab="marca")
abline(h=c(-2,0,2),lty=2)
plot(fitted(modelo),rstandard(modelo),ylim=c(-2.5,2.5))
abline(h=c(-2,0,2),lty=2)
qqnorm(rstandard(modelo))
qqline(rstandard(modelo),lty=2)
legend("topleft",legend=c("Shapiro-Wilk Test",expression(W==0.96922),
    expression(PValue== 0.2712)),cex=1.1)

detach(reparación)
```

8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA

Consideraremos ahora el caso en el cual los niveles de los factores A y B son muestras aleatorias de una población de posibles niveles, respectivamente. Por tanto cada tratamiento resultante de la combinación de niveles de A con niveles de B hace parte también de una muestra aleatoria de posibles tratamientos. Una vez elegidos los a niveles del factor A y los b niveles del factor B, a cada tratamiento se asignan en forma completamente aleatorizada, n unidades experimentales obteniéndose por tanto de cada tratamiento n observaciones de la variable respuesta. En total habrán

$N = abn$ observaciones de la variable respuesta. Un tabla de entrada de datos del experimento será similar a la presentada en el caso del diseño factorial de efectos fijos, sin embargo, el modelo con interacción de los factores para la respuesta k -ésima en el tratamiento $A_i B_j$ (nivel i del factor A combinado con nivel j del factor B), es el que se presenta a continuación:

8.6.1. Modelo ANOVA

Tanto los efectos principales como los de interacción son variables aleatorias y el modelo corresponde a

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n \quad (8.20)$$

con los siguientes supuestos, $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$, $\forall i, j, k$, $A_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$, $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$ y $(AB)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$. Además ε_{ijk} , A_i , B_j y $(AB)_{ij}$ son mutuamente independientes.

NOTE QUE: La media en el tratamiento $A_i B_j$ es $M_{ij} = \mu + A_i + B_j + (AB)_{ij}$, es decir, es una variable aleatoria! Similarmente las medias por nivel de cada factor, también son variables aleatorias. Por tanto, no tiene sentido ni es de interés, hacer inferencias sobre medias de tratamientos ni medias de niveles de cada factor.

TENGA EN CUENTA QUE: La variable respuesta bajo el modelo anterior tiene las siguientes características:

- Y_{ijk} es una variable aleatoria normal, pues es la suma de una constante más cuatro variables aleatorias normales independientes.
- $E(Y_{ijk}) = \mu$ y $V(Y_{ijk}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2$ que corresponde a la varianza total.
- $COV(Y_{ijk}, Y_{i'j'k'}) = 0$, para todo $i \neq i'$, $j \neq j'$, es decir, observaciones que están en tratamientos con distintos niveles en A y B son independientes.
- $COV(Y_{ijk}, Y_{ij'k'}) = \sigma_\alpha^2$, para todo $j \neq j'$, es decir, observaciones que están en mismo nivel de A pero diferente nivel de B tienen covarianza igual a σ_α^2 .

8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA249

- $COV(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2$, para todo $i \neq i'$, es decir, observaciones que están en el mismo nivel de B pero diferente nivel de A tienen covarianza igual a σ_β^2 .
- $COV(Y_{ijk}, Y_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2$, para todo $k \neq k'$, es decir, observaciones que están en el mismo tratamiento (mismo nivel de A y de B) tienen covarianza igual a $\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2$.

8.6.2. Parámetros de interés

El problema que enfrentamos en este modelo es determinar si la varianza de cada efecto aleatorio es significativa, por tanto los parámetros de interés son las componentes de la varianza total. Estas componentes tienen la siguiente interpretación:

- σ^2 la variación entre muestras de tratamientos o celdas (de la tabla de datos).
- σ_α^2 la variación del desempeño del factor A, o variación por cambios del factor A.
- σ_β^2 la variación del desempeño del factor B, o variación por cambios del factor B.
- $\sigma_{\alpha\beta}^2$ su existencia implica que el desempeño relativo de varios niveles del factor A en relación a cambios en el factor B no varía de forma congruente con dichos cambios.

Nota 8.6. *Los experimentos factoriales de efectos aleatorios permiten al investigador establecer el factor o factores que contribuyen significativamente a la variabilidad total de una variable respuesta, y determinar si estas fuentes de variabilidad exceden un nivel aceptable, y de ser necesario y posible, corregir las deficiencias en estos factores.*

8.6.3. ANOVA

Para la tabla ANOVA tenemos en cuenta los siguientes estadísticos,

- Y_{ijk} es la respuesta en la k -ésima réplica del tratamiento $A_i B_j$
- $\bar{Y}_{i..}$ es la media muestral de las nb réplicas en el nivel i del factor A, es decir
$$\bar{Y}_{i..} = \frac{1}{nb} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij.}$$

- $\bar{Y}_{\bullet j \bullet}$ es la media muestral de las na réplicas en el nivel j del factor B, es decir,

$$\bar{Y}_{\bullet j \bullet} = \frac{1}{na} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij \bullet}$$
- con $\bar{Y}_{ij \bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, la media muestral en el tratamiento $A_i B_j$,
- $\bar{Y}_{\dots} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$, la media muestral global de las abn observaciones.

El ANOVA descompone la variabilidad total de la misma manera que el caso con efectos fijos, es decir:

$$SST = SSA + SSB + SS(AB) + SSE$$

y las fórmulas de estas sumas de cuadrados y sus grados de libertad son las mismas que en el caso de dos factores con efectos fijos bajo interacción. Sin embargo, los valores esperados de los cuadrados medios no son iguales a los dados cuando los efectos son fijos. Vea en Tabla 8.11 los valores esperados, los estadísticos F y las expresiones probabilísticas de los correspondientes valores P. Note que para los estadísticos F asociados a los efectos aleatorios (principales) de A y de B, se utiliza en el denominador el $MS(AB)$ en lugar del MSE , en cambio, para el estadístico F asociado a los efectos aleatorios de la interacción, se usa en el denominador al MSE . Por tanto, los grados de libertad para el denominador de las distribuciones f asociadas a A y B son $(a-1)(b-1)$, y para la interacción estos grados de libertad en el denominador son $ab(n-1)$ (desde que los diseños sean balanceados).

Observe además que los valores esperados de los cuadrados medios son funciones de las componentes de varianza y a través de éstas expresiones es posible definir los estimadores de estas varianzas.

8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA251

Tabla 8.11: ANOVA en un DCA balanceado, con dos factores de efectos aleatorios con interacción

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F	Valor P
A	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_{\alpha}^2$	$\frac{MSA}{MS(AB)}$	$P(f_{a-1, dfi} > F_0)$
B	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + n\sigma_{\alpha\beta}^2 + na\sigma_{\beta}^2$	$\frac{MSB}{MS(AB)}$	$P(f_{b-1, dfe} > F_0)$
AB	dfi	$SS(AB)$	$MS(AB) = \frac{SSAB}{dfi}$	$E[MS(AB)] = \sigma^2 + n\sigma_{\alpha\beta}^2$	$\frac{MS(AB)}{MSE}$	$P(f_{dfi, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abn - 1$	SST				

Fórmulas	
dfi = $(a - 1)(b - 1)$, los grados de libertad de la interacción AB	
dfe = $(abn - 1) - (a - 1) - (b - 1) - (a - 1)(b - 1) = ab(n - 1)$, los grados de libertad del error.	
$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$	
$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$	
$SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$	
$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$	
$SSE = SST - SSA - SSB - SSAB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij\bullet})^2$	
donde,	
$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, media muestral en tratamientos $A_i B_j$	
$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A.	
$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B.	
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ , o bien	
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien	
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,	
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.	

8.6.4. Pruebas de hipótesis

En su orden, deben realizarse los siguientes tests, con γ como nivel de significancia:

Significancia de la componente de varianza debida a la interacción

$$H_0 : \sigma_{\alpha\beta}^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_{\alpha\beta}^2 > 0. \quad (8.21)$$

Note que si H_0 fuese cierta, entonces $E[MS(AB)] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_3 = \frac{MS(AB)}{MSE} \stackrel{H_0}{\sim} f_{(a-1)(b-1), ab(n-1)}. \quad (8.22)$$

Rechazar H_0 a un nivel de significancia γ si $F_3 > f_{\gamma, (a-1)(b-1), ab(n-1)}$, o con valor P si $P(f_{(a-1)(b-1), ab(n-1)} > F_3)$ es pequeño.

Si la interacción es significativa (es decir, se rechaza H_0), proceder con precaución en la interpretación de los resultados de los test de significancia sobre las componentes de varianza de los factores A y B. Al respecto Algunos estadísticos afirman lo siguiente:

...No tiene sentido probar la significancia de las componentes σ_α^2 y σ_β^2 si la componente de varianza $\sigma_{\alpha\beta}^2$ es significativa. Consecuentemente tales pruebas sólo deberían realizarse cuando los factores respectivos no estén involucrados en interacciones significativas...([2])

Significancia de la componente de varianza del factor A

$$H_0 : \sigma_\alpha^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\alpha^2 > 0. \quad (8.23)$$

Note que si H_0 fuese cierta, entonces $E[MSA] = \sigma^2 + n\sigma_{\alpha\beta}^2$ que también corresponde al $E[MS(AB)]$. Por tanto el estadístico de la prueba a usar es

$$F_1 = \frac{MSA}{MS(AB)} \stackrel{H_0}{\sim} f_{a-1, (a-1)(b-1)}. \quad (8.24)$$

8.6. Modelo factorial dos factores de efectos aleatorios con interacción, en un DCA253

Rechazar H_0 a un nivel de significancia γ si $F_1 > f_{\gamma, a-1, (a-1)(b-1)}$, o con valor P si $P(f_{a-1, (a-1)(b-1)} > F_1)$ es pequeño.

Significancia de la componente de varianza del factor B

$$H_0 : \sigma_\beta^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\beta^2 > 0. \quad (8.25)$$

Note que si H_0 fuese cierta, entonces $E[MSB] = \sigma^2 + n\sigma_{\alpha\beta}^2$ que también corresponde al $E[MS(AB)]$. Por tanto el estadístico de la prueba a usar es

$$F_2 = \frac{MSB}{MS(AB)} \stackrel{H_0}{\sim} f_{b-1, (a-1)(b-1)}. \quad (8.26)$$

Rechazar H_0 a un nivel de significancia γ si $F_2 > f_{\gamma, b-1, (a-1)(b-1)}$, o con valor P si $P(f_{b-1, (a-1)(b-1)} > F_2)$ es pequeño.

8.6.5. Estimaciones

Interesa estimar las componentes de varianza. Las estimaciones de estas componentes se determinan igualando los cuadrados medios esperados a los correspondientes cuadrados medios observados y se despejan los valores desconocidos de las componentes. Las estimaciones de las cuatro componentes de la varianza son:

Tabla 8.12: Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos aleatorios con interacción

Fuente	Componente de varianza estimada	Proporción respecto a la varianza total
A	$\hat{\sigma}_\alpha^2 = \frac{MSA - MS(AB)}{nb}$	$\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
B	$\hat{\sigma}_\beta^2 = \frac{MSB - MS(AB)}{na}$	$\frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
AB	$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS(AB) - MSE}{n}$	$\frac{\hat{\sigma}_{\alpha\beta}^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
Error	$\hat{\sigma}^2 = MSE$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$

Nota 8.7. Bajo interacción significativa, puede ocurrir que las componentes de varianza estimadas $\hat{\sigma}_\alpha^2$ y/o $\hat{\sigma}_\beta^2$ resulten negativas. Obviamente no tiene sentido dar como resultado una varianza negativa. Así que es mejor no hacer inferencias ni estimaciones de estas componentes bajo interacción significativa.

8.6.6. Validación de supuestos

Podemos chequear los supuestos sobre los términos de error ε_{ijk} reemplazando temporalmente todos los efectos aleatorios por efectos fijos, calculando los residuales y examinando los gráficos de residuales (estandarizados) y los test de normalidad de la manera como se hizo en el caso del experimento factorial de efectos fijos. Sin embargo chequear los supuestos de cada uno de los efectos aleatorios A_i , B_j , $(AB)_{ij}$, no es fácil, dado que en un modelo factorial por lo general hay pocos niveles de cada factor observado, y no todos los promedios de celdas son independientes. Consecuentemente en general, se omiten los chequeos para los supuestos de tales efectos aleatorios y se espera que cualquier problema severo sea detectado en el análisis de los residuales.

8.7. Modelo factorial de dos factores con efectos aleatorios sin interacción en un diseño DCA

En caso de que los datos obtenidos presenten evidencia respecto a la no significancia de la componente de varianza debida al término de interacción, el modelo factorial debe replantearse de la siguiente manera:

$$Y_{ijk} = \mu + A_i + B_j + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n \quad (8.27)$$

con los siguientes supuestos, $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$, $\forall i, j, k$, $A_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$, $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$. Además ε_{ijk} , A_i y B_j son mutuamente independientes.

TENGA EN CUENTA QUE: La variable respuesta bajo el modelo anterior tiene las siguientes características:

- Y_{ijk} es una variable aleatoria normal, pues es la suma de una constante más cuatro variables aleatorias normales independientes.
- $E(Y_{ijk}) = \mu$ y $V(Y_{ijk}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$ que corresponde a la varianza total.
- $COV(Y_{ijk}, Y_{i'j'k'}) = 0$, para todo $i \neq i'$, $j \neq j'$, es decir, observaciones que están en tratamientos con distintos niveles en A y B son independientes.

8.7. Modelo factorial de dos factores con efectos aleatorios sin interacción en un diseño DCA255

- $COV(Y_{ijk}, Y_{ij'k'}) = \sigma_\alpha^2$, para todo $j \neq j'$, es decir, observaciones que están en mismo nivel de A pero diferente nivel de B tienen covarianza igual a σ_α^2 .
- $COV(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2$, para todo $i \neq i'$, es decir, observaciones que están en el mismo nivel de B pero diferente nivel de A tienen covarianza igual a σ_β^2 .
- $COV(Y_{ijk}, Y_{ijk'}) = \sigma_\alpha^2 + \sigma_\beta^2$, para todo $k \neq k'$, es decir, observaciones que están en el mismo tratamiento (mismo nivel de A y de B) tienen covarianza igual a $\sigma_\alpha^2 + \sigma_\beta^2$.

Obviamente en el ANOVA la suma de cuadrados debida al término de error, es diferente en un modelo sin interacción respecto a un modelo con interacción.

8.7.1. ANOVA

Con las mismas definiciones de las diferentes medias muestrales que se dieron previamente en este capítulo, ahora se procede a descomponer la variabilidad total de la siguiente manera

$$SST = SSA + SSB + SSE$$

en donde la tabla ANOVA queda como se muestra en Tabla 8.13. Observe que en este caso los estadísticos F asociados a los factores A y B , se construyen dividiendo las respectivas sumas de cuadrados medios por el MSE el cual, tiene ahora $abn - a - b + 1$ grados de libertad.

8.7.2. Hipótesis a probar

Significancia de la componente de varianza del factor A

$$H_0 : \sigma_\alpha^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\alpha^2 > 0. \quad (8.28)$$

Note que si H_0 fuese cierta, entonces $E[MSA] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_1 = \frac{MSA}{MSE} \stackrel{H_0}{\sim} f_{a-1, abn-a-b+1}. \quad (8.29)$$

Rechazar H_0 a un nivel de significancia γ si $F_1 > f_{\gamma, a-1, abn-a-b+1}$, o con valor P si $P(f_{a-1, abn-a-b+1} > F_1)$ es pequeño.

Tabla 8.13: ANOVA en un DCA balanceado, con dos factores de efectos aleatorios sin interacción

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F	Valor P
A	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + nb\sigma_\alpha^2$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
B	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + na\sigma_\beta^2$	$\frac{MSB}{MSE}$	$P(f_{b-1, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abn - 1$	SST				
Fórmulas						
<p>dfe = $(abn - 1) - (a - 1) - (b - 1) = abn - a - b + 1$, los grados de libertad del error.</p> <p>$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSE = SST - SSA - SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$</p> <p>donde,</p> <p>$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, media muestral en tratamientos $A_i B_j$</p> <p>$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A.</p> <p>$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B.</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.</p>						

Significancia de la componente de varianza del factor B

Corresponde al siguiente test:

$$H_0 : \sigma_\beta^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\beta^2 > 0. \quad (8.30)$$

Note que si H_0 fuese cierta, entonces $E[MSB] = \sigma^2$ que también corresponde al

$E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_2 = \frac{MSB}{MSE} \stackrel{H_0}{\sim} f_{b-1, abn-a-b+1}. \quad (8.31)$$

Rechazar H_0 a un nivel de significancia γ si $F_2 > f_{\gamma, b-1, abn-a-b+1}$, o con valor P si $P(f_{b-1, abn-a-b+1} > F_2)$ es pequeño.

8.7.3. Estimaciones

Las estimaciones de las tres componentes de la varianza son:

Tabla 8.14: Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos aleatorios sin interacción

Fuente	Componente de varianza estimada	Proporción respecto a la varianza total
A	$\hat{\sigma}_\alpha^2 = \frac{MSA - MSE}{nb}$	$\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}^2}$
B	$\hat{\sigma}_\beta^2 = \frac{MSB - MS(AB)}{na}$	$\frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}^2}$
Error	$\hat{\sigma}^2 = MSE$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}^2}$

8.7.4. Validación de supuestos

En el modelo sin interacción la respuesta ajustada y los residuos de ajuste corresponden a $\hat{Y}_{ijk} = \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}$ y $\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$, es decir, asumiendo temporalmente que los efectos son fijos podemos chequear los supuestos sobre los términos de error examinando los gráficos de residuales y los test de normalidad de la manera como se hizo en el caso del experimento factorial de efectos fijos sin interacción. De nuevo, chequear los supuestos de cada uno de los efectos aleatorios básicamente no es posible mientras se tengan pocos niveles en cada factor y además por la no independencia entre los promedios muestrales de las celdas o tratamientos.

8.8. Ejemplo dos factores de efectos aleatorios en un DCA

En los laboratorios de Hewlet-Packard se estudió el efecto de la temperatura de impresión (factor A), y el tipo de tinta (factor B), en la nitidez de la impresión en un nuevo tipo de impresoras para fotografías. Se escogieron aleatoriamente cuatro niveles de temperatura, indicados por A1, A2, A3, A4. Se escogieron, también aleatoriamente,

tres tipos de tintas, indicados por B1, B2, B3. La nitidez es la variable respuesta Y , y a mayor valor indica mejor calidad de la impresión. Se obtuvieron cuatro réplicas en cada combinación de niveles. Se quiere conocer si existe interacción entre el tipo de tinta y la temperatura con respecto a la nitidez.

A: Temp. de impresión	B: Tipo de tinta		
	B1	B2	B3
A1	252.23	296.20	288.59
	253.32	327.97	299.12
	239.63	295.67	283.26
	247.48	328.08	279.47
A2	241.72	261.04	258.29
	256.31	209.03	212.89
	241.85	261.59	233.84
	257.55	240.16	247.95
A3	88.95	204.56	183.43
	90.48	172.36	161.39
	67.24	240.02	200.66
	96.44	208.66	189.55
A4	233.15	329.03	242.48
	236.80	316.88	232.05
	251.28	335.31	253.80
	240.96	332.46	223.89

- Escriba el modelo apropiado.
- Haga un análisis descriptivo del conjunto de datos.
- Pruebe la hipótesis de componente de varianza de interacción entre temperatura y tipo de tinta, al 5 %.
- Pruebe la hipótesis de que la temperatura y los tipos de tinta no tienen efecto en la variabilidad de la nitidez de impresión, a un 5 %.
- Estime las componentes de varianza.
- Validación de supuestos sobre los errores.

8.8.1. Modelo para el experimento

De acuerdo al planteamiento del problema, se trata de un experimento bifactorial, ambos factores con efectos aleatorios, en un diseño completamente aleatorizado. Si tomamos como factor A: Temperatura de impresión, y B: Tipo de tinta, el modelo correspondiente es:

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}$$

con niveles $a = 4$ y $b = 3$ respectivamente y $n = 4$, Y_{ijk} es la nitidez de la impresión en la k -ésima réplica del tratamiento ij . μ es el promedio global de la nitidez de impresión.

A_i es el efecto aleatorio de la temperatura sobre el promedio global de la nitidez de impresión, $A_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$. B_j es el efecto aleatorio de los tipos de tinta sobre el promedio global de la nitidez de impresión, $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$. $(AB)_{ij}$ es el efecto aleatorio de la interacción de las temperaturas con los tipos de tinta, sobre el promedio global de la nitidez de impresión, $(AB)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$. ε_{ijk} es el error aleatorio en la k -ésima réplica del tratamiento ij . Supuestos $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$, ε_{ijk} , A_i , B_j y $(AB)_{ij}$ son mutuamente independientes.

8.8.2. Análisis descriptivo

En la Figura 8.7 se presentan las gráficas correspondientes. En las gráficas de interacción, para los tratamientos que fueron observados, los perfiles de medias no son paralelos, esto puede ser evidencia de una componente de varianza significativa debido a la interacción de los dos factores, lo cual implicaría que los resultados relativos a varios niveles de temperatura en relación a cambios en el tipo de tinta no varían de forma congruente con dichos cambios, o viceversa, que el desempeño relativo de varios tipos de tinta en relación a cambios en la temperatura de impresión, no varían de forma congruente con tales cambios. Los resultados de nitidez varían tanto con el tipo de tinta como con la temperatura de impresión, esto es igualmente corroborado por las figuras de los boxplots. En los boxplots para niveles del factor temperatura, además de ver variaciones en los resultados promedios según varía la temperatura, también se observa cambios en la dispersión en cada nivel de temperatura, por ejemplo, en el segundo nivel de temperatura que resultó seleccionado para el experimento, la dispersión de los resultados fueron mucho menores comparativamente con los otros niveles. En los boxplots para niveles del factor tinta, también se notan variaciones en los promedios de nitidez como en la dispersión de los resultados por tipo de tinta, en particular con la tinta tres que fue seleccionada, la dispersión es menor respecto a la de las otras dos. Sin embargo tenga en cuenta que estos análisis marginales con cada factor deben realizarse con precaución en presencia de interacción significativa.

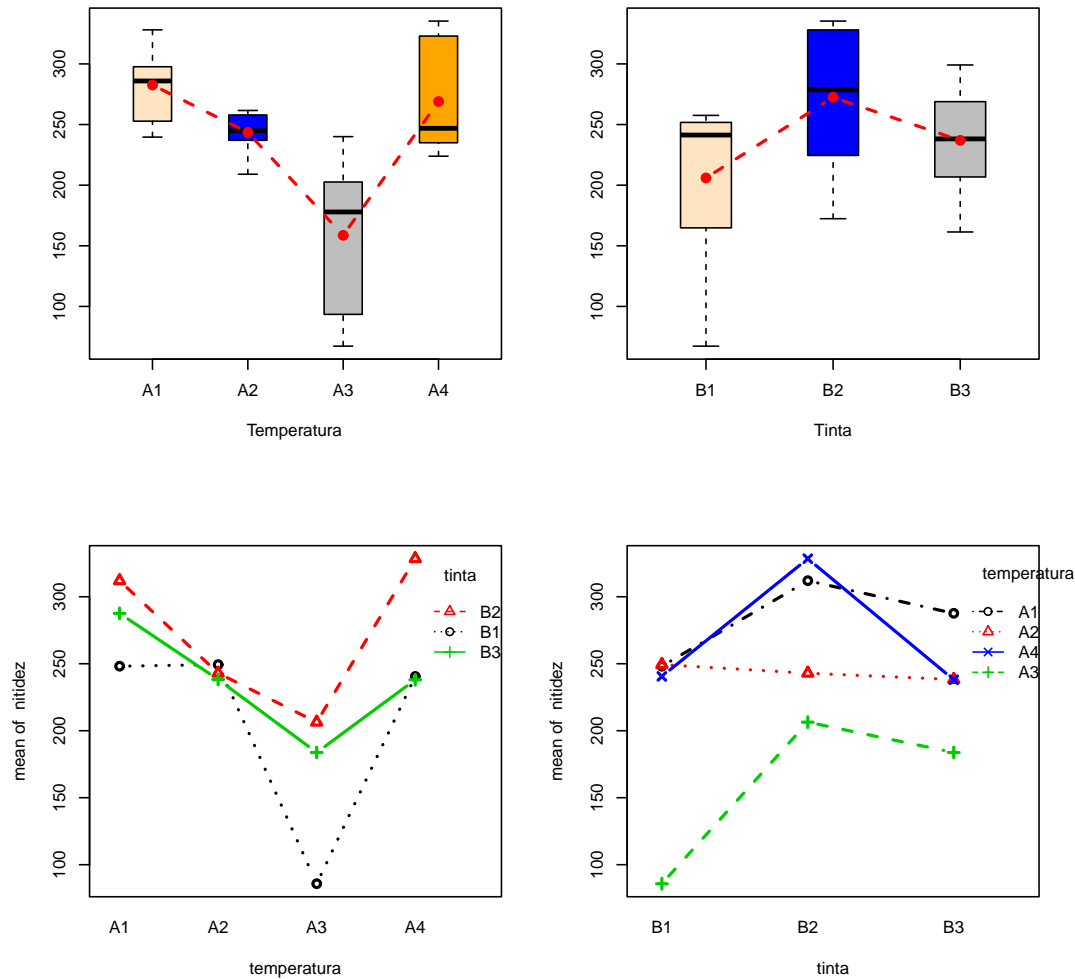


Figura 8.7: Boxplots según factores y gráficos de interacción, en datos del Ejemplo 8.8

8.8.3. ANOVA y estimación de componentes de varianza

En R, la función `aov()` permite correr modelo de efectos aleatorios pero no permite construir la tabla Anova como esperamos, pues no calcula los estadísticos F ni sus valores P. Los resultados R que se exhiben a continuación fueron obtenidos con esta función.

Salida R 8.8.1. Resultados modelo factorial dos factores aleatorios, con función `aov()`.

```
> modelo=aov(nitidez~Error(temperatura*tinta));summary(modelo)
Error: temperatura
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3 111314    37105
Error: tinta
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  2  35409    17705
Error: temperatura:tinta
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  6  27211     4535
Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals 36   9079     252.2
```

Una manera de subsanar la falencia de las funciones R para obtener la tabla ANOVA correcta en modelos de dos factores aleatorios con interacción, es crear una función de usuario que lo haga. La siguiente salida R fue obtenida con la función usuario `Anovafactorialaleatoriosconinteraccion()` la cual maneja como único argumento a `modelo`, en el cual se especifica un objeto R que haya guardado el ajuste del modelo ANOVA de dos factores en un DCA balanceado, pero asumiendo efectos fijos. Internamente la función hace los cálculos necesarios para producir una ANOVA tal como lo hace el paquete estadístico SAS, y el resultado en este ejemplo es el siguiente

Salida R 8.8.2. Resultados para modelo factorial de dos factores aleatorios, con función usuario `Anovafactorialaleatoriosconinteraccion()`.

```
Tests of Hypotheses for Random Model Analysis of Variance
Dependent Variable: nitidez

      Source DF      Sum.Sq    Mean.Sq F.Value P.value
1      temperatura  3 111314.21 37104.738    8.18 0.01531
2              tinta  2  35409.05 17704.523    3.9 0.08219
3  temperatura:tinta  6  27211.05  4535.176
      Source DF      Sum.Sq    Mean.Sq F.Value P.value
1  temperatura:tinta  6  27211.053 4535.1755   17.98      0
2              Error 36   9078.755  252.1876
```

NOTE QUE en el modelo con interacción los estadísticos F para probar la significancia de las componentes de varianza debidas a la Temperatura y a la Tinta, se construyen usando como denominador al MS(AB), en tanto que para probar la significancia de la componente de varianza debida a la interacción, el respectivo estadístico se construye usando como denominador al MSE.

Significancia de la componente de varianza de la interacción Temperatura de impresión con Tinta

El test de hipótesis sobre la significancia de la componente de varianza debida a la interacción es

$$H_0 : \sigma_{\alpha\beta}^2 = 0 \text{ vs. } H_1 : \sigma_{\alpha\beta}^2 > 0$$

El estadístico de la prueba a usar es $F_3 = \frac{MS(AB)}{MSE} = 17.98 \stackrel{\text{bajo } H_0}{\sim} f_{6,36}$ con $VP = P(f_{6,36} > F_3) < 0.0001$, luego, los factores Temperatura y Tinta sí tienen interacción que aporta significativamente a la varianza total de la nitidez de impresión, por tanto, hay que ser cautos en la interpretación de los test de significancia sobre las componentes de varianza aportadas individualmente por cada factor. Es más, recuerde que algunos estadísticos recomiendan no llevar a cabo tales pruebas si la componente de varianza $\sigma_{\alpha\beta}^2$ es significativa, pero procedamos con cada una de ellas como se pide en el literal d).

Significancia de la componente de varianza del factor Temperatura de impresión

$$H_0 : \sigma_{\alpha}^2 = 0 \text{ vs. } H_1 : \sigma_{\alpha}^2 > 0$$

El estadístico de la prueba a usar es $F_1 = \frac{MSA}{MS(AB)} = 8.18 \stackrel{\text{bajo } H_0}{\sim} f_{3,6}$ y como $VP = P(f_{3,6} > F_1) = 0.0153$ es pequeño, se concluye que la variación de temperatura aporta significativamente a la varianza total de la nitidez de las impresiones.

Significancia de la componente de varianza debida al factor Tinta

$$H_0 : \sigma_{\beta}^2 = 0 \text{ vs. } H_1 : \sigma_{\beta}^2 > 0$$

El estadístico de la prueba a usar es $F_2 = \frac{MSB}{MS(AB)} = 3.90 \stackrel{\text{bajo } H_0}{\sim} f_{2,6}$ y como $VP = P(f_{2,6} > F_2) = 0.0821$, Se concluye que la variación del tipo de tinta bajo una inter-

acción significativa, no es posible medirla efectivamente, hay enmascaramiento!! y por tanto tenga en cuenta que el hecho de no rechazar H_0 no significa que esta componente de varianza sea igual a cero.

Estimaciones de ls componentes de varianza

`lmer()` de la librería `lme4` estima las componentes de varianza (y sus raíces cuadradas) como también la media global:

Salida R 8.8.3. Resultados para modelo factorial de dos factores aleatorios, con función `lmer()`.

```
Linear mixed model fit by REML ['lmerMod']
Formula: nitidez ~ 1 + (1 | temperatura) + (1 | tinta) + (1 | temperatura:tinta)
REML criterion at convergence: 438
Random effects:
Groups          Name          Variance Std.Dev.
temperatura:tinta (Intercept) 1070.7   32.72
temperatura      (Intercept) 2714.1   52.10
tinta            (Intercept) 823.1    28.69
Residual                            252.2   15.88
Number of obs: 48, groups:  temperatura:tinta, 12; temperatura, 4; tinta, 3
Fixed effects:
              Estimate Std. Error t value
(Intercept)   238.44      32.36    7.368
```

De los análisis previos tenemos a partir del modelo con interacción las estimaciones de la Tabla 8.15. Como puede observarse, de la variación total, cerca del 94.81 % es debida a la variación de los factores Temperatura y Tinta y de la interacción, y el 55.84 % específicamente es debida a la variación de la temperatura de impresión, y un 22.03 % es causada por la interacción de los factores. A pesar que para lograr una menor variabilidad en la nitidez es necesario considerar simultáneamente a los dos factores por la presencia de la interacción, como la mayor proporción de varianza es debida a la temperatura, esto nos indica que a partir de la reducción de la variabilidad en esta fuente de variación es posible lograr una menor varianza de la nitidez.

Tabla 8.15: Estimaciones en el ejemplo

Componente de varianza debida a	Estimación	Porcentaje
La variación de la temperatura de impresión	$\hat{\sigma}_{\alpha}^2 = \frac{MSA - MS(AB)}{nb} = \frac{37105 - 4535.175510}{4 \times 3} = 2714.152$	$\frac{\hat{\sigma}_{\alpha}^2}{\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 55.84 \%$
La variación de los tipos de tinta	$\hat{\sigma}_{\beta}^2 = \frac{MSB - MS(AB)}{na} = \frac{17705 - 4535.175510}{4 \times 4} = 823.114$. Esta estimación puede ser dudosa bajo una interacción altamente significativa que encubrió la significancia de esta fuente de variabilidad!	$\frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 16.94 \%$
La variación en la interacción	$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS(AB) - MSE}{n} = \frac{4535.175510 - 252.187647}{4} = 1070.747$	$\frac{\hat{\sigma}_{\alpha\beta}^2}{\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 22.03 \%$
La variación aleatoria	$\hat{\sigma}^2 = MSE = 252.188$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 5.19 \%$
Varianza total	$\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2 = 4860.201$ Esta estimación puede ser dudosa bajo una interacción altamente significativa que encubrió la significancia de σ_{β}^2	100 %

8.8.4. Validación de supuestos sobre los errores

Considere las gráficas en la Figura 8.8. En éstas usamos los residuos del modelo ajustado como si fuese de efectos fijos, para el cálculos de residuos y valores ajustados (residuales internamente estudentizados vs. temperatura, residuales internamente estudentizados vs. tipo de tinta y residuales internamente estudentizados vs. valores ajustados) es clara la presencia de tres observaciones outliers (las que tienen residuales que están por fuera de las bandas ± 2), pero aún sin considerar estos residuales, parece haber cierta desviación del supuesto de varianza constante para los errores. En particular, la gráfica de residuales internamente estudentizados vs. temperatura muestra una menor dispersión en los niveles 1 y 4 de este factor comparados con sus niveles 2 y 3. En la gráfica de residuales internamente estudentizados vs. tinta, los residuales del nivel 1 presentan menor dispersión en relación a los otros dos niveles. La gráfica residuales internamente estudentizados vs. valores

ajustados no muestra problemas con la varianza como en las dos gráficas anteriores. Sin embargo, la violación del supuesto de varianza constante en este problema no es tan grave. En cuanto al supuesto de normalidad de los errores de este modelo, el test de Shapiro Wilk con un valor P de 0.5312, indica que tal supuesto es razonable pero ¿qué dice la gráfica de probabilidad normal?

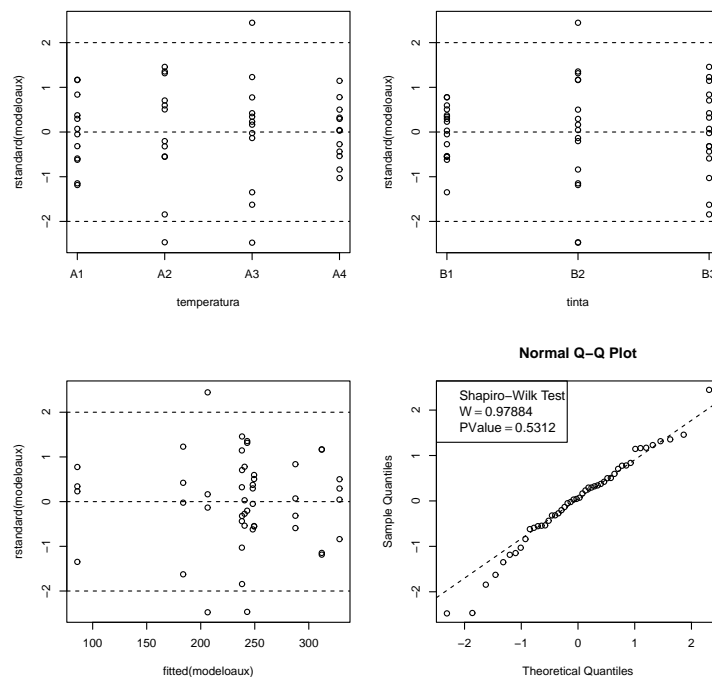


Figura 8.8: Gráficas para evaluación de supuestos sobre los errores, en datos del Ejemplo 8.8

Código R 8.8.1. Programa R usado en Ejemplo 8.8.

```
library(daewr)
library(lme4)
library(MASS)

impresion=data.frame(temperatura=factor(rep(c("A1","A2","A3","A4"),each=12)),
                     tinta=factor(rep(c("B1","B2","B3"),times=16)),nitidez=scan())

252.23 296.20 288.59
253.32 327.97 299.12
239.63 295.67 283.26
247.48 328.08 279.47
```

```

241.72 261.04 258.29
256.31 209.03 212.89
241.85 261.59 233.84
257.55 240.16 247.95
88.95 204.56 183.43
90.48 172.36 161.39
67.24 240.02 200.66
96.44 208.66 189.55
233.15 329.03 242.48
236.80 316.88 232.05
251.28 335.31 253.80
240.96 332.46 223.89

```

```

impresion
attach(impresion)
#GRÁFICOS PARA ANALIZAR LOS DATOS EXPERIMENTALES
mediastemperatura=sapply(split(nitidez,temperatura),mean)
mediastinta=sapply(split(nitidez,tinta),mean)
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
boxplot(nitidez~temperatura,xlab="Temperatura",boxwex=0.4,col=c("bisque","blue","gray","orange"))
lines(1:4,mediastemperatura,type="b",pch=19,col=2,lty=2,lwd=2)
boxplot(nitidez~tinta,xlab="Tinta",boxwex=0.4,col=c("bisque","blue","gray"))
lines(1:3,mediastinta,type="b",pch=19,col=2,lty=2,lwd=2)
interaction.plot(temperatura,tinta,nitidez,type="b",lwd=2,col=1:3,pch=1:3)
interaction.plot(tinta,temperatura,nitidez,type="b",lwd=2,col=1:4,pch=1:4)

#sumas de cuadrados y cuadrado medios
modelo=aov(nitidez~Error(temperatura*tinta))
summary(modelo)

#componentes de varianza
modelob=lmer(nitidez~1+(1|temperatura)+(1|tinta)+(1|temperatura:tinta))
summary(modelob)

#Creación función de usuario para calcular de forma completa y correcta tabla ANOVA
Anovafactorialaleatoriosconinteraccion=function(modelo){
  dfA=anova(modelo)[1][1,]
  dfB=anova(modelo)[1][2,]
  dfi=anova(modelo)[1][3,]
  dfe=anova(modelo)[1][4,]
  SSA=anova(modelo)[2][1,]

```

```

SSB=anova(modelo)[2][2,]
SSAB=anova(modelo)[2][3,]
SSE=anova(modelo)[2][4,]
MSA=anova(modelo)[3][1,]
MSB=anova(modelo)[3][2,]
MSAB=anova(modelo)[3][3,]
MSE=anova(modelo)[3][4,]
F1=round(MSA/MSAB,digits=2)
F2=round(MSB/MSAB,digits=2)
F3=round(MSAB/MSE,digits=2)
vp1=round(pf(F1,df1=dfA,df2=dfi,lower.tail=F),digits=5)
vp2=round(pf(F2,df1=dfB,df2=dfi,lower.tail=F),digits=5)
vp3=round(pf(F3,df1=dfi,df2=dfe,lower.tail=F),digits=5)
nombres=attr(attr(modelo$model,"terms"),"term.labels")
anova1=data.frame(Source=nombres,DF=c(dfA,dfB,dfi),"Sum\ Sq"=c(SSA,SSB,SSAB),"Mean Sq"=c(MSA,MSB,MSAB),
                  "F\ Value"=c(F1,F2," "),"P\ value"=c(vp1,vp2," "))
anova2=data.frame(Source=c(nombres[3],"Error"),DF=c(dfi,dfe),"Sum\ Sq"=c(SSAB,SSE),"Mean Sq"=c(MSAB,MSE),
                  "F\ Value"=c(F3," "),"P\ value"=c(vp3," "))
cat("Tests of Hypotheses for Random Model Analysis of Variance","\n")
cat("Dependent Variable:",names(modelo$model)[1],"","\n")
print(anova1)
cat("\n")
print(anova2)
}

modeloaux=aov(nitidez~temperatura*tinta) #correr modelo como si fuese efectos fijos
Anovafactorialaleatoriosconinteraccion(modelo=modeloaux) #Tabla ANOVA correcta

#Validar supuestos sobre errores usando residuos estandarizados
#del modelo ajustado como si fuese factorial de efectos fijos
shapiro.test(rstandard(modeloaux))
#OBTENIENDO GRÁFICOS DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS,
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
stripchart(rstandard(modeloaux)~temperatura,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,
           xlab="temperatura")
abline(h=c(-2,0,2),lty=2)
stripchart(rstandard(modeloaux)~tinta,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,xlab="tinta")
abline(h=c(-2,0,2),lty=2)
plot(fitted(modeloaux),rstandard(modeloaux),ylim=c(-2.5,2.5))
abline(h=c(-2,0,2),lty=2)
qqnorm(rstandard(modeloaux))

```

```
qqline(rstandard(modeloaux),lty=2)
legend("topleft",
      legend=c("Shapiro-Wilk Test",expression(W==0.97884),expression(PValue== 0.5312)),cex=1.1)

detach(impresion)
```

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado

Considere el caso de un experimento factorial, con dos factores A y B, en el cual uno de los dos (digamos A) es de efectos fijos y el otro (digamos B) es de efectos aleatorios. Por tanto cada tratamiento resultante de la combinación de niveles de A con niveles de B hace parte también de una muestra aleatoria de posibles tratamientos. Respecto al factor de efectos fijos el interés es estudiar los efectos de sus a niveles (efectos fijos) sobre la respuesta promedio de la variable de interés, en tanto que del factor de efectos aleatorios se toman aleatoriamente b niveles con el fin de estudiar el aporte de dicho factor a la varianza total de dicha variable de interés, $V(Y_{ijk})$.

Una vez elegidos aleatoriamente los b niveles del factor B, a cada tratamiento o combinación de niveles de los dos factores, se asignan en forma completamente aleatorizada, n unidades experimentales obteniéndose por tanto de cada tratamiento n observaciones de la variable respuesta. En total habrán $N = abn$ observaciones. Un tabla de entrada de datos del experimento será similar a la presentada en el caso del diseño factorial de efectos fijos, sin embargo, el modelo con interacción de los factores para la respuesta k -ésima en el tratamiento $A_i B_j$ (nivel i del factor A combinado con nivel j del factor B), es el que se presenta a continuación:

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n \quad (8.32)$$

donde

- μ es la media global de la respuesta
- α_i es el efecto fijo principal del nivel i factor A sobre la respuesta media

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado 269

- B_j es el efecto aleatorio del factor B sobre la respuesta media
- $(\alpha B)_{ij}$ es el efecto aleatorio de la interacción del nivel i del factor A con el factor B
- E_{ijk} Error aleatorio del modelo

Dado que el factor B tiene efectos aleatorios, la interacción entre A y B también es una variable aleatoria.

Existen dos modelos ANOVA's de efectos mixtos: Modelo mixto restringido y modelo mixto no restringido. Veamos cada uno:

8.9.1. Modelo Mixto restringido

De los dos modelos éste es el más general en el sentido de que en él las covarianzas entre observaciones en el mismo nivel del factor aleatorio (o sea, entre Y_{ijk} y $Y_{i'jk'}$, con $i \neq i'$) puede ser positiva o negativa, mientras que en el mixto no restringido estas covarianzas sólo pueden ser positivas, [10].

Sus supuestos y restricciones son como sigue:

- $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2), \forall i, j, k,$
- $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2), \forall j,$
- $(\alpha B)_{ij} \sim N\left(0, \frac{a-1}{a} \sigma_{\alpha\beta}^2\right)$, con $COV((\alpha B)_{ij}, (\alpha B)_{i'j}) = -\frac{1}{a} \sigma_{\alpha\beta}^2$, para $i \neq i'$, por tanto, los efectos de interacción en el mismo nivel del factor aleatorio no son independientes y están negativamente correlacionados, En cambio, tenemos que $COV((\alpha B)_{ij}, (\alpha B)_{i'j'}) = 0$, para $j \neq j'$.
- ε_{ijk}, B_j y $(AB)_{ij}$ son mutuamente independientes.
- Restricciones $\sum_{i=1}^a \alpha_i = 0, \sum_{i=1}^a (\alpha B)_{ij} = 0, j = 1, \dots, b$. Esta última restricción es la que causa la covarianza entre los efectos aleatorios $(\alpha B)_{ij}$ y $(\alpha B)_{i'j}$ para $i \neq i'$.

Nota 8.8. La salida media en el tratamiento $AiBj$ es $M_{ij} = \mu + \alpha_i + B_j + (\alpha B)_{ij}$, es decir, es una variable aleatoria! Similarmente las medias por nivel del factor B, también son

variables aleatorias. Por tanto, no tiene sentido ni es de interés, hacer inferencias sobre medias de tratamientos ni medias de niveles del factor B. Por otra parte, los efectos de A son fijos, interesa entonces inferir sobre las medias de niveles de A.

Nota 8.9. También note que $\sigma_{\alpha\beta}^2$ no es la varianza de los términos de interacción, sino que es proporcional a tal varianza. La razón de que la varianza de las interacciones sea $\frac{a-1}{a}\sigma_{\alpha\beta}^2$ es que ésta hace que los valores esperados de los cuadrados medios sean expresiones relativamente fáciles.

TENGA EN CUENTA QUE: La variable respuesta bajo el modelo anterior tiene las siguientes características:

- Y_{ijk} es una variable aleatoria normal, pues es la suma de dos constantes más tres variables aleatorias normales independientes.
- $E(Y_{ijk}) = \mu + \alpha_i$, y $V(Y_{ijk}) = \sigma_{\beta}^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2 + \sigma^2$ que corresponde a la varianza total.
- Al avanzar los ensayos aleatorios, diferentes respuestas son independientes si ellas no pertenecen al mismo nivel j del factor aleatorio B (OJO asumiendo una tabla de datos como la presentada en la Tabla 8.2):
 - $COV(Y_{ijk}, Y_{i'jk'}) = \sigma_{\beta}^2 - \frac{1}{a}\sigma_{\alpha\beta}^2$, para $i \neq i'$, es decir, observaciones en la misma columna (mismo nivel de B) pero diferente fila (diferente nivel de A), tienen covarianza igual a $\sigma_{\beta}^2 - \frac{1}{a}\sigma_{\alpha\beta}^2$
 - $COV(Y_{ijk}, Y_{ijk'}) = \sigma_{\beta}^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2$ para $k \neq k'$, es decir, observaciones en la misma celda (mismo tratamiento) tienen covarianza $\sigma_{\beta}^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2$
 - $COV(Y_{ijk}, Y_{i'j'k'}) = 0$, para todo $j \neq j'$, es decir, observaciones que están en distinta columna o nivel j de B, son independientes.

Parámetros de interés

El problema que enfrentamos en este modelo es determinar si las componentes de varianza de los efectos aleatorios de la interacción y del factor B son significativas, y

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado 271

además determinar si los efectos fijos del factor A son significativos sobre la respuesta promedio. Por tanto los parámetros de interés son los siguientes:

- σ^2 la variación entre muestras de tratamientos o celdas.
- σ_β^2 la variación del desempeño del factor B, o variación por cambios del factor B.
- $\frac{a-1}{a}\sigma_{\alpha\beta}^2$ La variación por cambios en las interacciones; su existencia implica que las diferencias entre media de los niveles del factor A varían con B, es decir, el desempeño relativo en los niveles del factor A en relación a cambios en el factor B no varía de forma congruente con dichos cambios.
- Los efectos principales α_i e incluso las medias $\mu_{i\bullet}$ de la respuesta en cada nivel de A. Sin embargo las inferencias sobre estos parámetros es más complicada que en un experimento de efectos fijos, debido a que ciertos pares de observaciones están correlacionadas en el modelo mixto, lo cual dificulta las estimaciones de las varianzas de los estimadores de efectos principales y medias de nivel de A.

ANOVA

Para el ANOVA de nuevo tendremos en cuenta los siguientes estadísticos muestrales, con

- Y_{ijk} es la respuesta en la k-esima réplica del tratamiento A_iB_j
- $\bar{Y}_{i\bullet\bullet}$ es la media muestral de las nb réplicas en el nivel i del factor A, es decir
$$\bar{Y}_{i\bullet\bullet} = \frac{1}{nb} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet},$$
 que estima a la media $\mu_{i\bullet}$.
- $\bar{Y}_{\bullet j\bullet}$ es la media muestral de las na réplicas en el nivel j del factor B, es decir,
$$\bar{Y}_{\bullet j\bullet} = \frac{1}{na} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}.$$
- con $\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, la media muestral en el tratamiento A_iB_j ,
- $\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$, la media muestral global de las abn observaciones y que estima a la media global μ .

El ANOVA descompone la variabilidad total de la misma manera que el caso con efectos fijos, es decir:

$$SST = SSA + SSB + SS(AB) + SSE$$

Las fórmulas de estas sumas de cuadrados y sus grados de libertad son las mismas que en el caso de dos factores de efectos fijos bajo interacción. Sin embargo, los valores esperados de los respectivos cuadrados medios cambian respecto a los dados en los casos de efectos fijos y de efectos aleatorios. Vea con cuidado la Tabla 8.16.

Pruebas de hipótesis

En su orden, deben realizarse los siguientes tests, con γ como nivel de significancia:

Significancia de la componente de varianza debida a la interacción

$$H_0 : \sigma_{\alpha\beta}^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_{\alpha\beta}^2 > 0. \quad (8.33)$$

Note que si H_0 fuese cierta, entonces $E[MS(AB)] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_3 = \frac{MS(AB)}{MSE} \stackrel{H_0}{\sim} f_{(a-1)(b-1), ab(n-1)}. \quad (8.34)$$

Rechazar H_0 a un nivel de significancia γ si $F_3 > f_{\gamma, (a-1)(b-1), ab(n-1)}$, o con valor P si $P(f_{(a-1)(b-1), ab(n-1)} > F_3)$ es pequeño.

Si la interacción es significativa (es decir, se rechaza H_0), en general no tiene sentido hacer pruebas generalizadas sobre la significancia de los efectos del factor fijo A, o sobre la componente de varianza debida al factor B. Una componente $\sigma_{\alpha\beta}^2 > 0$ implica que las diferencias entre las medias del factor A varían con B. Sólo en ausencia de interacción sería adecuada una prueba de la hipótesis nula de que no hay diferencia entre las medias marginales del factor de efectos fijos (o equivalentemente, que todos sus efectos son simultáneamente cero). Sin embargo, Con fines ilustrativos se muestran los siguientes tests en presencia de interacción:

Tabla 8.16: ANOVA en un DCA balanceado, con dos factores de efectos Mixtos (A factor de efectos fijos y B factor de efectos aleatorios) con interacción, modelo restringido

Fórmulas
$dfi = (a - 1)(b - 1)$, los grados de libertad de la interacción AB
$dfe = (abn - 1) - (a - 1) - (b - 1) - (a - 1)(b - 1) = ab(n - 1)$, los grados de libertad del error.
$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$
$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SSE = SST - SSA - SSB - SSAB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij\bullet})^2$
donde,
$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, media muestral en tratamientos $A_i B_j$
$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A.
$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B.
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ , o bien
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.

Significancia de los efectos principales fijos del factor A

$$H_0 : \alpha_1 = \alpha_2 = \dots, \alpha_a = 0 \quad \text{vs.} \quad H_1 : \text{al menos un } \alpha_i \neq 0. \quad (8.35)$$

Note que si H_0 fuese cierta, entonces $E[MSA] = \sigma^2 + n\sigma_{\alpha\beta}^2$ que también corresponde al $E[MS(AB)]$. Por tanto el estadístico de la prueba a usar es

$$F_1 = \frac{MSA}{MS(AB)} \stackrel{H_0}{\sim} f_{a-1, (a-1)(b-1)}. \quad (8.36)$$

Rechazar H_0 a un nivel de significancia γ si $F_1 > f_{\gamma, a-1, (a-1)(b-1)}$, o con valor P si $P(f_{a-1, (a-1)(b-1)} > F_1)$ es pequeño.

Significancia de la componente de varianza del factor B

$$H_0 : \sigma_\beta^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\beta^2 > 0. \quad (8.37)$$

Note que si H_0 fuese cierta, entonces $E[MSB] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_2 = \frac{MSB}{MSE} \stackrel{H_0}{\sim} f_{b-1, ab(n-1)}. \quad (8.38)$$

Rechazar H_0 a un nivel de significancia γ si $F_2 > f_{\gamma, b-1, ab(n-1)}$, o con valor P si $P(f_{b-1, ab(n-1)} > F_2)$ es pequeño.

Estimaciones

Interesa estimar σ_β^2 , $\frac{a-1}{a}\sigma_{\alpha\beta}^2$, α_i y $\mu_{i\bullet}$. Las estimaciones de las componentes de varianza se determinan igualando los cuadrados medios esperados a los correspondientes cuadrados medios observados y se despejan los valores desconocidos de las componentes. Las estimaciones de las tres componentes de la varianza son:

Nota 8.10. *Bajo interacción significativa, puede ocurrir que la componente de varianza estimada $\hat{\sigma}_\beta^2$ resulte negativa. Obviamente no tiene sentido dar como resultado una varianza negativa. Así que es mejor no hacer inferencias ni estimaciones de esta componente bajo interacción significativa.*

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado 275

Tabla 8.17: Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos mixtos con interacción, modelo restringido

Fuente	Componente de varianza estimada	Proporción respecto a la varianza total
B	$\hat{\sigma}_\beta^2 = \frac{MSB - MSE}{na}$	$\frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + \frac{a-1}{a}\hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
AB	$\frac{a-1}{a}\hat{\sigma}_{\alpha\beta}^2 = \left(\frac{a-1}{a}\right) \left[\frac{MS(AB) - MSE}{n} \right]$	$\frac{\frac{a-1}{a}\hat{\sigma}_{\alpha\beta}^2}{\hat{\sigma}_\beta^2 + \frac{a-1}{a}\hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
Error	$\hat{\sigma}^2 = MSE$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_\beta^2 + \frac{a-1}{a}\hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$

En cuanto a los efectos y medias de nivel del factor A, en un diseño balanceado se tiene lo siguiente

Tabla 8.18: Estimadores de efectos fijos y medias del factor A, factorial dos factores de efectos mixtos con interacción

Parámetro	Estimador	Varianza	Error estándar	I.C (1 - γ)100 %
α_i	$\hat{\alpha}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$	$V(\hat{\alpha}_i) = \frac{(a-1)E(MS(AB))}{abn}$	$S_{\hat{\alpha}_i} = \sqrt{\frac{(a-1)MS(AB)}{abn}}$	$\hat{\alpha}_i \pm t_{\gamma/2, (a-1)(b-1)} S_{\hat{\alpha}_i}$
$\mu_{i\bullet}$	$\hat{\mu}_i = \bar{Y}_{i\bullet\bullet}$	$V(\hat{\mu}_{i\bullet}) = c_1 E(MS(AB)) + c_2 E(MSB)$	$S_{\hat{\mu}_{i\bullet}} = \sqrt{c_1 MS(AB) + c_2 MSB}$	$\hat{\mu}_{i\bullet} \pm t_{\gamma/2, \nu} S_{\hat{\mu}_{i\bullet}}$ aprox.
Donde: $c_1 = \frac{a-1}{abn}$, $c_2 = \frac{1}{abn}$, y $\nu \simeq \frac{\left(\frac{a-1}{abn}MS(AB) + \frac{1}{abn}MSB\right)^2}{\frac{\left(\frac{a-1}{abn}MS(AB)\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{1}{abn}MSB\right)^2}{(b-1)}}$				

Contrastes de medias $\mu_{i\bullet}$

Para contrastes entre medias del factor A, se tiene lo siguiente: Sea $W = \sum_{i=1}^a c_i \mu_{i\bullet}$ con $\sum_{i=1}^a c_i = 0$; su estimador corresponde a $\widehat{W} = \sum_{i=1}^a c_i \bar{Y}_{i\bullet\bullet}$ el cual tiene como varianza,

$$V(\widehat{W}) = \frac{E(MS(AB))}{bn} \times \sum_{i=1}^a c_i^2 \quad (8.39)$$

por tanto un intervalo de confianza del (1 - γ)100 % para W es,

$$\widehat{W} \pm t_{\gamma/2, (a-1)(b-1)} \times \sqrt{\frac{MS(AB)}{bn} \sum_{i=1}^a c_i^2} \quad (8.40)$$

y el estadístico para la prueba $H_0 : W = 0$ vs. $H_1 : W \neq 0$ es,

$$T_0 = \frac{\widehat{W}}{\sqrt{\frac{MS(AB)}{bn} \sum_{i=1}^a c_i^2}} \stackrel{H_0}{\sim} t_{(a-1)(b-1)} \quad (8.41)$$

Para las comparaciones múltiples de Tukey de las medias de A, los I.C simultáneos del $(1 - \gamma)100\%$ están dados por:

$$(\mu_{i\bullet} - \mu_{j\bullet}) \in (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet}) \pm q(\gamma, a, (a-1)(b-1)) \times \sqrt{\frac{MS(AB)}{bn}} \quad (8.42)$$

Validación de supuestos

Podemos chequear los supuestos sobre los términos de error ε_{ijk} reemplazando temporalmente todos los efectos aleatorios por efectos fijos, calculando los residuales y examinando los gráficos de residuales y los test de normalidad de la manera como se hizo en el caso del experimento factorial de efectos fijos. Sin embargo chequear los supuestos de cada uno de los efectos aleatorios B_j , y $(\alpha B)_{ij}$, no es fácil, dado que en un modelo factorial por lo general hay pocos niveles de cada factor observado, y los promedios de celdas no son independientes. Consecuentemente, en general se omite los chequeos para los supuestos de tales efectos aleatorios, y se espera que cualquier problema severo sea detectado en el análisis de los residuales.

8.9.2. Modelo mixto no restringido

Sus supuestos y restricciones son,

- $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ para todo i, j, k ,
- $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$ para todo j ,
- $(\alpha B)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$ para todo i, j . (he aquí una de las diferencias con el modelo restringido).
- Además las variables aleatorias ε_{ijk} , B_j y $(\alpha B)_{ij}$ son mutuamente independientes.
- Restricciones: La única restricción en este modelo es que $\sum_{i=1}^a \alpha_i = 0$. No hay restricciones sobre las interacciones, lo cual constituye la segunda diferencia con los modelos mixtos restringidos.

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado 277

NOTE QUE: La salida media en el tratamiento $A_i B_j$ es $M_{ij} = \mu + \alpha_i + B_j + (\alpha\beta)_{ij}$, es decir, es una variable aleatoria! Similarmente las medias por nivel del factor B, también son variables aleatorias. Por tanto no tiene sentido ni es de interés, hacer inferencias sobre medias de tratamientos ni medias de niveles del factor B. Por otra parte, los efectos de A son fijos, interesa entonces inferir sobre las medias de niveles de A. También note que $\sigma_{\alpha\beta}^2$ es la varianza de los términos de interacción.

TENGA EN CUENTA QUE: La variable respuesta bajo el modelo anterior tiene las siguientes características:

- Y_{ijk} es una variable aleatoria normal, pues es la suma de dos constantes más tres variables aleatorias normales independientes.
- $E(Y_{ijk}) = \mu + \alpha_i$, y $V(Y_{ijk}) = \sigma_{\beta}^2 + \sigma_{\alpha\beta}^2 + \sigma^2$ que corresponde a la varianza total.
- Al avanzar los ensayos aleatorios, diferentes respuestas son independientes si ellas no pertenecen al mismo nivel j del factor aleatorio B (OJO asumiendo una tabla de datos como la presentada en la Tabla 8.2):
 - $COV(Y_{ijk}, Y_{i'jk'}) = \sigma_{\beta}^2$, para $i \neq i'$, es decir, observaciones en la misma columna (mismo nivel de B) pero diferente fila (diferente nivel de A), tienen covarianza igual a σ_{β}^2 .
 - $COV(Y_{ijk}, Y_{ijk'}) = \sigma_{\beta}^2 + \sigma_{\alpha\beta}^2$ para $k \neq k'$, es decir, observaciones en la misma celda (mismo tratamiento) tienen covarianza $\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2$
 - $COV(Y_{ijk}, Y_{i'j'k'}) = 0$, para todo $j \neq j'$, es decir, observaciones que están en distinta columna o nivel j de B, son independientes.

Nota 8.11. Cuando el diseño es desbalanceado, es decir, no hay igual número de réplicas por tratamientos, el modelo mixto no restringido es preferido sobre el restringido.

Parámetros de interés

El problema que enfrentamos en este modelo es determinar si las componentes de varianza de los efectos aleatorios de la interacción y del factor B son significativas, y

además determinar si los efectos fijos principales del factor A son significativos sobre la respuesta promedio. Por tanto los parámetros de interés son los siguientes:

- σ^2 la variación entre muestras de tratamientos o celdas.
- σ_β^2 la variación del desempeño del factor B, o variación por cambios del factor B.
- $\sigma_{\alpha\beta}^2$ La variación por cambios en las interacciones; su existencia implica que las diferencias entre media de los niveles del factor A varían con B, es decir, el desempeño relativo en los niveles del factor A en relación a cambios en el factor B no varía de forma congruente con dichos cambios.
- Los efectos α_i e incluso las medias $\mu_{i\bullet}$ de la respuesta en cada nivel de A. Sin embargo como ya se ha dicho, las inferencias sobre estos parámetros es más complicada que en un experimento de efectos fijos, debido a que ciertos pares de observaciones están correlacionadas en el modelo mixto, lo cual dificulta las estimaciones de las varianzas de los estimadores de efectos principales y medias de nivel de A.

ANOVA

Con las mismas definiciones de las diferentes medias muestrales dadas previamente, de nuevo descomponemos la variabilidad total en la forma

$$SST = SSA + SSB + SS(AB) + SSE$$

Observe en la Tabla 8.19 cómo cambian las esperanzas de los cuadrados medios y los estadísticos F en este modelo con respecto al mixto no restringido. Si observa con cuidado, la principal diferencia radica en el $E[MSB]$ y el correspondiente estadístico F, en donde ahora el MSB es dividido por el $MS(AB)$.

Tabla 8.19: ANOVA en un DCA balanceado, con dos factores de efectos Mixtos (A factor de efectos fijos y B factor de efectos aleatorios) con interacción, modelo no restringido

Fórmulas
$dfi = (a - 1)(b - 1)$, los grados de libertad de la interacción AB
$dfe = (abn - 1) - (a - 1) - (b - 1) - (a - 1)(b - 1) = ab(n - 1)$, los grados de libertad del error.
$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SS(AB) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$
$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$
$SSE = SST - SSA - SSB - SSAB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij\bullet})^2$
donde,
$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, media muestral en tratamientos $A_i B_j$
$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A.
$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B.
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ , o bien
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,
$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.

Pruebas de hipótesis

En su orden, deben realizarse los siguientes tests, con γ como nivel de significancia:

Significancia de la componente de varianza debida a la interacción

$$H_0 : \sigma_{\alpha\beta}^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_{\alpha\beta}^2 > 0. \quad (8.43)$$

Note que si H_0 fuese cierta, entonces $E[MS(AB)] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_3 = \frac{MS(AB)}{MSE} \stackrel{H_0}{\sim} f_{(a-1)(b-1), ab(n-1)}. \quad (8.44)$$

Rechazar H_0 a un nivel de significancia γ si $F_3 > f_{\gamma, (a-1)(b-1), ab(n-1)}$, o con valor P si $P(f_{(a-1)(b-1), ab(n-1)} > F_3)$ es pequeño.

De nuevo, bajo interacción significativa no se recomienda en general hacer inferencias marginales sobre los factores A y B, sin embargo si se hacen, proceda con precaución en las conclusiones.

Significancia de los efectos fijos del factor A

$$H_0 : \alpha_1 = \alpha_2 = \dots, \alpha_a = 0 \quad \text{vs.} \quad H_1 : \text{al menos un } \alpha_i \neq 0. \quad (8.45)$$

Note que si H_0 fuese cierta, entonces $E[MSA] = \sigma^2 + n\sigma_{\alpha\beta}^2$ que también corresponde al $E[MS(AB)]$. Por tanto el estadístico de la prueba a usar es

$$F_1 = \frac{MSA}{MS(AB)} \stackrel{H_0}{\sim} f_{a-1, (a-1)(b-1)}. \quad (8.46)$$

Rechazar H_0 a un nivel de significancia γ si $F_1 > f_{\gamma, a-1, (a-1)(b-1)}$, o con valor P si $P(f_{a-1, (a-1)(b-1)} > F_1)$ es pequeño.

Significancia de la componente de varianza del factor B

$$H_0 : \sigma_{\beta}^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_{\beta}^2 > 0. \quad (8.47)$$

8.9. Modelo factorial de dos factores con efectos mixtos con interacción, en un DCA balanceado 281

Note que si H_0 fuese cierta, entonces $E[MSB] = \sigma^2 + n\sigma_{\alpha\beta}^2$ que también corresponde al $E[MS(AB)]$. Por tanto el estadístico de la prueba a usar es

$$F_2 = \frac{MSB}{MS(AB)} \stackrel{H_0}{\sim} f_{b-1, (a-1)(n-1)}. \quad (8.48)$$

Rechazar H_0 a un nivel de significancia γ si $F_2 > f_{\gamma, b-1, (a-1)(n-1)}$, o con valor P si $P(f_{b-1, (a-1)(n-1)} > F_2)$ es pequeño.

Estimaciones

Interesa estimar σ_β^2 , $\sigma_{\alpha\beta}^2$, α_i y $\mu_{i\bullet}$. Las estimaciones de las componentes de varianza se determinan igualando los cuadrados medios esperados a los correspondientes cuadrados medios observados y se despejan los valores desconocidos de las componentes. Las estimaciones de las tres componentes de la varianza son:

Tabla 8.20: Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos mixtos con interacción, modelo no restringido

Fuente	Componente de varianza estimada	Proporción respecto a la varianza total
B	$\hat{\sigma}_\beta^2 = \frac{MSB - MS(AB)}{na}$	$\frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
AB	$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS(AB) - MSE}{n}$	$\frac{\hat{\sigma}_{\alpha\beta}^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$
Error	$\hat{\sigma}^2 = MSE$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2}$

Nota 8.12. De nuevo tenga en cuenta que bajo interacción significativa, puede ocurrir que la componente de varianza estimada $\hat{\sigma}_\beta^2$ resulte negativa y como no tiene sentido dar como resultado una varianza negativa, es mejor no hacer inferencias ni estimaciones de esta componente bajo interacción significativa.

En cuanto a los efectos y medias de nivel del factor A en un diseño balanceado de efectos mixtos modelo no restringido, se obtienen los mismos resultados reportados previamente en el caso mixto restringido, es decir, los estimadores, sus varianzas y sus I.C siguen las ecuaciones presentadas en la Tabla 8.18. De la misma forma, las inferencias sobre contrastes de medias del factor A y las comparaciones múltiples

de Tukey, siguen exactamente las fórmulas previamente dadas en el modelo mixto restringido, es decir, las ecuaciones (8.39), (8.40), (8.41) en los contrastes y (8.42) en las comparaciones por I.C en Tukey, aplican tal cual en el modelo mixto no restringido.

Validación de supuestos

Al igual que en el modelo restringido, podemos chequear los supuestos sobre los términos de error ε_{ijk} reemplazando temporalmente todos los efectos aleatorios por efectos fijos, calculando los residuales y examinando sus gráficos y el test de normalidad de la manera como se hizo en el caso del experimento factorial de efectos fijos. También aquí como en todos los modelos vistos con efectos aleatorios, en general se omiten los chequeos para los supuestos de los efectos aleatorios presentes en el modelo, y se espera que cualquier problema severo sea detectado en el análisis de los residuales.

8.10. Modelo factorial de dos factores de efectos mixtos sin interacción, en un DCA balanceado

En caso de que los datos obtenidos presenten evidencia respecto a la no significancia de la componente de varianza debida al término de interacción, el modelo factorial debe replantearse de la siguiente manera:

$$Y_{ijk} = \mu + \alpha_i + B_j + \varepsilon_{ijk} \quad (8.49)$$

Dado que no hay interacción sólo se postula un modelo. los supuestos y restricciones son como sigue,

- $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ para todo i, j, k ,
- $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$ para todo j ,
- Las variables aleatorias ε_{ijk} , B_j son mutuamente independientes.
- Restricciones: $\sum_{i=1}^a \alpha_i = 0$.

TENGA EN CUENTA QUE:

- Y_{ijk} es una variable aleatoria normal, pues es la suma de una constante más dos variables aleatorias independientes.
- $E(Y_{ijk}) = \mu + \alpha_i$ y $V(Y_{ijk}) = \sigma_\beta^2 + \sigma^2$ que corresponde a la varianza total.
- Al avanzar los ensayos aleatorios, diferentes respuestas son independientes, excepto aquellas que correspondan al mismo nivel j del factor B las cuales están correlacionadas porque contienen términos aleatorios en común:
 - $COV(Y_{ijk}, Y_{ijk}) = \sigma_\beta^2$ para $i \neq i'$, es decir, observaciones en la misma columna (mismo nivel de B) tienen covarianza σ_β^2 .
 - $COV(Y_{ijk}, Y_{ijk}) = 0$ para $j \neq j'$, es decir, observaciones que están en distinto nivel de B son independientes.

Obviamente en el ANOVA la suma de cuadrados debida al término de error, es diferente en un modelo con interacción respecto a un modelo sin interacción.

8.10.1. Parámetros de interés

El problema que enfrentamos en este modelo es determinar si la componente de varianza de los efectos aleatorios del factor B es significativa, y además determinar si los efectos fijos del factor A son significativos sobre la respuesta promedio. Por tanto los parámetros de interés son los siguientes:

- σ^2 la variación entre muestras de tratamientos o celdas.
- σ_β^2 la variación del desempeño del factor B, o variación por cambios del factor B.
- Los efectos α_i e incluso las medias $\mu_{i\bullet}$ de la respuesta en cada nivel de A. Sin embargo las inferencias sobre estos parámetros sigue siendo complicada, debido a que ciertos pares de observaciones están correlacionadas en el modelo mixto, lo cual dificulta las estimaciones de las varianzas de los estimadores de efectos y medias de nivel de A.

8.10.2. ANOVA

La variabilidad total se descompone como

$$SST = SSA + SSB + SSE,$$

y la tabla ANOVA queda como se muestra en la Tabla 8.21.

8.10.3. Pruebas de hipótesis

Significancia de los efectos fijos principales del factor A

$$H_0 : \alpha_1 = \alpha_2 = \dots, \alpha_a = 0 \quad \text{vs.} \quad H_1 : \text{al menos un } \alpha_i \neq 0. \quad (8.50)$$

Note que si H_0 fuese cierta, entonces $E[MSA] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_1 = \frac{MSA}{MSE} \stackrel{H_0}{\sim} f_{a-1, abn-a-b+1}. \quad (8.51)$$

Rechazar H_0 a un nivel de significancia γ si $F_1 > f_{\gamma, a-1, abn-a-b+1}$, o con valor P si $P(f_{a-1, a-1, abn-a-b+1} > F_1)$ es pequeño.

Significancia de la componente de varianza del factor B

$$H_0 : \sigma_\beta^2 = 0 \quad \text{vs.} \quad H_1 : \sigma_\beta^2 > 0. \quad (8.52)$$

Note que si H_0 fuese cierta, entonces $E[MSB] = \sigma^2$ que también corresponde al $E[MSE]$. Por tanto el estadístico de la prueba a usar es

$$F_2 = \frac{MSB}{MSE} \stackrel{H_0}{\sim} f_{b-1, a-1, abn-a-b+1}. \quad (8.53)$$

Rechazar H_0 a un nivel de significancia γ si $F_2 > f_{\gamma, b-1, a-1, abn-a-b+1}$, o con valor P si $P(f_{b-1, a-1, abn-a-b+1} > F_2)$ es pequeño.

8.10. Modelo factorial de dos factores de efectos mixtos sin interacción, en un DCA balanceado 285

Tabla 8.21: ANOVA en un DCA balanceado, con dos factores de efectos mixtos sin interacción

Fuente	g.l	Suma de cuadrados	Cuadrados medios	Valores esperados	F	Valor P
A	$a - 1$	SSA	$MSA = \frac{SSA}{a - 1}$	$E[MSA] = \sigma^2 + nb \frac{\sum_{i=1}^a \alpha_i^2}{a - 1}$	$\frac{MSA}{MSE}$	$P(f_{a-1, dfe} > F_0)$
B	$b - 1$	SSB	$MSB = \frac{SSB}{b - 1}$	$E[MSB] = \sigma^2 + na \sigma_\beta^2$	$\frac{MSB}{MSE}$	$P(f_{b-1, dfe} > F_0)$
Error	dfe	SSE	$MSE = \frac{SSE}{dfe}$	$E[MSE] = \sigma^2$		
Total	$abn - 1$	SST				
Fórmulas						
<p>dfe = $(abn - 1) - (a - 1) - (b - 1) = abn - a - b + 1$, los grados de libertad del error.</p> <p>$SSA = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = bn \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = an \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - abn \bar{Y}_{\bullet\bullet\bullet}^2$</p> <p>$SSE = SST - SSA - SSB = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$</p> <p>donde,</p> <p>$\bar{Y}_{ij\bullet} = \frac{1}{n} \sum_{k=1}^n Y_{ijk}$, media muestral en tratamientos $A_i B_j$</p> <p>$\bar{Y}_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n Y_{ijk} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{ij\bullet}$. Media muestral en el nivel i del factor A. Es el estimador de la media $\mu_{i\bullet}$</p> <p>$\bar{Y}_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n Y_{ijk} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{ij\bullet}$. Media muestral en el nivel j del factor B.</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}$ Media muestral de todas las abn observaciones, es el estimador de la media μ, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \bar{Y}_{ij\bullet}$ es decir el promedio de las medias muestrales de los ab tratamientos, o bien</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{a} \sum_{i=1}^a \bar{Y}_{i\bullet\bullet}$ es decir es el promedio de las medias muestrales de los a niveles del factor A, o bien,</p> <p>$\bar{Y}_{\bullet\bullet\bullet} = \frac{1}{b} \sum_{j=1}^b \bar{Y}_{\bullet j\bullet}$ es decir es el promedio de las medias muestrales de los b niveles del factor B.</p>						

8.10.4. Estimaciones

Interesa estimar σ_β^2 , α_i y $\mu_{i\bullet}$. Las estimaciones de las componentes de varianza se determinan igualando los cuadrados medios esperados a los correspondientes cuadrados medios observados y se despejan los valores desconocidos de las componentes. Las estimaciones de las tres componentes de la varianza son:

Tabla 8.22: Estimaciones de las componentes y proporciones de varianza en modelo factorial, dos factores de efectos mixtos sin interacción.

Fuente	Componente de varianza estimada	Proporción respecto a la varianza total
B	$\hat{\sigma}_\beta^2 = \frac{MSB - MSE}{na}$	$\frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}^2}$
Error	$\hat{\sigma}^2 = MSE$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}^2}$

Nota 8.13. Puede ocurrir que la componente de varianza estimada $\hat{\sigma}_\beta^2$ resulte negativa. Obviamente no tiene sentido dar como resultado una varianza negativa. Así esto debe considerarse como evidencia de que tal componente no es significativa.

En cuanto a los efectos principales y medias de nivel del factor A, en un diseño balanceado se tiene lo siguiente:

Tabla 8.23: Estimadores de efectos fijos y medias del factor A, factorial dos factores de efectos mixtos sin interacción

Parámetro	Estimador	Varianza	Error estándar	I.C. $(1 - \gamma)100\%$
α_i	$\hat{\alpha}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$	$V(\hat{\alpha}_i) = \frac{(a-1)E(MSE)}{abn}$	$S_{\hat{\alpha}_i} = \sqrt{\frac{(a-1)MSE}{abn}}$	$\hat{\alpha}_i \pm t_{\gamma/2, abn-a-b+1} S_{\hat{\alpha}_i}$
$\mu_{i\bullet}$	$\hat{\mu}_{i\bullet} = \bar{Y}_{i\bullet\bullet}$	$V(\hat{\mu}_{i\bullet}) = c_1 E(MSE) + c_2 E(MSB)$	$S_{\hat{\mu}_{i\bullet}} = \sqrt{c_1 MSE + c_2 MSB}$	$\hat{\mu}_{i\bullet} \pm t_{\gamma/2, \nu} S_{\hat{\mu}_{i\bullet}}$ aprox.
Donde: $c_1 = \frac{a-1}{abn}$, $c_2 = \frac{1}{abn}$, y $\nu \simeq \frac{\left(\frac{a-1}{abn} MSE + \frac{1}{abn} MSB\right)^2}{\frac{\left(\frac{a-1}{abn} MSE\right)^2}{abn-a-b+1} + \frac{\left(\frac{1}{abn} MSB\right)^2}{(b-1)}}$				

8.10.5. Contrastes de medias $\mu_{i\bullet}$

Para contrastes entre medias del factor A, se tiene lo siguiente: Sea $W = \sum_{i=1}^a c_i \mu_{i\bullet}$ con $\sum_{i=1}^a c_i = 0$; su estimador corresponde a $\widehat{W} = \sum_{i=1}^a c_i \bar{Y}_{i\bullet\bullet}$ el cual tiene como varianza,

$$V(\widehat{W}) = \frac{E(MSE)}{bn} \times \sum_{i=1}^a c_i^2 \quad (8.54)$$

por tanto un intervalo de confianza del $(1 - \gamma)100\%$ para W es,

$$\widehat{W} \pm t_{\gamma/2, abn-a-b+1} \times \sqrt{\frac{MSE}{bn} \sum_{i=1}^a c_i^2} \quad (8.55)$$

y el estadístico para la prueba $H_0 : W = 0$ vs. $H_1 : W \neq 0$ es,

$$T_0 = \frac{\widehat{W}}{\sqrt{\frac{MSE}{bn} \sum_{i=1}^a c_i^2}} \stackrel{H_0}{\sim} t_{abn-a-b+1} \quad (8.56)$$

Para las comparaciones múltiples de Tukey de las medias de A, los I.C simultáneos del $(1 - \gamma)100\%$ están dados por:

$$(\mu_{i\bullet} - \mu_{j\bullet}) \in (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet}) \pm q(\gamma, a, abn - a - b + 1) \times \sqrt{\frac{MSE}{bn}} \quad (8.57)$$

8.10.6. Validación de supuestos

Al igual que en los casos anteriores, para los supuestos sobre los errores reemplazamos temporalmente todos los efectos aleatorios por efectos fijos, calculamos los residuales y los examinamos como en el factorial de efectos fijos. Para los efectos aleatorios se continúa omitiendo los chequeos referentes a sus supuestos.

8.11. Ejemplo dos factores de efectos mixtos en un DCA

El objetivo de un experimento agrícola fue determinar si existían diferencias apreciables en la cantidad de trigo cosechado, entre cuatro variedades de trigo y por tipos de fertilizantes.

Fertilizante	Variedad de trigo			
	A	B	C	D
1	35	45	24	55
	26	39	23	48
	38	39	36	39
	20	43	29	49
2	55	64	58	68
	44	57	74	61
	68	62	49	60
	64	61	69	75
3	97	93	89	82
	89	91	98	78
	92	82	85	89
	99	98	87	92

Para el experimento se encontró un área muy grande de siembra, en la que las condiciones del suelo eran prácticamente homogéneas. Tres clases de fertilizante se seleccionaron aleatoriamente de entre un número relativamente grande de fertilizantes, pero el interés no se extendió más allá de las cuatro variedades de trigo seleccionadas para el experimento. El área disponible fue dividida en 12 zonas de igual tamaño para las 12 combinaciones de variedad de trigo y tipo de fertilizante. Para medir el error experimental, cada zona se dividió a su vez en cuatro y cada una de éstas recibió el mismo tratamiento. En el momento de la cosecha se observaron los datos que aparecen en la tabla adjunta. Con base en esta información responda a las siguientes preguntas:

- a) Escriba el modelo apropiado para este problema (mixto no restringido).
- b) Haga un análisis descriptivo del conjunto de datos.
- c) Calcule el ANOVA y determine si la componente de varianza de la interacción es significativa. Use un nivel de significancia de 0.05.
- d) Discuta la significancia de los factores fertilizante y variedad de trigo.
- e) Construya la tabla de medias por niveles del factor variedad de trigo, compare las medias según este factor. Interprete los resultados.
- f) Estime e interprete las componentes de varianza.
- g) Valide supuestos sobre los errores del modelo (normalidad y varianza constante) usando residuos internamente estudentizados.

8.11.1. Modelo para el experimento

Considerando el modelo mixto no restringido tome como factor A: variedad de trigo (pues es el factor de efectos fijos) y B: tipo de fertilizante (pues es el factor de efectos aleatorios), entonces con $a = 4$, $b = 3$ y $n = 4$, el modelo corresponde a:

$$Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}$$

donde:

Y_{ijk} es la cantidad cosechada de trigo en la k -ésima parcela asignada a la variedad i y fertilizante j .

μ es el promedio global de la cantidad cosechada de trigo.

α_i es el efecto fijo principal de la variedad de trigo i (1=A, 2=B, 3=C, y 4=D) sobre el promedio global de la cantidad cosechada de trigo.

B_j es el efecto aleatorio del tipo de fertilizante j sobre el promedio global de la cantidad cosechada de trigo, $B_j \stackrel{iid}{\sim} N(0, \sigma_\beta^2)$.

$(\alpha B)_{ij}$ es el efecto aleatorio de la interacción de la variedad de trigo i con el tipo de fertilizante j sobre el promedio global de la cantidad cosechada de trigo, $(\alpha B)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$.

ε_{ijk} es el error aleatorio en la k -ésima réplica del tratamiento $A_i B_j$, $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$.

Además se supone que $(\alpha B)_{ij}$ y ε_{ijk} son mutuamente independientes.

Restricción para estimación: $\sum_i^4 \alpha_i = 0$.

8.11.2. Análisis descriptivo

En la Figura 8.9 aparecen en la parte inferior las dos gráficas posibles de perfiles de medias, para diagnosticar si existe interacción. En ambas gráficas los perfiles no son paralelos, lo cual indica la posibilidad de una componente de varianza significativa debida a la interacción, es decir $\sigma_{\alpha\beta}^2$. Las diferencias de las medias de la cantidad cosechada por variedad de trigo no ocurren de manera congruente con la variación de los tipos de fertilizantes usados. Por otra parte en los boxplots según factor fertilizante es claro que existe variación de las medias y por tanto tal suceso es indicativo de que la variación de fertilizante puede estar contribuyendo significativamente en la varianza total de la cantidad de trigo cosechada, es decir, σ_β^2 puede ser estimado, pues no ha sido enmascarado por la interacción. En cambio en los boxplots según factor variedad, aparentemente el promedio de la cantidad cosechada de trigo es prácticamente el mismo para todas las 4 variedades, sin embargo, si la interacción es significativa, no se puede concluir tal cosa, y lo que se observa en los boxplots según factor variedad es un enmascaramiento de los efectos principales de la variedad de trigo por

parte de la interacción observada. Sobre la varianza, a nivel del factor fertilizante, se observa aproximadamente la misma varianza, pero a nivel de las variedades de trigo no hay igualdad de la varianza de las observaciones; sin embargo, para validar el supuesto de varianza constante sobre los errores es mejor diagnosticar con los gráficos de residuales.

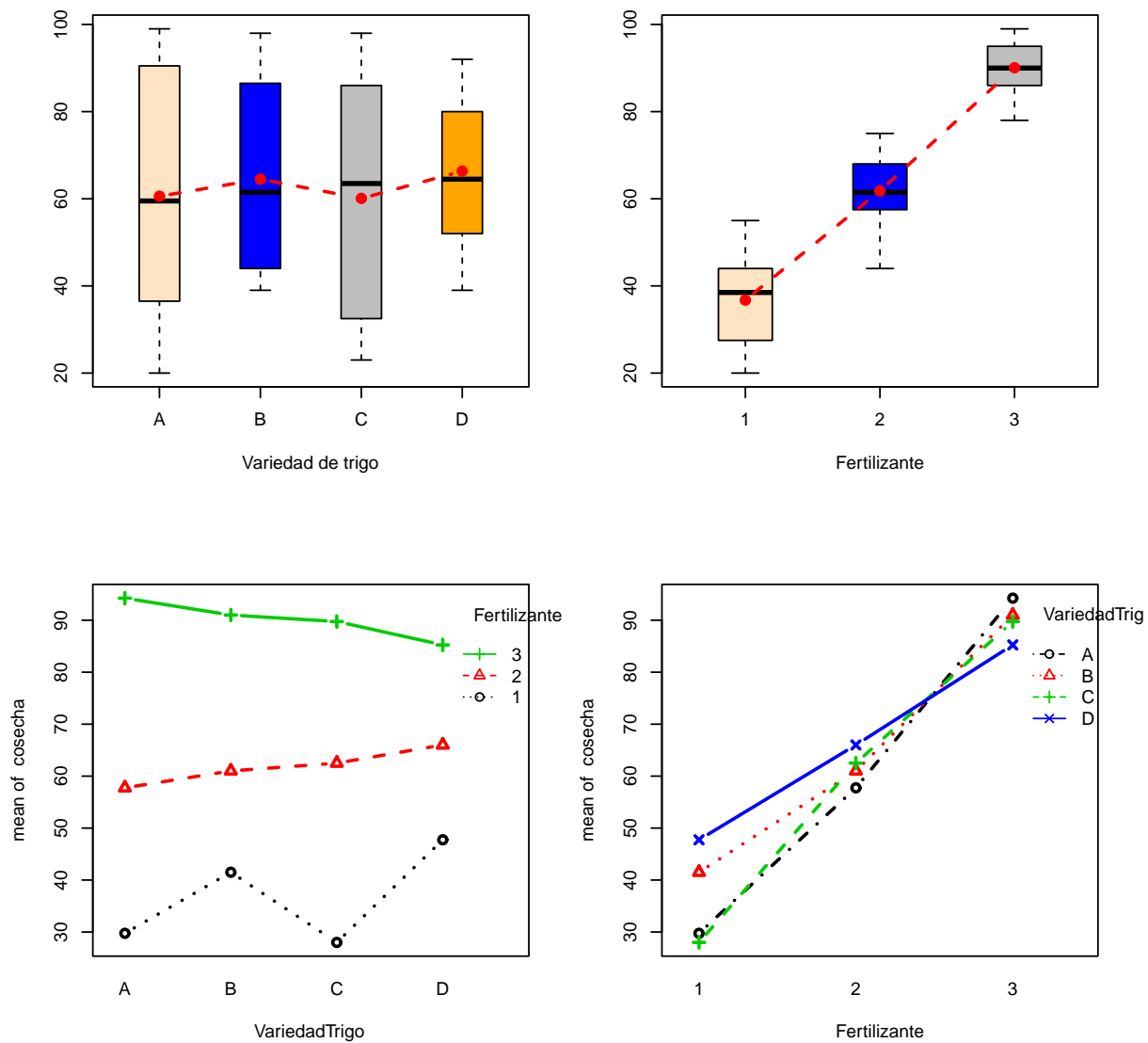


Figura 8.9: Boxplots según factores y gráficos de interacción, en datos del Ejemplo 8.11

8.11.3. ANOVA y estimaciones

La siguiente Tabla ANOVA se construyó en R usando la función usuario denominada como `Anovafactorialmixtoconinteraccion()`, la cual permite obtener los resultados tanto en el mixto restringido como no restringido, a partir de la ANOVA del modelo factorial de efectos fijos con interacción (vea Programa R). Si bien el ejercicio pide sólo el caso no restringido, aquí se exhiben los dos casos con propósito de ilustrar el manejo de la función.

Salida R 8.11.1. Resultados para modelo factorial de dos factores de efectos mixtos, modelo no restringido, con función usuario `Anovafactorialmixtoconinteraccion()`.

```
> modeloaux=aov(cosecha~VariedadTrigo*Fertilizante) #Modelo de efectos fijos con interacción
> Anovafactorialmixtoconinteraccion(modelo=modeloaux) #ANOVA mixto no restringido
The Unrestricted Mixed Model Analysis of Variance
Dependent Variable: cosecha
```

	Source	DF	Sum.Sq	Mean.Sq	F.Value	P.value
1	VariedadTrigo	3	331.750	110.5833	0.63	0.6219
2	Fertilizante	2	22764.875	11382.4375	64.91	9e-05
3	VariedadTrigo:Fertilizante	6	1052.125	175.3542		

```
Source DF Sum.Sq Mean.Sq F.Value P.value
1 VariedadTrigo:Fertilizante 6 1052.125 175.35417 3.55 0.00729
2 Error 36 1776.500 49.34722
```

Salida R 8.11.2. Resultados para modelo factorial de dos factores de efectos mixtos, modelo restringido, con función usuario `Anovafactorialmixtoconinteraccion()`.

```
> modeloaux=aov(cosecha~VariedadTrigo*Fertilizante) #Modelo de efectos fijos con interacción
> Anovafactorialmixtoconinteraccion(modelo=modeloaux,norestringido="FALSE") #ANOVA mixto restringido
The Restricted Mixed Model Analysis of Variance
Dependent Variable: cosecha
```

	Source	DF	Sum.Sq	Mean.Sq	F.Value	P.value
1	VariedadTrigo	3	331.750	110.5833	0.63	0.6219
2	VariedadTrigo:Fertilizante	6	1052.125	175.3542		

```
Source DF Sum.Sq Mean.Sq F.Value P.value
1 Fertilizante 2 22764.875 11382.43750 230.66 0
2 VariedadTrigo:Fertilizante 6 1052.125 175.35417 3.55 0.00729
3 Error 36 1776.500 49.34722
```

ADVERTENCIA: el modelo ajustado como de efectos fijos, que en la programación R suministrada es denominado `modeloaux`, debe invocar como primer factor aquel que realmente sea de efectos fijos, de lo contrario los resultados de la función usuario `Anovafactorialmixtoconinteraccion()` no serán correctos.

Con base en la Salida R 8.11.1, es decir, con ANOVA del mixto no restringido, procedemos a los tests de hipótesis siguientes.

Significancia de la componente de varianza debida a la interacción

$$H_0 : \sigma_{\alpha\beta}^2 = 0 \text{ vs. } H_1 : \sigma_{\alpha\beta}^2 > 0$$

El estadístico de prueba es $F_3 = MS(AB) / MSE = 3.55$, $F_3 \stackrel{H_0}{\sim} f_{6,36}$, y $VP = P(f_{6,36} > 3.55) = 0.0072$. Se concluye con una significancia de 0.05 que sí hay componente de varianza significativa aportada por la interacción entre los factores VARIEDAD DE TRIGO y TIPO DE FERTILIZANTE, por lo tanto las diferencias entre las medias del factor VARIEDAD DE TRIGO variarán con las variaciones del factor TIPO DE FERTILIZANTE. Como consecuencia de esto es necesario interpretar con precaución los resultados de los tests sobre la significancia de los efectos principales del factor VARIEDAD DE TRIGO y sobre la significancia de la componente de varianza debida a la variación del TIPO DE FERTILIZANTE.

Significancia de los efectos fijos del factor VARIEDAD DE TRIGO

Se prueba que

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (efectos principales del factor variedad trigo no son significativos)

$H_1 : \text{algún } \alpha_i \neq 0, i = 1, 2, 3, 4$ (algunos efectos principales del factor son significativos)

o bien, en términos de las medias de este factor:

$$H_0 : \mu_{1\bullet} = \mu_{2\bullet} = \mu_{3\bullet} = \mu_{4\bullet} \left(\begin{array}{l} \text{Las medias de cantidad cosechada según el factor variedad trigo} \\ \text{no difieren significativamente} \end{array} \right)$$

$$H_1 : \mu_{i\bullet} \neq \mu_{j\bullet} \left(\begin{array}{l} \text{Al menos dos de las medias de las cantidades cosechadas según el factor} \\ \text{variedad trigo difieren significativamente} \end{array} \right)$$

En cualquier caso el estadístico de prueba corresponde a $F_1 = MSA/MS(AB) = 0.63$, $F_1 \stackrel{H_0}{\sim} f_{3,6}$ y $VP = P(f_{3,6} > 0.63) = 0.6216$. Se concluye que con una significancia de 0.05, ante interacción significativa, los efectos del factor VARIEDAD DE TRIGO son enmascarados por la interacción y por lo tanto no pueden ser medidos sus efectos principales!. Esto quiere decir que no tiene sentido estimar ni efectos, ni medias para los niveles de este factor y tampoco tiene sentido construir I.C, o hacer comparaciones de Tukey o contrastes de sus medias. Como los niveles estudiados del factor TIPO DE FERTILIZANTE son una muestra aleatoria, tampoco tiene sentido comparar en cada uno de los tres fertilizantes seleccionados, las medias del factor VARIEDAD DE TRIGO, aunque teóricamente se puedan deducir los estadísticos para tales tipos de pruebas, las conclusiones a las cuales se pudiesen llegar con éstas sólo son aplicables a los tipos de fertilizantes estudiados y no a toda la población de fertilizantes disponibles en el mercado.

Significancia de la componente de varianza debida al factor TIPO DE FERTILIZANTE

$$H_0 : \sigma_\beta^2 = 0 \text{ vs. } H_1 : \sigma_\beta^2 > 0$$

El estadístico de prueba corresponde a $F_0 = MSB/MS(AB) = 64.91$, $F_0 \stackrel{H_0}{\sim} f_{2,6}$ y $VP = P(f_{2,6} > 64.91) < 0.0001$. Con estos resultados se corrobora una componente de varianza altamente significativa aportada por la variación de los tipos de fertilizantes empleados. Observe la alta significancia de este término, además que este factor posee la mayor suma de cuadrados y de cuadrados medios. Se determinará luego qué proporción de la varianza total es aportada por éste factor.

Medias del factor variedad de trigo

Ya se ha concluido del ANOVA que los efectos de este factor han sido enmascarados y que las inferencias sobre estos no son recomendables. Sin embargo vea a continuación los resultados de las estimaciones de las medias $\mu_{i\bullet}$ con sus I.C de acuerdo a las ecuaciones dadas en la Tabla 8.18, las cuales aplican tanto en el modelo restringido como en el no restringido. Para la obtención de estos resultados se construyó en

R la función usuario denominada `mediasfactorialmixtointerv02()` la cual tiene como argumento principal a `modelo`, el cual se refiere al objeto producido con la función `lmer()` en la estimación de las componentes de varianza, pero únicamente en el modelo mixto no restringido.

Salida R 8.11.3. *Medias estimadas para modelo factorial de dos factores de efectos mixtos, modelo no restringido, con función usuario `mediasfactorialmixtointerv02()`.*

```
> modelob=lmer(cosecha~1+VariedadTrigo+(1|Fertilizante)+(1|VariedadTrigo:Fertilizante))
> mediasfactorialmixtointerv02(modelo=modelob)
Least Square Means
```

	Estimate	Std. Error	df	t value	lower	upper
VariedadTrigoA	60.58333	15.75102	2.187569	3.846311	14.59058	106.5761
VariedadTrigoB	64.50000	15.75102	2.187569	4.094972	18.50724	110.4928
VariedadTrigoC	60.08333	15.75102	2.187569	3.814567	14.09058	106.0761
VariedadTrigoD	66.33333	15.75102	2.187569	4.211367	20.34058	112.3261

Nota 8.14. En el programa R también se ha creado la función usuario denominada `mediasfactorialmixtonointerv02()` con la cual se estimarían las medias del factor de efectos fijos y sus I.C, cuando el modelo mixto no considera interacción, y su principal argumento `modelo` se refiere al objeto producido con la función `lmer()` en la estimación de las componentes de varianza en el modelo mixto sin interacción. Vea en programa R cómo se usaría esta función si fuese el modelo mixto sin interacción.

Nota 8.15. Tenga en cuenta además que se necesita la librería `lmerTest` para poder hacer uso de las dos funciones de usuario para la estimación de medias.

En la salida R 8.11.3 aparecen las medias estimadas $\bar{Y}_{i..}$ en la columna Estimate, sus errores estándar en *Standard Error* los cuales se calculan como $S_{\bar{Y}_{i..}} = \sqrt{c_1 MS(AB) + c_2 MSB} = 15.7510$ donde $c_1 = (a - 1) / abn$, $c_2 = 1 / abn$; los límites con base en la distribución t - Student de los respectivos intervalos de confianza $\mu_{i.} \in \bar{Y}_{i..} \pm t_{0.025, \nu} \times S_{\bar{Y}_{i..}}$, aparecen en las columnas lower y upper. Los grados de libertad DF de la variable t -student, se obtienen mediante la aproximación Satterthwaite, según la fórmula:

$$\nu \simeq \frac{[c_1 MS(AB) + c_2 MSB]^2}{\frac{[c_1 MS(AB)]^2}{(a-1)(b-1)} + \frac{[c_2 MSB]^2}{(b-1)}} \simeq 2.2$$

Adicionalmente, se presentan los valores en t Value que corresponden a los valores observados de los estadísticos $T_0 = \bar{Y}_{i\bullet\bullet} / S_{\bar{Y}_{i\bullet\bullet}} \stackrel{approx}{\sim} t_\nu$ para las pruebas $H_0 : \mu_{i\bullet} = 0$ vs. $H_1 : \mu_{i\bullet} \neq 0$, lo cual no es de nuestro interés. Lo que interesa analizar son las medias estimadas y sus I.C. Pero ante el problema de enmascaramiento cualquier inferencia generalizada sobre las medias de la cantidad cosechada según variedades de trigo, no resultarían correctas.

Comparaciones de pares de medias del factor variedad de trigo

Los siguientes resultados también fueron obtenidos mediante una función usuario creada para obtener los I.C del 95 % para las diferencias de medias del factor fijo, mediante métodos LSD y Tukey, en el modelo mixto con interacción (aplica tanto en el caso no restringido como en el restringido). El nombre de la función es `LSDTukeymixtosinter()`, la cual maneja como argumentos a respuesta para especificar la variable respuesta, `factorfijo` para especificar el factor de efectos fijos, `nivelesA` y `nivelesB` para especificar el número de niveles de los factores fijo y aleatorio, respectivamente, y `MSAB` para especificar el valor de las sumas de cuadrados medios de la interacción en el ANOVA.

Salida R 8.11.4. *Comparaciones de pares de medias del factor de efectos fijos en modelo factorial de dos factores de efectos mixtos con interacción (estos resultados aplican tanto en el modelo restringido como en el no restringido), con función usuario `LSDTukeymixtosinter()`.*

```
> LSDTukeymixtosinter(respuesta=cosecha,factorfijo=VariedadTrigo,nivelesA=4,nivelesB=3,MSAB=175.3542)
LSD Test for fixed factor in mixed model with interaction

      test p.adjusted      name.t ntr alpha
Fisher-LSD      none factorfijo    4  0.05

Comparison between treatments means

      difference pvalue signif.      LCL      UCL
A - B   -3.916667 0.4960      -17.144868  9.311534
A - C    0.500000 0.9293      -12.728201 13.728201
A - D   -5.750000 0.3284      -18.978201  7.478201
B - C    4.416667 0.4452       -8.811534 17.644868
```

B - D	-1.833333	0.7461	-15.061534	11.394868
C - D	-6.250000	0.2916	-19.478201	6.978201

HSD Test for fixed factor in mixed model with interaction

test name.t ntr StudentizedRange alpha

Tukey factorfijo 4 4.895599 0.05

Comparison between treatments means

	difference	pvalue	signif.	LCL	UCL
A - B	-3.916667	0.8840		-22.63096	14.79762
A - C	0.500000	0.9997		-18.21429	19.21429
A - D	-5.750000	0.7220		-24.46429	12.96429
B - C	4.416667	0.8447		-14.29762	23.13096
B - D	-1.833333	0.9853		-20.54762	16.88096
C - D	-6.250000	0.6724		-24.96429	12.46429

Nota 8.16. En la programación R también se ha creado la función usuario denominada `LSDTukeymixtosnointer()`, para las comparaciones de pares de medias en el modelo mixto sin interacción. Vea en programa R cómo se usaría esta función.

Nota 8.17. Ambas funciones de usuario para las comparaciones de medias dependen de la librería `agricolae`.

En la Salida R 8.11.4 se tienen las estimaciones de las diferencias de medias (sin repetir casos) de los niveles del factor VARIEDAD DE TRIGO; en la columna Difference aparecen los valores correspondientes a $\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet}$; los errores estándar de estas diferencias, se calculan como $S_{(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet})} = \sqrt{2MS(AB)/bn}$; en la columna pvalue sig. se dan los valores P para la prueba de la igualdad de medias en cada comparación: $H_0 : \mu_{i\bullet} = \mu_{j\bullet}$ vs. $H_1 : \mu_{i\bullet} \neq \mu_{j\bullet}$, de acuerdo a la distribución $t_{(a-1)(b-1)}$ en el método LSD y de acuerdo a la distribución de los rangos estudentizados $q(a, (a-1)(b-1))$ en el método de Tukey. En el método LSD los intervalos de confianza del 95 % son calculados como

$$(\mu_{i\bullet} - \mu_{j\bullet}) \in (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet}) \pm t_{0.025, (a-1)(b-1)} \times S_{(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet})}$$

mientras que en Tukey corresponden a

$$(\mu_{i\bullet} - \mu_{j\bullet}) \in (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{j\bullet\bullet}) \pm q_{0.05}(a, (a-1)(b-1)) \times \sqrt{MS(AB)/bn}$$

Como puede observarse, todos los valores P son grandes y el cero está contenido en cada I.C de LSD y de Tukey, con lo cual se concluiría erróneamente que todas las variedades producen en promedio la misma cantidad de cosecha, sin embargo dado que hay interacción significativa, estos resultados sólo nos dicen que hay enmascaramiento de los efectos de las variedades de trigo y por tanto tales efectos no pueden medirse. Así que de nada sirve la interpretación de las diferencias e I.C reportados! como ya se había advertido desde el test ANOVA.

Estimaciones de las componentes de varianza

A continuación se dan las estimaciones correspondientes,

Salida R 8.11.5. *Estimación de las componentes de varianza con función `lmer()`. Esta función sólo considera el modelos mixto no restringido. De los distintos resultados exhibidos con el `summary()`, sólo nos interesa los que se muestran a continuación, en donde *Variance* es el estimador de la componente de varianza y *Std.Dev* su raíz cuadrada.*

```
> modelob=lmer(cosecha~1+VariedadTrigo+(1|Fertilizante)+(1|VariedadTrigo:Fertilizante))
> summary(modelob) #tomar sólo lo referente a Random effects:
Linear mixed model fit by REML t-tests use Satterthwaite approximations to degrees of freedom [lmerMod]
Random effects:

```

Groups	Name	Variance	Std.Dev.
VariedadTrigo:Fertilizante	(Intercept)	31.50	5.613
Fertilizante	(Intercept)	700.44	26.466
Residual		49.35	7.025

```
Number of obs: 48, groups: VariedadTrigo:Fertilizante, 12; Fertilizante, 3
```

De esta salida R se tiene que:

Como puede observarse, la componente de varianza que más pesa en este problema, es la debida a la variación de los tipos de fertilizantes, componente de varianza mucho más importante que la debida a la interacción.

Si bien en este problema no resultó posible medir los efectos de las variedades de trigo sobre el promedio de la cantidad cosechada, es claro que es crítico controlar al factor tipo de fertilizante y buscar a través de otros experimentos, el tipo de fertilizante

Componente de varianza debida a	Estimación	Porcentaje
La variación de los tipos de fertilizantes	$\hat{\sigma}_{\beta}^2 = \frac{MSB - MS(AB)}{na} = \frac{11382.4375 - 175.354167}{16} = 700.44$	$\frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 89.65 \%$
La variación en la interacción	$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS(AB) - MSE}{n} = \frac{175.35417 - 49.34722}{4} = 31.5017$	$\frac{\hat{\sigma}_{\alpha\beta}^2}{\hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 4.03 \%$
La variación aleatoria	$\hat{\sigma}^2 = MSE = 49.3472$	$\frac{\hat{\sigma}^2}{\hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2} \times 100 \% = 6.32 \%$
Varianza total	$\hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2 = 781.2889$	100 %

que debe aplicarse (no utilizar cualquier tipo ni dejar que este factor esté variando, dado que los distintos tipos no proporcionan los mismos resultados), junto con el tipo de variedad más apropiado. Es decir, se tendrá que recurrir a la experimentación secuencial buscando la mejor combinación de estos dos factores para maximizar la producción.

8.11.4. Validación de supuestos sobre los errores

Para esto recuerde que se usan los residuos internamente estudentizados del modelo factorial de efectos fijos, con interacción.

En la Figura 8.10, puede observarse que el valor del estadístico del test Shapiro Wilk $W=0.97944$ con un valor P de 0.5554 nos lleva a concluir que el supuesto de normalidad para los errores del modelo, es un supuesto razonable. Aunque en el gráfico de probabilidad normal se observa alguna desviación en la cola superior, no es una desviación grave como para rechazar.

En todas las gráficas se observan tres residuales que pueden corresponder a valores extremos. Sin considerar a tales, puede decirse que aproximadamente se cumple el supuesto de varianza constante para los errores del modelo mixto ajustado.

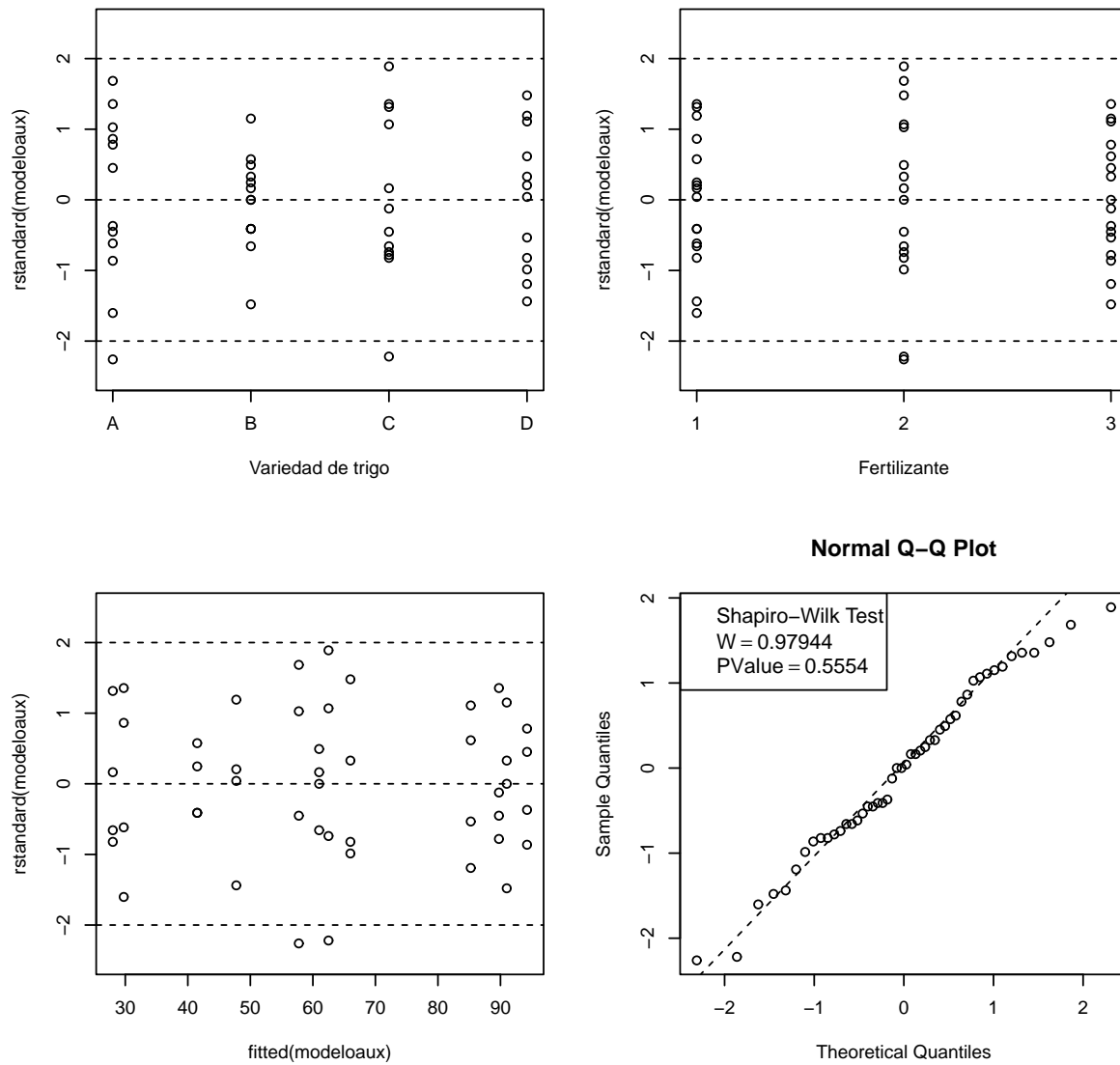


Figura 8.10: Gráficos de residuales internamente estudentizados para validar supuestos en modelo factorial mixto, en datos del Ejemplo 8.11

A continuación vea el código R usado.

Código R 8.11.1. Programa R usado en Ejemplo 8.11.

```

library(daewr)
library(lme4)
library(MASS)
library(lmerTest)

datos=data.frame(VariedadTrigo=factor(rep(c("A", "B", "C", "D"), times=12)),
                  Fertilizante=factor(rep(1:3, each=16)), cosecha=scan())

35 45 24 55
26 39 23 48
38 39 36 39
20 43 29 49
55 64 58 68
44 57 74 61
68 62 49 60
64 61 69 75
97 93 89 82
89 91 98 78
92 82 85 89
99 98 87 92

datos
attach(datos)
#GRÁFICOS PARA ANALIZAR LOS DATOS EXPERIMENTALES
mediasvariedad=sapply(split(cosecha, VariedadTrigo), mean)
mediasfertilizante=sapply(split(cosecha, Fertilizante), mean)
layout(rbind(c(1,1,2,2), c(3,3,4,4)))
boxplot(cosecha~VariedadTrigo, xlab="Variedad de trigo", boxwex=0.4, col=c("bisque", "blue", "gray", "orange"))
lines(1:4, mediasvariedad, type="b", pch=19, col=2, lty=2, lwd=2)
boxplot(cosecha~Fertilizante, xlab="Fertilizante", boxwex=0.4, col=c("bisque", "blue", "gray"))
lines(1:3, mediasfertilizante, type="b", pch=19, col=2, lty=2, lwd=2)
interaction.plot(VariedadTrigo, Fertilizante, cosecha, type="b", lwd=2, col=1:3, pch=1:3)
interaction.plot(Fertilizante, VariedadTrigo, cosecha, type="b", lwd=2, col=1:4, pch=1:4)

#CREANDO FUNCIÓN DE USUARIO PARA ANOVA CORRECTA EN MODELOS MIXTOS
Anovafactorialmixtoconinteraccion=function(modelo, norestringido="TRUE") {
  dfA=anova(modelo)[1][1,]
  dfB=anova(modelo)[1][2,]
  dfI=anova(modelo)[1][3,]
  dfE=anova(modelo)[1][4,]

```



```

SSA=anova(modelo)[2][1,]
SSB=anova(modelo)[2][2,]
SSAB=anova(modelo)[2][3,]
SSE=anova(modelo)[2][4,]
MSA=anova(modelo)[3][1,]
MSB=anova(modelo)[3][2,]
MSAB=anova(modelo)[3][3,]
MSE=anova(modelo)[3][4,]
if(norestringido=="TRUE"){
F1=round(MSA/MSAB,digits=2)
F2=round(MSB/MSAB,digits=2)
F3=round(MSAB/MSE,digits=2)
vp1=round(pf(F1,df1=dfA,df2=dfi,lower.tail=F),digits=5)
vp2=round(pf(F2,df1=dfB,df2=dfi,lower.tail=F),digits=5)
vp3=round(pf(F3,df1=dfi,df2=dfe,lower.tail=F),digits=5)
nombres=attr(attr(modelo$model,"terms"),"term.labels")
anova1=data.frame(Source=nombres,DF=c(dfA,dfB,dfi),"Sum\ Sq"=c(SSA,SSB,SSAB),"Mean Sq"=c(MSA,MSB,MSAB),
                  "F\ Value"=c(F1,F2," "),"P\ value"=c(vp1,vp2," "))
anova2=data.frame(Source=c(nombres[3],"Error"),DF=c(dfi,dfe),"Sum\ Sq"=c(SSAB,SSE),"Mean Sq"=c(MSAB,MSE),
                  "F\ Value"=c(F3," "),"P\ value"=c(vp3," "))
cat("The Unrestricted Mixed Model Analysis of Variance","\n")
cat("Dependent Variable:",names(modelo$model)[1],"\n")
print(anova1)
cat("\n")
print(anova2)
}
if(norestringido=="FALSE"){
F1=round(MSA/MSAB,digits=2)
F2=round(MSB/MSE,digits=2)
F3=round(MSAB/MSE,digits=2)
vp1=round(pf(F1,df1=dfA,df2=dfi,lower.tail=F),digits=5)
vp2=round(pf(F2,df1=dfB,df2=dfe,lower.tail=F),digits=5)
vp3=round(pf(F3,df1=dfi,df2=dfe,lower.tail=F),digits=5)
nombres=attr(attr(modelo$model,"terms"),"term.labels")
anova1=data.frame(Source=nombres[c(1,3)],DF=c(dfA,dfi),"Sum\ Sq"=c(SSA,SSAB),"Mean Sq"=c(MSA,MSAB),
                  "F\ Value"=c(F1," "),"P\ value"=c(vp1," "))
anova2=data.frame(Source=c(nombres[2:3],"Error"),DF=c(dfB,dfi,dfe),"Sum\ Sq"=c(SSB,SSAB,SSE),
                  "Mean Sq"=c(MSB,MSAB,MSE),"F\ Value"=c(F2,F3," "),"P\ value"=c(vp2,vp3," "))
cat("The Restricted Mixed Model Analysis of Variance","\n")
cat("Dependent Variable:",names(modelo$model)[1],"\n")
print(anova1)

```

```

cat("\n")
print(anova2)
}
}

#correr modelo como de efectos fijos pero colocar de primero el factor que realmente es fijo
modeloaux=aov(cosecha~VariedadTrigo*Fertilizante)
Anovafactorialmixtoconinteraccion(modelo=modeloaux) #ANOVA mixto no restringido
Anovafactorialmixtoconinteraccion(modelo=modeloaux,norestringido="FALSE") #ANOVA mixto restringido

#Validación de supuestos usando residuos estandarizados del modelo factorial
#como si fuese de efectos fijos
shapiro.test(rstandard(modeloaux))
#OBTENIENDO GRÁFICOS DE RESIDUOS INTERNAMENTE ESTUDENTIZADOS,
layout(rbind(c(1,1,2,2),c(3,3,4,4)))
stripchart(rstandard(modeloaux)~VariedadTrigo,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,
           xlab="Variedad de trigo")
abline(h=c(-2,0,2),lty=2)
stripchart(rstandard(modeloaux)~Fertilizante,vertical=TRUE,ylim=c(-2.5,2.5),pch=1,cex=1,
           xlab="Fertilizante")
abline(h=c(-2,0,2),lty=2)
plot(fitted(modeloaux),rstandard(modeloaux),ylim=c(-2.5,2.5));abline(h=c(-2,0,2),lty=2)
qqnorm(rstandard(modeloaux))
qqline(rstandard(modeloaux),lty=2)
legend("topleft",legend=c("Shapiro-Wilk Test",
                           expression(W==0.97944),expression(PValue== 0.5554)),cex=1.1)

#Estimación de efectos y componentes de varianza en modelo mixto no restringido
modelob=lmer(cosecha~1+VariedadTrigo+(1|Fertilizante)+(1|VariedadTrigo:Fertilizante))
summary(modelob) #tomar sólo lo referente a Random effects:

#Función usuario para construcción de los I.C para medias del factor de efectos fijos,
#ejecutable en R >=3.5.1
#En modelo mixto con interacción. El argumento modelo hace referencia al modelo ajustado
#con la función lmer
mediasfactorialmixtointerv02=function(modelo,alpha=0.05){
  medias=lsmeansLT(modelo)
  df=round(medias$df)
  tcrit=qt(alpha,df=df,lower.tail=F)
  lower=medias$Estimate-tcrit*medias$"Std. Error"
  upper=medias$Estimate+tcrit*medias$"Std. Error"

```

```
mediasf=data.frame(medias$Estimate,medias$"Std. Error",medias$df,medias$"t value",lower,upper,
                  row.names=dimnames(medias)[[1]])
names(mediasf)=dimnames(medias)[[2]][1:6]
cat("Least Square Means","\n")
mediasf
}
```

```
#En la función, el argumento modelo debe ser el objeto obtenido con lmer en modelo mixto con interacción
mediasfactorialmixtointerv02(modelo=modelob)
```

```
#Función usuario para construcción de los I.C para medias del factor de efectos fijos
#ejecutable en R >=3.5.1
#En modelo mixto sin interacción. El argumento modelo hace referencia al modelo ajustado con
#la función lmer
mediasfactorialmixtonointerv02=function(modelo,alpha=0.05){
  medias=lsmeansLT(modelo)
  df=round(medias$df)
  tcrit=qt(alpha,df=df,lower.tail=F)
  lower=medias$Estimate-tcrit*medias$"Std. Error"
  upper=medias$Estimate+tcrit*medias$"Std. Error"
  mediasf=data.frame(medias$Estimate,medias$"Std. Error",medias$df,medias$"t value",lower,upper,
                    row.names=dimnames(medias)[[1]])
  names(mediasf)=dimnames(medias)[[2]][1:6]
  cat("Least Square Means","\n")
  mediasf
}
```

```
#Lo siguiente no lo ejecute, sólo es para ilustrar cómo hacer uso de la
#función usuario mediasfactorialmixtonointerv02()
#En la función, el argumento modelo debe ser el objeto ajustado con lmer
modeloc=lmer(cosecha~1+VariedadTrigo+(1|Fertilizante)) #modelo mixto sin interacción
summary(modeloc)
mediasfactorialmixtonointerv02(modelo=modeloc) #Medias y sus I.C modelo mixto sin interacción
```

```
#Ejecute lo siguiente:
#Creando función para comparaciones por LSD y Tukey en modelo mixto con interacción
#Se asume que el factor A es el de efectos fijos
LSDTukeymixtosinter=function(respuesta,factorfijo,nivelesA,nivelesB,MSAB){
  library(agricolae)
  DError=(nivelesA-1)*(nivelesB-1)
  lsd=LSD.test(respuesta,factorfijo,DError=DError,MSerror=MSAB,console=F,group=F)
  hsd=HSD.test(respuesta,factorfijo,DError=DError,MSerror=MSAB,console=F,group=F)
```


Anexo A

Complementos

A.1. Funciones estimables de los parámetros

De acuerdo a [2], una función de los parámetros de cualquier modelo se dice estimable si y sólo sí, puede escribirse como el valor esperado de una combinación lineal de las variables respuesta. Sólo las funciones estimables de los parámetros tienen estimadores lineales insesgados. Es importante que el análisis de un experimento incluya sólo funciones estimables. Para el modelo ANOVA de un solo factor de efectos fijos dado en (3.2), cada función estimable es de la forma:

$$E \left[\sum_{i=1}^a \sum_{j=1}^{n_i} a_{ij} Y_{ij} \right] = \sum_{i=1}^a b_i (\mu + \alpha_i) \quad (\text{A.1})$$

con $b_i = \sum_{j=1}^{n_i} a_{ij}$ y los a_{ij} números reales.

No hay valores de b_i que den $\mu, \alpha_1, \alpha_2, \dots, \alpha_a$ separadamente como valores esperados, por tanto, estos parámetros no son individualmente estimables, sino cada $\mu + \alpha_i$.

A.2. Test de Bartlett de homogeneidad de varianzas

Suponga que se tienen a poblaciones o tratamientos independientes, cada uno con distribución normal $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, a$, con varianzas desconocidas. Se quiere probar la hipótesis de igualdad de varianzas:

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 = \dots = \sigma_a^2 \text{ vs.} \\ H_1 : &\text{algún par } \sigma_i^2 \neq \sigma_j^2 \end{aligned} \quad (\text{A.2})$$

El estadístico de prueba está dado por $\chi_0^2 = M/C \sim \chi_{a-1}^2$ con

$$M = (N - a) \log S_p^2 - \sum_{i=1}^a (n_i - 1) \log S_i^2, \quad C = 1 + \frac{1}{3(a-1)} \left(\sum_{i=1}^a \frac{1}{n_i - 1} - \frac{1}{N-a} \right), \quad S_p^2 = MSE = \frac{\sum_{i=1}^a (n_i - 1) S_i^2}{N - a},$$

donde S_i^2 es la varianza muestral de los n_i valores observados de la respuesta con el tratamiento i . Se rechaza H_0 a un nivel de significancia γ si $\chi_0^2 > \chi_{\gamma, a-1}^2$, o bien, usando valor P, si $P(\chi_{a-1}^2 > \chi_0^2)$ es pequeño. Esta prueba es altamente sensible a no normalidad por lo cual debe comprobarse primero tal supuesto.

Ver función R `bartlett.test()` de la librería `stats` (se carga por defecto al abrir R).

```
#ENTRADA DE DATOS:
ni=c(6,9,12) #No. réplicas por nivel del factor
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"), times=ni)), mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
bartlett.test(mprod~ID, data=diseño)

Bartlett test of homogeneity of variances

data:  mprod by ID
Bartlett's K-squared = 0.12936, df = 2, p-value = 0.9374
detach(diseño)
```

A.3. Test de Levene de Homogeneidad de varianza

Esta prueba fue propuesta por Levene (1960) y es robusta al supuesto de normalidad. Para su ejecución se debe reemplazar cada valor observando Y_{ij} por $Z_{ij} = |Y_{ij} - \bar{Y}_{i\bullet}|$ o bien, $Z_{ij}^2 = (Y_{ij} - \bar{Y}_{i\bullet})^2$, y luego se ejecuta el análisis de varianza sobre estas nuevas observaciones. También existe la versión basada en $\tilde{Z}_{ij} = |Y_{ij} - \tilde{Y}_{i\bullet}|$ donde $\tilde{Y}_{i\bullet}$ es la mediana del i -ésimo nivel del factor. Se rechaza H_0 Si la prueba es significativa:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2 \text{ vs.}$$

$$H_1 : \text{algún par } \sigma_i^2 \neq \sigma_j^2.$$

Ver funciones R `leveneTest()` de la librería `car` y `levene.test()` de la librería `lawstat`.

```

#ENTRADA DE DATOS:
ni=c(6,9,12) #No. réplicas por nivel del factor
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"),times=ni)),mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
library(car)
leveneTest(mprod~ID,data=diseño,center="median") #usando la mediana para centrar
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2   0.0245 0.9758
      24

leveneTest(mprod~ID,data=diseño,center="mean") #usando la media para centrar
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  2   0.1472 0.8639
      24

leveneTest(mprod~ID,data=diseño,center="mean",trim=0.1) #usando media recortada al 10% para centrar
Levene's Test for Homogeneity of Variance (center = "mean": 0.1)
      Df F value Pr(>F)
group  2   0.1471 0.864
      24

#O bien
library(lawstat)
levene.test(diseño$mprod,diseño$ID,location="median") #usando la mediana para centrar
      modified robust Brown-Forsythe Levene-type test based on the absolute
      deviations from the median
data:  diseño$mprod
Test Statistic = 0.024491, p-value = 0.9758

levene.test(diseño$mprod,diseño$ID,location="mean")#usando la media para centrar
      classical Levene's test based on the absolute deviations from the mean
      ( none not applied because the location is not set to median )
data:  diseño$mprod
Test Statistic = 0.14719, p-value = 0.8639

levene.test(diseño$mprod,diseño$ID,location="trim.mean",trim.alpha=0.1) #usando media recortada al 10% para centrar
      modified robust Levene-type test based on the absolute deviations from
      the trimmed mean ( none not applied because the location is not set to
      median )
data:  diseño$mprod
Test Statistic = 0.14708, p-value = 0.864
detach(diseño)

```

A.4. Test de Cochran para homogeneidad de varianza

A.4.1. Test C de Cochran

Este test es usado para evaluar la homogeneidad de varianzas en experimentos de un sólo factor de efectos fijos en diseños balanceados (es decir, los tamaños de muestra n_i son iguales). Se asume que se cumple que las distribuciones en cada nivel del factor son normales y que es válido el supuesto de independencia. Además, este test prueba únicamente sobre la máxima varianza (es decir, no prueba sobre la mínima varianza de las a poblaciones). El test de hipótesis es como sigue:

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 = \dots = \sigma_a^2 \text{ vs.} \\ H_1 : \sigma_i^2 &> \sigma_{i'}^2, \forall i' \neq i, \text{ para al menos un } i; i, i' = 1, 2, \dots, a. \end{aligned} \quad (\text{A.3})$$

Observe entonces que la hipótesis alternativa es que al menos una de las varianzas es significativamente más grande que las demás. Por tanto, la prueba es unilateral de cola superior. El estadístico de prueba es como sigue,

$$C = \frac{\max_{i=1, \dots, a} S_i^2}{\sum_{i=1}^a S_i^2}, \quad (\text{A.4})$$

es decir, el estadístico de prueba C corresponde a la razón de la varianza muestral más grande respecto la suma de las varianzas muestrales en los a niveles del factor. Los valores críticos de nivel γ de significancia pueden ser calculados de la siguiente forma,

$$C_\gamma(n, a) = \left[1 + \frac{a-1}{f_{\gamma/a, n-1, (a-1)(n-1)}} \right]^{-1}, \quad (\text{A.5})$$

donde $f_{\gamma/a, n-1, (a-1)(n-1)}$ es el valor crítico de nivel γ/a de una distribución F con $n-1$ y $(a-1)(n-1)$ grados de libertad en el denominador y numerador, respectivamente. Luego, se rechaza H_0 si $C > C_\gamma(n, a)$.

Ver función R `cochran.test()` de la librería `outliers`. Esta función corre aún cuando los tamaños de muestras n_i no son iguales; para ello usa como n la media aritmética de los n_i , lo cual puede ser aplicado desde que no haya mucho desbalance, en caso contrario no se recomienda su uso.


```
#ENTRADA DE DATOS:
ni=c(6,9,12) #No. réplicas por nivel
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"), times=ni)), mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
##En este ejemplo puede no ser apropiado este test por el desbalance en las réplicas
library(outliers)
cochran.test(mprod~ID, data=diseño)

      Cochran test for outlying variance
data:  mprod ~ ID
C = 0.37841, df = 9, k = 3, p-value = 1
alternative hypothesis: Group alto has outlying variance
sample estimates:
      alto      bajo      medio
0.7520000 0.6619444 0.5733333

detach(diseño)
```

A.4.2. Test G de Cochran

Es una extensión del Test C de Cochran propuesta por [11] para el caso de un DCA desbalanceado y también tiene la posibilidad de realizar prueba sobre la mínima varianza. El estadístico de prueba es denominado estadístico G . Para el test sobre la máxima varianza formulado en (A.3), el estadístico de la prueba es,

$$G = \frac{\max_{i=1, \dots, a} (n_i - 1) S_i^2}{\sum_{i=1}^a (n_i - 1) S_i^2}, \quad (\text{A.6})$$

y el valor crítico de nivel γ es dado por,

$$G_\gamma(\nu_{\text{máx}}, \nu_{\text{pool}}, a) = \left[1 + \frac{(\nu_{\text{pool}}/\nu_{\text{máx}}) - 1}{f_{\gamma/a, \nu_{\text{máx}}, \nu_{\text{pool}} - \nu_{\text{máx}}}} \right]^{-1}, \quad (\text{A.7})$$

donde $\nu_{\text{máx}}$ son los grados de libertad de la varianza muestral más grande y $\nu_{\text{pool}} = \sum_{i=1}^a (n_i - 1)$. Se rechaza H_0 si $G > G_\gamma(\nu_{\text{máx}}, \nu_{\text{pool}}, a)$. Note que cuando $n_i = n \ \forall \ i = 1, \dots, a$, el estadístico G es igual al estadístico C dado en (A.4).

Una implementación en R de este test puede ser a través de una función usuario como la que se muestra a continuación,

```

#ENTRADA DE DATOS:
ni=c(6,9,12) #No. réplicas por nivel
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"), times=ni)), mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)
#FUNCIÓN USUARIO PARA TEST G DE COCHRAN
micochranG=function(factor, respuesta, alpha=0.05){
  s2.i=sapply(split(respuesta, factor), var)
  ni=ftable(factor)
  num.G=max((ni-1)*s2.i)
  den.G=sum((ni-1)*s2.i)
  a=length(ni)
  G=num.G/den.G
  vpool=sum((ni-1))
  vmax=(ni-1)[which.max(s2.i)]
  f=qf((alpha/a), df1=vmax, df2=(vpool-vmax), lower.tail=FALSE)
  G.crit=1/(1+(vpool/vmax-1)/f)
  cat("Test G Cochran de Homogeneidad de Varianza", "\n")
  cat("Hipótesis Alternativa: Varianza outlying es significativamente mayor", "\n")
  cat("alpha", alpha, "\n")
  cat("df max.", vmax, "\n")
  cat("df pool.", vpool, "\n")
  cat("valor crítico G", G.crit, "\n")
  cat("valor estadístico G", G, "\n", "\n")
  cat("Varianzas muestrales", "\n")
  s2.i
}

#aplicando la función a los datos del DOE
micochranG(diseño$ID, diseño$mprod)
Test G Cochran de Homogeneidad de Varianza
Hipótesis Alternativa: Varianza outlying es significativamente mayor
alpha 0.05
df max. 5
df pool. 24
valor crítico G 0.4930141
valor estadístico G 0.4105309

```

```

Varianzas muestrales
      alto      bajo      medio
0.7520000 0.6619444 0.5733333
detach(diseño)

```

A.5. Transformaciones para homogeneizar varianza

Según [2], las transformaciones orientadas a homogeneizar varianzas implican hallar una función $h(Y_{ij})$ tal que $h(Y_{ij}) = \mu^* + \alpha_i^* + \varepsilon_{ij}^*$ con $\varepsilon_{ij}^* \sim N(0, \sigma^2)$ y mutuamente independientes. Una transformación apropiada puede hallarse por lo general cuando hay una clara relación entre la varianza del error $\text{Var}[\varepsilon_{ij}] = \sigma_i^2$, y la respuesta media $E[Y_{ij}] = \mu + \alpha_i$. Las relaciones son a menudo de la forma $\sigma_i^2 = k(\mu + \alpha_i)^q$, donde k y q son constantes. La función h es elegida entre las siguientes:

$$h(Y_{ij}) = \begin{cases} Y_{ij}^{1-q/2} & \text{si } q \neq 2 \\ \log(Y_{ij}) & \text{si } q = 2 \text{ y todas las } Y_{ij} \text{ son no nulas} \\ \log(Y_{ij} + 1) & \text{si } q = 2 \text{ y algunas } Y_{ij} \text{ son nulas} \end{cases} \quad (\text{A.8})$$

Por lo general q es desconocido pero la siguiente es una aproximación razonable para hallarla. Si se usan los estimadores muestrales en la expresión $\sigma_i^2 = k(\mu + \alpha_i)^q$ y tomando logaritmo natural a ambos lados de la igualdad, se obtiene: $\log(S_i^2) = \log(k) + q \log(\bar{Y}_{i\bullet})$. Por tanto, la pendiente de la línea de regresión obtenida de $\log(S_i^2)$ vs. $\log(\bar{Y}_{i\bullet})$ da una estimación para q . Pero a veces el valor de q es sugerido por consideraciones teóricas, por ejemplo cuando los datos son Poisson, en cuyo caso el valor de q es 1 (la varianza y la media son iguales); luego, la transformación a usar es $h(Y_{ij}) = Y_{ij}^{1/2}$. Con datos de una distribución binomial(m, p), se tiene que la varianza es $(1-p)$ veces la media, luego una transformación que puede estabilizar varianza es $h(Y_{ij}) = \sin^{-1} \sqrt{Y_{ij}/m}$.

Tenga en cuenta que cuando se halla una transformación que homogenice las varianzas, es necesario también chequear los otros supuestos del modelo para los datos transformados. Si no hay problemas, el ANOVA opera sobre los datos transformados.

Según [6], el método empírico basado en regresión para estimar la potencia en transformaciones de la forma $h(Y_{ij}) = y_{ij}^p$, son eficaces sólo en el DCA pues la deter-

minación de p en diseños de experimentos más complejos con bloques requieren de un enfoque más riguroso.

La transformación original de Box-Cox es de la forma $h(Y_{ij}) = (Y_{ij}^p - 1)/p$ para $p \neq 1$ y $h(Y_{ij}) = \log Y_{ij}$ para $p = 0$. En este caso la estimación de p se puede calcular maximizando,

$$L(p) = -\frac{1}{2} \log [MSE(p)], \quad (\text{A.9})$$

donde $MSE(p)$ es el MSE del ANOVA sobre los datos con la transformación $h(Y_{ij}) = (Y_{ij}^p - 1)/p$ si $p \neq 1$ o $h(Y_{ij}) = \log Y_{ij}$ si $p = 0$. Sin embargo, según [6], es común probar con algunas de las transformaciones estándar que se listan a continuación:

Tabla A.1: Transformaciones de potencia Y^p usuales. Fuente [6].

p	Y^p	Nombre	Observaciones
2	Y^2	Cuadrada	La más usada
1	Y	Datos originales	No hay transformación
1/2	\sqrt{Y}	Raíz cuadrada	Distribución de Poisson
0	$\log(Y)$	Logaritmo natural	Distribución Lognormal.
-1/2	$1/\sqrt{Y}$	Raíz cuadrada recíproca	El signo menos preserva el orden de las observaciones.
-1	$1/Y$	Recíproca	Reexpresa el tiempo como una tasa.

Ver función R `boxCox()` de la librería `car`, para transformaciones de Box-Cox de la forma y^λ y la función `transform()` de la librería `base` (o sea, la librería que por defecto se carga al abrir R) para crear `data.frames`'s con los datos transformados.

```
#ENTRADA DE DATOS:
ni=c(6,9,12) #No. Réplicas por nivel del factor
diseño=data.frame(ID=factor(rep(c("alto", "bajo", "medio"), times=ni)), mprod=scan())
8.5 9.7 10.1 7.8 9.6 9.5
7.6 8.2 6.8 5.8 6.9 6.6 6.3 7.7 6.0
6.7 8.1 9.4 8.6 7.8 7.7 8.9 7.9 8.3 8.7 7.1 8.4

attach(diseño)

#TRANSFORMACIONES DE BOX-COX
#Ojo en este conjunto de datos no se justifica la transformación pues realmente
#el log de verosimilitud con lambda=1 y el lambda óptimo son prácticamente
```

```
#iguales y lambda=1 está dentro del I.C del 95% para los lambdas
library(car)
bcx=boxCox(mprod~ID,data=diseño,family="bcPower",param="lambda")#grafica log-verosimilitud vs. lambda
lambda=bcx$x[which.max(bcx$y)];lambda #extrayendo lambda óptimo
nuevo.diseño=transform(diseño,mprod=mprod^lambda) #Conjunto de datos transformados
detach(diseño)
```

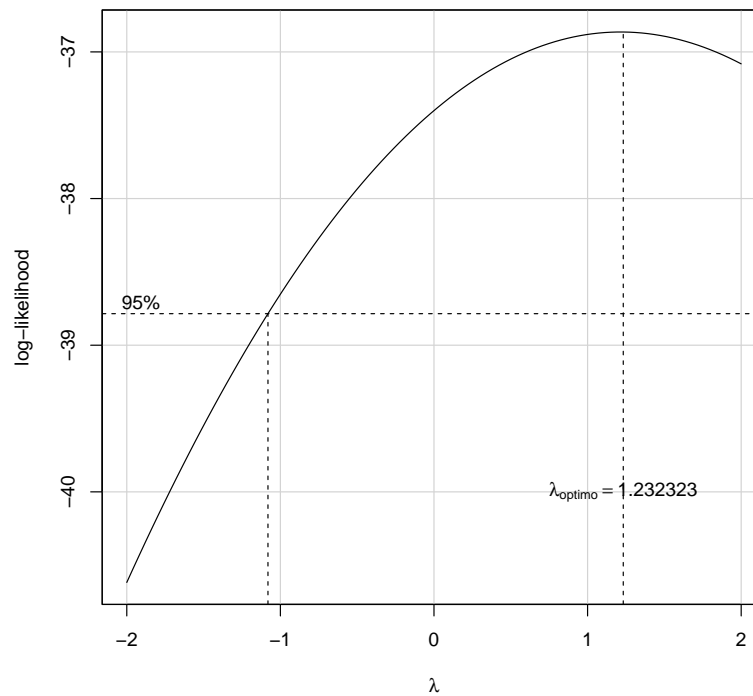


Figura A.1: Gráfico del log-verosimilitud vs. λ producido por función R `boxCox` sobre los datos del ejemplo 3.2.6. Observe que el valor de λ que maximiza el log-verosimilitud corresponde a $\lambda_{\text{óptimo}} = 1.232323$, sin embargo, $\lambda = 1$ está dentro del intervalo de confianza del 95 % construido por la función para el parámetro λ y además el valor del log-verosimilitud del modelo con $\lambda = 1$ es muy cercano al valor máximo de esta función.

Nota A.1. El IC del 95 % para λ en la Figura A.1 corresponde a un resultado basado en propiedades del estadístico conocido como LR, específicamente para el parámetro λ de la transformación Box-Cox,

$$\text{LR} = -2 \left[l(\lambda) - l(\hat{\lambda}) \right] \overset{a}{\sim} \chi_1^2, \quad (\text{A.10})$$

donde $l(\cdot)$ es el logaritmo natural de la función de verosimilitud para λ y $\hat{\lambda}$ es el estimador de máxima verosimilitud. Un IC aproximado de $(1 - \alpha) 100\%$ corresponde al intervalo de valores de λ tal que $LR \leq \chi_{\alpha,1}^2$, es decir, tales que $l(\lambda) \geq l(\hat{\lambda}) - \frac{1}{2}\chi_{\alpha,1}^2$, con $\chi_{\alpha,1}^2$ el percentil $(1 - \alpha) 100\%$ de la distribución chi cuadrado con un grado de libertad.

A.6. Análisis con varianzas desiguales

De acuerdo a [2], una alternativa a la aplicación de una transformación para homogenizar varianzas, es usar un método de análisis de datos que esté diseñado para el caso de varianza no constante. Por ejemplo, para la construcción de I.C. El método es aproximado y menos potente, sin embargo tiene la ventaja de mantener a las observaciones en sus unidades originales de medición. Para el ANOVA sin la presunción de varianza constante el modelo es:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \text{ con } \varepsilon_{ij} \sim N(0, \sigma_i^2) \text{ e independientes.} \quad (\text{A.11})$$

Para este modelo cada “contraste” $w = \sum_{i=1}^a c_i \alpha_i$, con $\sum_{i=1}^a c_i = 0$, permanece estimable, pero su estimador de mínimos cuadrados $\hat{w} = \sum_{i=1}^a c_i \hat{\alpha}_i = \sum_{i=1}^a c_i \bar{Y}_{i\bullet}$ tiene ahora varianza $\text{var}(\hat{w}) = \text{var}\left(\sum_{i=1}^a c_i \bar{Y}_{i\bullet}\right) = \sum_{i=1}^a c_i^2 \frac{\sigma_i^2}{n_i}$, la cual se estima usando para cada varianza su correspondiente estimador muestral. Entonces

$$T_0 = \frac{\sum_{i=1}^a c_i \hat{\alpha}_i - \sum_{i=1}^a c_i \alpha_i}{\sqrt{\widehat{\text{var}}\left(\sum_{i=1}^a c_i \hat{\alpha}_i\right)}} = \frac{\sum_{i=1}^a c_i \hat{\alpha}_i - \sum_{i=1}^a c_i \alpha_i}{\sqrt{\sum_{i=1}^a c_i^2 S_i^2 / n_i}} \underset{\text{aprox.}}{\sim} t_\nu, \quad (\text{A.12})$$

en donde los grados de libertad ν son calculados con la aproximación Satterthwaite,

$$\nu = \frac{\left(\sum_{i=1}^a c_i^2 S_i^2 / n_i\right)^2}{\sum_{i=1}^a \frac{(c_i^2 S_i^2 / n_i)^2}{n_i - 1}} \quad (\text{A.13})$$

Por tanto, un intervalo de confianza aproximado de $(1 - \gamma)100\%$ para w es,

$$w = \sum_{i=1}^a c_i \alpha_i \in \left(\sum_{i=1}^a c_i \hat{\alpha}_i \mp t_{\gamma/2, \nu} \sqrt{\widehat{\text{var}}\left[\sum_{i=1}^a c_i \hat{\alpha}_i\right]} \right). \quad (\text{A.14})$$

A.7. Deducción de los intervalos de confianza para los parámetros en modelo DCA un factor de efectos aleatorios

¿Cómo evaluar la igualdad de medias cuando no se cumple el supuesto de varianza constante?

Para probar la hipótesis de igualdad de medias,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a \text{ vs. } H_1 : \text{algún par } \mu_i \neq \mu_j,$$

en un DCA balanceado cuando no se cumple el supuesto de homogeneidad de varianza, se puede seguir la sugerencia de [1] la cual consiste en ajustar los grados de libertad del estadístico de prueba F de la siguiente manera:

$$F_0 = MSA/MSE \stackrel{\text{aprox.}}{\sim} f_{v_1, v_2} \text{ con} \quad (\text{A.15})$$

$$v_1 = \frac{a-1}{1 + c^2 \frac{(a-2)}{a-1}}, \quad v_2 = \frac{a(n-1)}{1 + c^2}, \quad \text{con } c^2 = \frac{\sum_{i=1}^a (S_i^2 - \bar{S}^2)^2}{a(\bar{S}^2)^2}, \quad \text{donde } \bar{S}^2 = \frac{1}{a} \sum_{i=1}^a S_i^2 \quad (\text{A.16})$$

También se puede probar la hipótesis de igualdad de medias mediante el test de Welch, tanto en un DCA balanceado como desbalanceado (ver [13]), en el cual el estadístico de la prueba es

$$F_0^* = \frac{\sum_{i=1}^a \omega_i (\bar{Y}_{i\bullet} - \hat{\mu})^2 / (a-1)}{1 + 2 \frac{(a-2)}{a^2-1} \sum_{i=1}^a h_i} \stackrel{\text{aprox.}}{\sim} f_{a-1, \eta}, \quad (\text{A.17})$$

donde

$$\begin{aligned} \omega_i &= \frac{n_i}{S_i^2}, \quad W = \sum_{i=1}^a \omega_i, \quad \hat{\mu} = \sum_{i=1}^a \omega_i \bar{Y}_{i\bullet} / W, \quad \eta = \frac{a^2 - 1}{3 \sum_{i=1}^a h_i} \\ h_i &= \frac{(1 - \omega_i / W)^2}{(n_i - 1)}, \quad S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^a (Y_{ij} - \bar{Y}_{i\bullet})^2. \end{aligned} \quad (\text{A.18})$$

A.7. Deducción de los intervalos de confianza para los parámetros en modelo DCA un factor de efectos aleatorios

Teniendo como válidos los supuestos sobre los efectos aleatorios A_i y los errores E_{ij} , se tienen los siguientes resultados.

A.7.1. I.C para σ^2

Desde que $SSE/\sigma^2 \sim \chi^2_{N-a}$, un intervalo del $(1 - \gamma)100\%$ para σ^2 se deduce despejando de la siguiente ecuación probabilística, el intervalo para el parámetro,

$$P\left(\chi^2_{1-\gamma/2, N-a} < \frac{SSE}{\sigma^2} < \chi^2_{\gamma/2, N-a}\right) = 1 - \gamma \quad (\text{A.19})$$

y por tanto, obtenemos

$$\frac{SSE}{\chi^2_{\gamma/2, N-a}} < \sigma^2 < \frac{SSE}{\chi^2_{1-\gamma/2, N-a}} \quad (\text{A.20})$$

A.7.2. I.C para σ_α^2

Dado que el estimador puntual de este parámetro es una combinación lineal de dos sumas de cuadrados (independientes), no existe una forma cerrada para deducir un intervalo de confianza exacto. Hay varios métodos para obtener intervalos de confianza aproximados del $(1 - \gamma)100\%$, veamos dos de ellos:

1. Se sabe que bajo la validez de los supuestos del modelo ANOVA un factor de efectos aleatorios en un DCA, $SSA/(\sigma^2 + c\sigma_\alpha^2) \sim \chi^2_{a-1}$ y que $SSE/\sigma^2 \sim \chi^2_{N-a}$, y además son mutuamente independientes. La distribución exacta de $\hat{\sigma}_\alpha^2$ está por tanto basada en la diferencia de dos variables aleatorias chi cuadrado cada una multiplicada respectivamente por una constante de valor desconocido, y para ésta no hay una distribución estándar conocida; sin embargo, una aproximación razonable a la verdadera distribución de $\hat{\sigma}_\alpha^2/\sigma_\alpha^2$ es una distribución chi cuadrado dividida por sus grados de libertad de κ , donde κ es estimada por:

$$\kappa = \frac{(MSA - MSE)^2}{\frac{MSA^2}{a-1} + \frac{MSE^2}{N-a}} \quad (\text{A.21})$$

Es decir, la distribución de $\kappa \frac{\hat{\sigma}_\alpha^2}{E[\hat{\sigma}_\alpha^2]} \stackrel{approx.}{\sim} \chi^2_\kappa$. Este resultado está relacionado a la aproximación Satterthwaite (ver Apéndice A.8). Usando esta aproximación podemos escribir la siguiente declaración de probabilidad aproximada:

$$P\left(\chi^2_{1-\gamma/2, \kappa} < \frac{\kappa \hat{\sigma}_\alpha^2}{\sigma_\alpha^2} < \chi^2_{\gamma/2, \kappa}\right) \approx 1 - \gamma \quad (\text{A.22})$$

A.7. Deducción de los intervalos de confianza para los parámetros en modelo DCA un factor de efectos a

Despejando la desigualdad para σ_α^2 a partir de ésta expresión, obtenemos el intervalo de confianza aproximado (*) presentado en la Tabla 5.2,

$$\frac{\kappa \hat{\sigma}_\alpha^2}{\chi_{\gamma/2, \kappa}^2} < \sigma_\alpha^2 < \frac{\kappa \hat{\sigma}_\alpha^2}{\chi_{1-\gamma/2, \kappa}^2} \quad (\text{A.23})$$

2. Otra aproximación un poco más precisa que la anterior da un intervalo de confianza de al menos $(1 - 2\gamma)100\%$, y es la dada en la ecuación (**) de la Tabla 5.2,

$$\frac{(F_0 - f_{\gamma/2, a-1, N-a})}{cF_0 \chi_{\gamma/2, a-1}^2} SSA < \sigma_\alpha^2 < \frac{(F_0 - f_{1-\gamma/2, a-1, N-a})}{cF_0 \chi_{1-\gamma/2, a-1}^2} SSA \quad (\text{A.24})$$

3. Otra aproximación es el procedimiento MLS (*Modified Large Sample Method*), el cual es más complejo que la aproximación Satterthwaite y que ha sido propuesto para estimar una combinación lineal de dos cuadrados medios esperados en diseños balanceados, puede consultarse en Kutner, et. al. (2005). *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill Irwin, Capítulo 25.

A.7.3. I.C para σ_α^2/σ^2

Bajo los supuestos del modelo ANOVA, se tiene que:

$$\frac{MSA/(\sigma^2 + c\sigma_\alpha^2)}{MSE/\sigma^2} = \frac{MSA}{MSE \left(1 + c \frac{\sigma_\alpha^2}{\sigma^2}\right)} \sim f_{a-1, N-a} \quad (\text{A.25})$$

Por tanto, podemos formular la siguiente declaración probabilística:

$$P \left(f_{1-\gamma/2, a-1, N-a} < \frac{F_0}{\left(1 + c \frac{\sigma_\alpha^2}{\sigma^2}\right)} < f_{\gamma/2, a-1, N-a} \right) = 1 - \gamma \quad (\text{A.26})$$

donde $F_0 = MSA/MSE$. A partir de lo cual, haciendo la manipulación algebraica apropiada, se llega a que un I.C del $(1 - \gamma)100\%$, para la razón σ_α^2/σ^2 está dado por:

$$\frac{1}{c} \left[\frac{F_0}{f_{\gamma/2, a-1, N-a}} - 1 \right] < \frac{\sigma_\alpha^2}{\sigma^2} < \frac{1}{c} \left[\frac{F_0}{f_{1-\gamma/2, a-1, N-a}} - 1 \right] \quad (\text{A.27})$$

A.7.4. I.C para $\sigma_\alpha^2/(\sigma^2 + \sigma_\alpha^2)$

Esta razón tomará valores entre 0 (cuando la componente σ_α^2 es cero) y 1 (cuando la componente σ_α^2 es mucho mayor que σ^2). Pero esta proporción tiene también otro

significado: Es el coeficiente de correlación entre cualesquiera dos respuestas en el mismo nivel del factor, y en este sentido es llamada el **coeficiente de correlación intra grupo o intra clase**, es decir, es una medida de similitud de las observaciones dentro de grupos. Partiendo de la siguiente declaración de probabilidad,

$$P \left(f_{1-\gamma/2, a-1, N-a} < \frac{MSA/(\sigma^2 + c\sigma_\alpha^2)}{MSE/\sigma^2} < f_{\gamma/2, a-1, N-a} \right) = 1 - \gamma \quad (\text{A.28})$$

Despejando para la razón $\sigma_\alpha^2/(\sigma^2 + \sigma_\alpha^2)$, se obtiene que el intervalo de confianza de $(1 - \gamma)100\%$ está dado por:

$$\frac{L}{1+L} < \frac{\sigma_\alpha^2}{\sigma^2 + \sigma_\alpha^2} < \frac{U}{1+U} \quad \text{donde} \quad (\text{A.29})$$

$$L = \frac{1}{c} \left(\frac{F_0}{f_{\gamma/2, a-1, N-a}} - 1 \right) \quad \text{y} \quad U = \frac{1}{c} \left(\frac{F_0}{f_{1-\gamma/2, a-1, N-a}} - 1 \right) \quad (\text{A.30})$$

Nota A.2. En los intervalos de confianza anteriores en los cuales aparece σ_α^2 , puede resultar el límite inferior negativo, caso en el cual éste se aproxima por cero.

A.7.5. I.C para μ

Para deducir este intervalo es necesario estudiar la distribución muestral de su estimador $\bar{Y}_{\bullet\bullet}$. Veamos.

Desde que $\bar{Y}_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}$, y usando la ecuación (5.1), tenemos que

$$\bar{Y}_{\bullet\bullet} = \mu + \frac{1}{N} \sum_{i=1}^a n_i A_i + \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} \varepsilon_{ij} \quad (\text{A.31})$$

Luego, la media y varianza de este estimador corresponden a $E[\bar{Y}_{\bullet\bullet}] = \mu$, $V[\bar{Y}_{\bullet\bullet}] = \left(\sum_{i=1}^a \frac{n_i^2}{N^2} \right) \sigma_\alpha^2 + \frac{1}{N} \sigma^2$, respectivamente, de donde un estimador de su varianza es,

$$S_{\bar{Y}_{\bullet\bullet}}^2 = \left(\sum_{i=1}^a \frac{n_i^2}{N^2} \right) \left(\frac{MSA - MSE}{c} \right) + \frac{1}{N} MSE \quad (\text{A.32})$$

$$S_{\bar{Y}_{\bullet\bullet}}^2 = \left(\sum_{i=1}^a \frac{n_i^2}{cN^2} \right) MSA + \left(\frac{1}{N} - \sum_{i=1}^a \frac{n_i^2}{cN^2} \right) MSE.$$

Aplicando la aproximación Satterthwaite (ver Apéndice A.8), se tiene que

$$\nu^* \times \frac{S_{\bar{Y}_{\bullet\bullet}}^2}{V[\bar{Y}_{\bullet\bullet}]} \stackrel{approx.}{\sim} \chi_{\nu^*}^2, \quad (\text{A.33})$$

donde

$$\nu^* = \frac{(a_1 MSA + a_2 MSE)^2}{\frac{(a_1 MSA)^2}{a-1} + \frac{(a_2 MSE)^2}{N-a}} \quad \text{con} \quad a_1 = \sum_{i=1}^a \frac{n_i^2}{cN^2}, \quad a_2 = \frac{1}{N} - \sum_{i=1}^a \frac{n_i^2}{cN^2}. \quad (\text{A.34})$$

Por lo tanto,

$$T = \frac{(\bar{Y}_{\bullet\bullet} - \mu) / \sqrt{V[\bar{Y}_{\bullet\bullet}]}}{\sqrt{S_{\bar{Y}_{\bullet\bullet}}^2 / V[\bar{Y}_{\bullet\bullet}]}} = \frac{\bar{Y}_{\bullet\bullet} - \mu}{S_{\bar{Y}_{\bullet\bullet}}} = \frac{Z}{\sqrt{\chi_{\nu^*}^2 / \nu^*}} \stackrel{\text{aprox.}}{\sim} t_{\nu^*}, \quad (\text{A.35})$$

donde Z es una variable aleatoria normal estándar. A partir de la siguiente declaración de probabilidad se deduce el I.C aproximado para la media global:

$$P(-t_{\gamma/2, \nu^*} < T < t_{\gamma/2, \nu^*}) \simeq 1 - \gamma. \quad (\text{A.36})$$

A.8. Procedimiento Satterthwaite, [7]

Éste es un procedimiento general para construir intervalos de confianza aproximados para combinaciones lineales de valores esperados de cuadrados medios. Note por ejemplo que, $\sigma_\alpha^2 = \frac{1}{c} E[MSA] + \left(-\frac{1}{c}\right) E[MSE]$. En general se hablará de una combinación lineal de valores esperados de cuadrados medios dada por:

$$L = a_1 E[MS_1] + a_2 E[MS_2] + \cdots + a_h E[MS_h], \quad (\text{A.37})$$

donde los a_i son los coeficientes de la combinación. Un estimador insesgado de L es

$$\hat{L} = a_1 MS_1 + a_2 MS_2 + \cdots + a_h MS_h. \quad (\text{A.38})$$

Denotemos por v_i los grados de libertad asociados a la suma de cuadrados medios MS_i . Satterthwaite sugirió la siguiente aproximación distribucional: si los h cuadrados medios son independientes, entonces:

$$\kappa \frac{\hat{L}}{L} \stackrel{\text{aprox.}}{\sim} \chi_\kappa^2, \quad (\text{A.39})$$

donde

$$\kappa = \frac{(a_1 MS_1 + a_2 MS_2 + \cdots + a_h MS_h)^2}{\frac{(a_1 MS_1)^2}{v_1} + \frac{(a_2 MS_2)^2}{v_2} + \cdots + \frac{(a_h MS_h)^2}{v_h}}. \quad (\text{A.40})$$

Entonces un intervalo de confianza aproximado de $(1 - \gamma)100\%$, para L , corresponde a:

$$\kappa \frac{\hat{L}}{\chi_{\gamma/2, \kappa}^2} < L < \kappa \frac{\hat{L}}{\chi_{1-\gamma/2, \kappa}^2}. \quad (\text{A.41})$$

Bibliografía

- [1] Box, G. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effects of inequality of variance in the one-way classification". *Annals of Mathematical Statistics*, 25(2), pp. 290-302.
- [2] Dean, A., Voss, D., and Draguljić, D. (2017). *Design and Analysis of Experiments*, 2nd Edition. Springer.
- [3] Gómez López, H. (1997). *Estadística Experimental Aplicada a las Ciencias Agrícolas*. Universidad Nacional de Colombia, Sede Medellín, Facultad de Ciencias Agropecuarias.
- [4] Gutiérrez Pulido, H. y de la Vara Salazar, R. (2012). *Análisis y Diseño de Experimentos*, 3^a Edición. McGraw-Hill.
- [5] Hinkelmann, K. and Kempthorne, O. (2008) *Design and Analysis of Experiments. Volume I: Introduction to Experimental Design*, 2nd Edition. Wiley-Interscience.
- [6] Kuehl, R. O. (2001). *Diseño de Experimentos. Principios estadísticos de diseño y análisis de investigación*, 2^a Edición, Thomson Learning.
- [7] Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill Irwin.
- [8] Mendoza, H., Bautista, G. (2002). *Diseño Experimental*, Universidad Nacional de Colombia, <http://www.virtual.unal.edu.co/cursos/ciencias/2000352/>. Licencia: Creative Commons BY-NC-ND.

- [9] Milliken, G. A. and Johnson, D. E. (2009). *Analysis of Messy Data, Volume 1: Designed Experiments*, 2nd Edition. CRC Press, Taylor and Francis Group.
- [10] Montgomery, D. C. (2020). *Design and Analysis of Experiments*, 10th Edition. John Wiley & Sons, Inc.
- [11] 't Lam, R.U.E. (2010). "Scrutiny of variance results for outliers: Cochran's test optimized". *Analytica Chimica Acta*, 659(1), pp. 68-84.
- [12] Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance components*. John Wiley & Sons, Inc.
- [13] Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach". *Biometrika*, 38, pp. 330-336.
- [14] Wu, C. F. J. and Hamada, M. (2009). *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley & Sons, Inc.