

Capítulo 3

Comparaciones múltiples

3.1. Introducción

En este capítulo explicaremos algunas técnicas para analizar con mayor detalle los datos de un experimento, con posterioridad a la realización del Análisis de la Varianza. Si dicho análisis confirma la existencia de diferencias significativas entre los tratamientos, es conveniente investigar qué medias son distintas. Para ello, emplearemos diversas técnicas cuyo objeto es identificar qué tratamientos son estadísticamente diferentes y en cuánto oscila el valor de esas diferencias. Consideraremos su aplicación únicamente al modelo de efectos fijos.

El uso de estas técnicas, en algunos casos, está supeditado al resultado del análisis de la varianza; en otros casos, las técnicas pueden emplearse directamente sin haber realizado previamente dicho análisis. Este conjunto de técnicas se engloba bajo la denominación de *contrastes para comparaciones múltiples* ya que su objetivo fundamental es comparar entre sí medias de tratamientos o grupos de ellas.

En primer lugar estudiamos un procedimiento intuitivo y cualitativo basado en la representación gráfica de los datos del experimento¹. Después del método gráfico consideramos la técnica de comparación por parejas introducida por Fisher en 1935. Dicha técnica, denominada *método de la diferencia mínima significativa o método LSD* (Least Significant Difference), se basa en la construcción de tests de hipótesis para la diferencia de cualquier par de medias.

Cuando el número posibles de comparaciones es elevado, la aplicación reiterada de este procedimiento, para un nivel de significación α dado, puede conducir a un número grande de rechazos de la hipótesis nula aunque no existan diferencias reales. El intento de

¹En cierto sentido, el método gráfico ya fué comentado en el capítulo anterior al dar la interpretación de algunas salidas gráficas de STATGRAPHICS.

paliar el problema de los falsos rechazos justifica la introducción de otros procedimientos para comparaciones múltiples. Entre estos métodos se estudia primero el basado en la *desigualdad de Bonferroni*. A continuación se aborda otra forma de solventar las deficiencias mencionadas anteriormente, basada en el *rango estudentizado*, que da lugar al *método de la diferencia significativa honesta* propuesto por Tukey o método HSD (Honestly Significant Difference) y a los procedimientos de rangos múltiples de *Newman-Keuls y Duncan*.

Finalmente se estudia el procedimiento general de *Scheffé*, basado en la construcción de intervalos de confianza simultáneos para todas las posibles diferencias de medias y que permite su extensión a comparaciones más generales denominadas *contrastos*.

En la sección 3.3 se justifica qué procedimientos pueden ser adecuados para cada situación y cuáles son las condiciones experimentales que los hacen recomendables.

Otro problema de índole diferente, también ligado a las comparaciones múltiples, consiste en las comparaciones de distintos tratamientos con un *control* que suele emplearse en determinados campos de investigación, por ejemplo en estudios epidemiológicos. Dicho problema se aborda mediante *el contraste de Dunnett*.

3.2. Comparaciones múltiples de medias

Al estudiar el comportamiento de los tratamientos de un factor, mediante un análisis de la varianza, el único objetivo es saber si, globalmente, dichos tratamientos difieren significativamente entre sí. Ahora estamos interesados, una vez aceptada la existencia de diferencias entre los efectos del factor, en conocer qué tratamientos concretos producen mayor efecto o cuáles son los tratamientos diferentes entre sí. En estas mismas condiciones, puede ser útil también realizar comparaciones adicionales entre grupos de medias de los tratamientos.

Antes de seguir con el desarrollo de esta sección, remitimos al lector al Capítulo 1, sección ??, donde presentamos el modelo de efectos fijos definido por la ecuación

$$y_{ij} = \mu_i + u_{ij} \quad . \quad (3.1)$$

Antes de estudiar analíticamente las distintas formas de comparar efectos, es conveniente examinar de forma cualitativa su comportamiento. Esto se efectúa de forma gráfica y constituye el primer método que pasamos a desarrollar.

3.2.1. Comparaciones gráficas de medias

Al igual que en el capítulo anterior vamos a concretar toda la explicación sobre un determinado ejemplo. En este caso, vamos a considerar el Ejemplo 1-1.

Para ese ejemplo obtuvimos, véase Tabla 1-6, que el rendimiento de la semilla de algodón depende significativamente del tipo de fertilizante utilizado. Pero esta conclusión no nos informa nada sobre una serie de preguntas tales como: ¿Producen algunos fertilizantes el mismo rendimiento?, en el caso de que dos fertilizantes produzcan distinto rendimiento ¿entre qué valores oscila esa diferencia?, etc. Una primera respuesta a estas preguntas se puede realizar de forma gráfica.

En primer lugar haremos unas consideraciones generales antes de estudiar el ejemplo de referencia. Para ello, consideremos el modelo equilibrado con n observaciones por nivel y sean $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_I$ las medias muestrales de los correspondientes tratamientos. Bajo las hipótesis del modelo, cualquier promedio \bar{y}_i de un tratamiento se distribuirá alrededor de su media μ_i con una desviación estándar σ/\sqrt{n} . En consecuencia, si aceptamos que las medias de los distintos niveles de un factor son todas idénticas, entonces las medias muestrales $\bar{y}_i, i = 1, 2, \dots, I$, se comportarán como un conjunto de I observaciones tomadas al azar de una misma distribución normal con media μ y desviación estándar σ/\sqrt{n} .

Supongamos que dibujamos la función de densidad de una distribución normal “deslizante” e imaginemos que situamos dicha distribución sobre los puntos que representan las medias de los tratamientos. Procediendo de esta manera y teniendo en cuenta los distintos comportamientos de los efectos, podemos distinguir dos casos relevantes:

a) *Los efectos de los tratamientos son iguales entre sí.*

Entonces, al deslizar la distribución a lo largo del eje OX , es probable que exista alguna posición para la cual todos los valores \bar{y}_i sean compatibles con dicha distribución; es decir, sean una muestra aleatoria de la misma distribución.

b) *Los efectos de los tratamientos no son iguales.*

Entonces al deslizar la distribución a lo largo del eje OX , es improbable que exista alguna posición para la cual todos los valores \bar{y}_i sean compatibles con dicha distribución. En este caso, es posible que existan situaciones que permitan detectar niveles del factor que producen respuestas medias diferentes; concretamente, éstos se identifican por medio de aquellos valores \bar{y}_i que están muy distanciados de los otros.

Como generalmente σ^2 es desconocida estimamos su valor por medio de la varianza residual \hat{S}_R^2 y, en lugar de la distribución normal, hay que emplear una distribución t deslizante con factor de escala $\sqrt{\hat{S}_R^2/n}$ y con los grados de libertad correspondientes a dicha varianza residual. Para dibujar dicha distribución t utilizamos la Tabla III del Apéndice C, que muestra el valor de la función de densidad, para determinadas abscisas, de una distribución t de Student con grados de libertad dados. Para obtener la distribución con

factor de escala hay que multiplicar las correspondientes abscisas por dicho factor de escala y representar sus valores frente a las correspondientes ordenadas.

Este método se ilustra en la Figura 2-1 donde se distinguen varios casos:

- a) Las cinco medias pueden considerarse como una muestra aleatoria de la misma distribución.

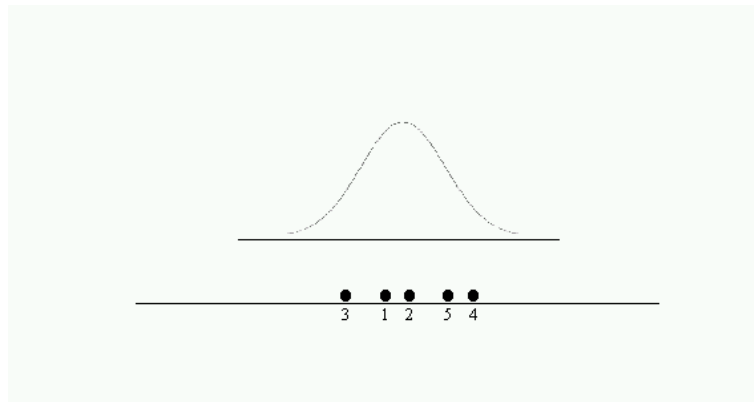


Figura 2-1a

- b) Al deslizar la distribución de referencia a lo largo del eje horizontal, nos permite detectar grupos homogéneos. Así, deducimos que las medias 2, 4 y 5 son probablemente mayores que 1 y 3 y que no hay distinción entre las medias de cada uno de los grupos.

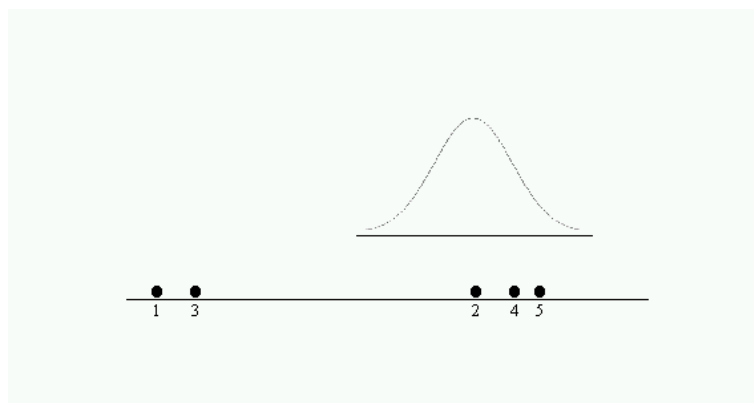


Figura 2-1b

- c) Podemos deslizar la distribución de referencia de forma que abarque a las cinco medias, pero no parece razonable concluir que las cinco medias sean compatibles con la misma población.

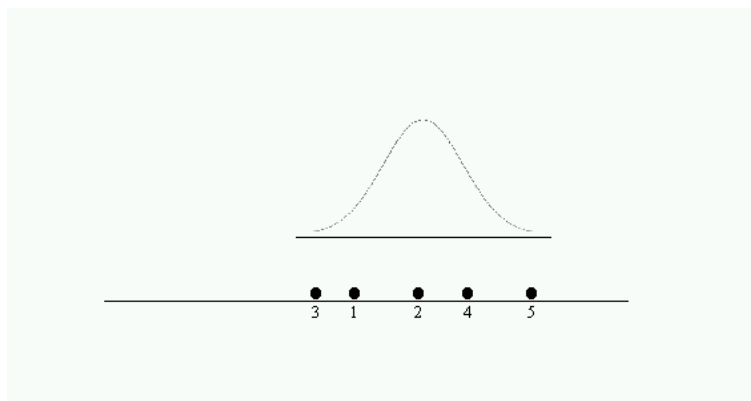


Figura 2-1c

En estos casos parece razonable que el resultado que proporciona el análisis de la varianza sea aceptar la hipótesis H_0 en el primer caso y rechazarla en los otros dos.

Este procedimiento gráfico se ha desarrollado para el caso en que todos los tamaños muestrales sean iguales. Sin embargo, de forma aproximada, se puede aplicar también cuando las diferencias entre los tamaños muestrales no sean muy grandes. En tal caso, debe usarse

$$\sqrt{\frac{\hat{S}_R^2}{\bar{n}}},$$

como factor de escala, siendo \bar{n} el tamaño medio de las muestras, $\bar{n} = \frac{1}{I} \sum_i n_i$.

La aplicación de este método a nuestro ejemplo de referencia se muestra en la Figura 2-2

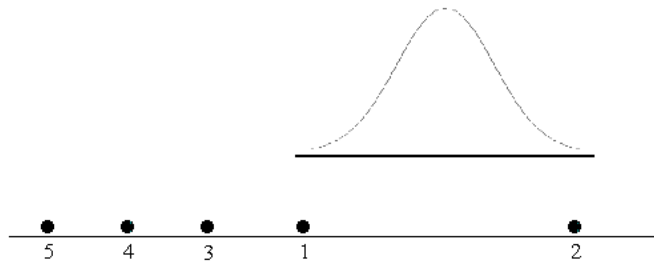


Figura 2-2: Comparación gráfica de los tratamientos del Ejemplo 1-1.

En dicho ejemplo, como la varianza es desconocida y los tamaños muestrales son distintos, el factor de escala es igual a

$$\sqrt{\frac{\widehat{S}_R^2}{\bar{n}}} = \sqrt{\frac{4,67}{26/5}} = 0,947 \quad .$$

Para construir la distribución *deslizante*, como el número de grados de libertad de la varianza residual es 21, tomamos los valores de la correspondiente fila de la Tabla III del Apéndice C, obteniendo:

Abcisa	0	0.5	1.0	1.5	2.0
Ordenada	0.394	0.346	0.236	0.129	0.058
Abcisa $\times 0,947$	0	0.473	0.947	1.420	1.894

Para dibujar la distribución de referencia, se elige un origen arbitrario, que debe estar próximo a las medias, nosotros hemos considerado como origen $\widehat{\mu}$. A continuación se representan las ordenadas de los puntos $\widehat{\mu} \pm 0$, $\widehat{\mu} \pm 0,473$, $\widehat{\mu} \pm 0,947$, $\widehat{\mu} \pm 1,420$, $\widehat{\mu} \pm 1,894$ que se unen en una línea continua, con lo que esbozamos la distribución de referencia. El deslizamiento de la distribución de referencia a lo largo del eje horizontal permite obtener las siguientes conclusiones:

- Las cinco medias no son iguales entre sí. No es posible situar la distribución de referencia en ninguna posición de forma que las abarque a todas.
- Los fertilizantes (5 y 4), (4 y 3) y (3 y 1) tienen un rendimiento similar y el fertilizante 2 tiene un rendimiento distinto de los otros cuatro.

Este procedimiento de comparación gráfica de las medias de los tratamientos mediante una distribución de referencia, es un método aproximado pero eficaz en muchos problemas de comparaciones múltiples. En las siguientes secciones se describen otros métodos estadísticos más rigurosos para realizar comparaciones múltiples. En primer lugar se abordan métodos creados específicamente para comparar por parejas los efectos de tratamientos, de entre ellos, comenzaremos con los basados en la distribución t .

3.2.2. Comparaciones basadas en la distribución t

Las pruebas estadísticas para comparaciones múltiples más frecuentemente utilizadas se basan en la distribución t de Student. Supongamos que interesa comparar por parejas los efectos de I tratamientos. Es decir, nos interesa contrastar cualquier hipótesis de la forma

$$\begin{aligned} H_0 : & \mu_i = \mu_j \\ H_1 : & \mu_i \neq \mu_j \end{aligned} \quad (3.2)$$

La técnica más antigua y popular para efectuar estas comparaciones múltiples es el procedimiento *LSD*, (*Least Significant Difference*), que pasamos a estudiar a continuación.

Contraste de la Mínima Diferencia Significativa

Este procedimiento fué sugerido por Fisher en 1935 y es el primer método de comparaciones múltiples que vamos a utilizar. Dicho procedimiento consiste en una prueba de hipótesis por parejas basada en la distribución t .

Este método debe aplicarse cuando previamente se haya rechazado la hipótesis nula del test F del análisis de la varianza. Para ello, se determina el siguiente estadístico

$$t = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{\hat{S}_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad (3.3)$$

que, por las hipótesis del modelo de ecuación (??), sigue una distribución t de Student con $N - I$ grados de libertad.

Por lo tanto, se concluye que la pareja de medias μ_i y μ_j son estadísticamente diferentes si

$$| \bar{y}_i - \bar{y}_j | > LSD \quad , \quad (3.4)$$

donde la cantidad *LSD*, denominada *mínima diferencia significativa*, viene dada por

$$LSD = t_{\alpha/2; N-I} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad , \quad (3.5)$$

siendo

- * n_i y n_j el número de observaciones correspondiente a cada media,
- * $N - I$ el número de grados de libertad de la varianza residual,
- * $t_{\alpha/2; N-I}$ el valor crítico de la distribución t con $N - I$ grados de libertad que deja una probabilidad a su derecha igual a $\alpha/2$.

Si el diseño es balanceado, el valor de LSD se reduce a

$$LSD = t_{\alpha/2; N-I} \sqrt{\frac{2\hat{S}_R^2}{n}} . \quad (3.6)$$

El procedimiento LSD es sencillo de utilizar; se puede aplicar tanto en modelos equilibrados como no-equilibrados. Además proporciona también intervalos de confianza para diferencias de medias. Dichos intervalos son de la forma

$$((\bar{y}_i. - \bar{y}_j.) - LSD , (\bar{y}_i. - \bar{y}_j.) + LSD) \quad (3.7)$$

Comentarios 3.1

C1) *Un problema que presenta la aplicación de este procedimiento, para un número relativamente grande de tratamientos, es que el número de posibles falsos rechazos de la hipótesis nula puede ser elevado aunque no existan diferencias reales. Así, por ejemplo, si I es igual a 10, hay $\binom{10}{2} = 45$ posibles parejas de comparaciones. Si tomamos $\alpha = 0,05$, entonces el número esperado de falsos rechazos será*

$$0,05 \times 45 = 2,25 ,$$

supuesta independencia estadística entre los distintos contrastes.

Otra forma de enfocar este problema es comprobando que la utilización reiterada de este procedimiento conduce a que la probabilidad de que se rechace al menos una de las posibles comparaciones sea bastante alta. En efecto, bajo la hipótesis nula

$$\Pr[\text{rechazar al menos una comparación}] = 1 - \Pr[\text{aceptar todas}] =$$

$$1 - (1 - 0,05)^{45} \simeq 0,90$$

El intento de paliar este problema justifica la introducción de otros procedimientos para comparaciones múltiples.

C2) Puede suceder que el método LSD falle al aceptar que todas las parejas son iguales, a pesar de que el estadístico F del análisis de la varianza resulte significativo; esto es debido a que la prueba F considera simultáneamente todas las posibles comparaciones entre las medias de los tratamientos y no sólo las comparaciones por parejas.

A fin de ilustrar este procedimiento, consideramos el ejemplo de referencia. Así, para un nivel de significación del 5 %, el valor de LSD es

$$LSD = t_{0,025;21} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = 2,080 \sqrt{4,67} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = 4,49 \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Por tanto, dos medias difieren significativamente entre sí, si el valor absoluto de la diferencia de sus promedios muestrales es mayor que el valor LSD . Recordemos que, en este ejemplo de referencia, los valores de los cinco promedios de los tratamientos son:

$$\bar{y}_{1.} = 50, \quad \bar{y}_{2.} = 57, \quad \bar{y}_{3.} = 48, \quad \bar{y}_{4.} = 47, \quad \bar{y}_{5.} = 45 \quad .$$

Para la aplicación de este procedimiento es recomendable ordenar las medias de menor a mayor. Una vez ordenadas se comparan la 1ª y la 2ª:

- a) Si hay diferencia significativa entre ellas, también la habrá entre la 1ª y todas las demás.
- b) Si no hay diferencia significativa entre ellas, se compara la 1ª con la 3ª y siguientes hasta encontrar la primera diferencia significativa.

De forma similar se realiza la comparación de la 2ª media con todas las demás y así sucesivamente se haría con las restantes medias.

Por lo tanto,

1º) Ordenamos las medias de menor a mayor.

$$\bar{y}_{5.} = 45, \quad \bar{y}_{4.} = 47 \quad \bar{y}_{3.} = 48, \quad \bar{y}_{1.} = 50, \quad \bar{y}_{2.} = 57 \quad .$$

2º) Comparamos las medias 1ª y 2ª.

$$LSD = 4,49 \sqrt{1/6 + 1/4} = 2,898 \quad ; \quad | \bar{y}_{5.} - \bar{y}_{4.} | = | 45 - 47 | = 2 \quad .$$

Como no hay diferencia significativa entre ellas, se compara la 1ª con la 3ª

Al haber diferencia significativa entre ellas, también la hay entre la 1ª y la 4ª y entre la 1ª y la 5ª.

$$LSD = 4,49\sqrt{1/4 + 1/5} = 3,011 \quad ; \quad |\bar{y}_4 - \bar{y}_3| = |47 - 48| = 1 \quad .$$
$$LSD = 4,49\sqrt{1/4 + 1/6} = 2,898 \quad ; \quad |\bar{y}_4. - \bar{y}_1. | = |47 - 50| = 3 \quad .$$
$$LSD = 4,49\sqrt{1/5 + 1/6} = 2,718 \quad ; \quad |\bar{y}_{3.} - \bar{y}_{1.}| = |48 - 50| = 2 \quad .$$
$$LSD = 4,49\sqrt{1/5 + 1/5} = 2,839 \quad ; \quad |\bar{y}_3. - \bar{y}_2. | = |48 - 57| = 9 \quad .$$
$$LSD = 4,49\sqrt{1/6 + 1/5} = 2,718 \quad ; \quad |\bar{y}_1. - \bar{y}_2. | = | 50 - 57 | = 7 \quad .$$

Normalmente el resultado de las comparaciones por parejas se suele representar en forma gráfica, como se muestra a continuación

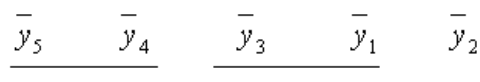


Figura 2-3

En dicha representación los valores de las medias de los tratamientos aparecen en orden creciente y se subrayan aquellos grupos de medias que resultan equivalentes. Esta representación gráfica es muy fácil de interpretar y es la utilizada por la mayoría de los paquetes estadísticos.

En la siguiente sección se describe el método de Bonferroni, que presenta una solución al problema de los *falsos rechazos* planteado en el método *LSD*.

Método de Bonferroni

En este procedimiento se fija un nivel de significación α que se reparte entre cada una de las comparaciones consideradas y se utiliza la desigualdad de Bonferroni

$$\Pr\left(\bigcup_{m=1}^M A_m\right) \leq \sum_{m=1}^M \Pr(A_m) \quad . \quad (3.8)$$

Consideremos que queremos realizar estimación por intervalos para las $M = \binom{I}{2}$ comparaciones posibles, cada una al nivel de significación $\alpha^* = \alpha/M$; esto da origen a M intervalos de confianza que contienen a cada una de las posibles diferencias $\mu_i - \mu_j$ con probabilidad $1 - \alpha^*$. Llamando C_m al intervalo m -ésimo se tiene que

$$\Pr[\mu_{1m} - \mu_{2m} \in C_m] = 1 - \alpha^* \quad m = 1, 2, \dots, M$$

siendo μ_{1m} y μ_{2m} la primera y segunda media de la correspondiente comparación, donde supondremos que $1 \leq 1m < 2m \leq I$.

Aplicando la desigualdad de Bonferroni, (expresión (3.8)), se tendrá

$$\Pr\left(\bigcap_{m=1}^M C_m\right) = 1 - \Pr\left(\bigcup_{m=1}^M \bar{C}_m\right) \geq 1 - \sum_{m=1}^M \Pr(\bar{C}_m) = 1 - \sum_{m=1}^M \alpha^* \quad ,$$

notamos por \bar{C}_m al complementario del intervalo C_m .

Teniendo en cuenta este resultado y si queremos garantizar un nivel de significación α para el conjunto de las M comparaciones por parejas, o un nivel de confianza $1 - \alpha$ para el conjunto de intervalos, basta tomar

$$\alpha^* = \frac{\alpha}{M} \quad .$$

Con ello, evidentemente, la probabilidad de que todos los intervalos, C_m , contengan a la correspondiente diferencia de medias, será al menos $1 - \alpha$. Así pues, los intervalos quedarán en la forma

$$\bar{y}_{1m} - \bar{y}_{2m} \pm t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}} \right)} , \quad (3.9)$$

donde \bar{y}_{1m} , \bar{y}_{2m} y n_{1m} , n_{2m} , son las medias y los tamaños muestrales correspondientes a la comparación m -ésima.

Denotamos por $\theta_m = \mu_{1m} - \mu_{2m}$, $m = 1, 2, \dots, M$, una de las M comparaciones lineales por parejas de medias, para las cuales interesa contrastar $H_0 : \theta_m = 0$ frente $H_1 : \theta_m \neq 0$.

Entonces, se rechaza H_0 si

$$|\theta_m| > B_m ,$$

y se acepta en caso contrario.

Donde

$$B_m = t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}} \right)} . \quad (3.10)$$

En el caso del modelo equilibrado los valores de B_m coinciden, dichos valores se denotan por BSD y tienen la siguiente expresión

$$BSD = t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n} + \frac{1}{n} \right)} , \quad (3.11)$$

siendo n el número de observaciones de cada grupo y $t_{\alpha/2M}$ el valor crítico de la distribución t , con el mismo número de grados de libertad que la varianza residual, que deja una probabilidad $\alpha/2M$ a su derecha.

Para ilustrar este método consideramos el ejemplo de referencia. Puesto que es un modelo no-equilibrado, los valores de los estadísticos de contraste vendrán dados por la siguiente expresión

$$B_m = t_{0.025, 21} \sqrt{(4,67) \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}} \right)} \quad m = 1, 2, \dots, 10 .$$

Así, por ejemplo, si comparamos

* 5 vs² 4, el valor de B_m es

$$B_1 = t_{\frac{0.025}{10}; 21} \sqrt{4,67 \left(\frac{1}{6} + \frac{1}{4} \right)} = (3,135)(1,394) = 4,370 \quad .$$

Entonces, como

$$| \bar{y}_{5.} - \bar{y}_{4.} | = 2 < B_1 \quad ,$$

las medias 5 y 4 no son significativamente distintas

* 5 vs 3, el valor de B_m es

$$B_2 = t_{\frac{0.025}{10}; 21} \sqrt{4,67 \left(\frac{1}{6} + \frac{1}{5} \right)} = (3,135)(1,308) = 4,100 \quad .$$

Entonces, como

$$| \bar{y}_{5.} - \bar{y}_{3.} | = 3 < B_2 \quad ,$$

las medias 5 y 3 no son significativamente distintas.

El resultado de las comparaciones por parejas realizadas con este procedimiento se muestra, en forma gráfica, a continuación

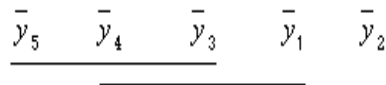


Figura 2-4

Seguidamente presentamos los contrastes de rangos múltiples. Dichos contrastes se basan en la distribución del rango estudentizado y se distinguen tres métodos debidos respectivamente a Tukey, Duncan-Newman y Keuls.

²Se utiliza “i vs j”, (i versus j), para denotar la comparación de μ_i frente a μ_j .

3.2.3. Tests de rangos múltiples

Estos contrastes se basan en la distribución del rango estudentizado, cuya definición se realiza en términos del número de grupos que hay que comparar y de los grados de libertad del estimador de la varianza.

Estos procedimientos, al igual que el procedimiento de Bonferroni, permiten superar las dificultades que surgen al aumentar el número de grupos a comparar y no poderse controlar los falsos rechazos de la hipótesis nula. Los métodos resultantes, en términos generales, son conservadores; es decir, la probabilidad real de rechazar la hipótesis nula cuando es cierta es menor que el nivel de significación α fijado.

Para definir el *recorrido* o *rango estudentizado*, supongamos que disponemos de k observaciones independientes y_1, y_2, \dots, y_k de una distribución Normal con media μ y varianza σ^2 . Supongamos que disponemos, también, de un estimador S^2 de σ^2 que tiene ν grados de libertad y es independiente de las y_i . Sea R el rango de este conjunto de observaciones,

$$R = \max(y_i) - \min(y_i) \quad . \quad (3.12)$$

Bajo estas condiciones, se define el rango estudentizado como el cociente

$$\frac{\max(y_i) - \min(y_i)}{S} = \frac{R}{S} \quad , \quad (3.13)$$

que se denota por $q_{k,\nu}$. La distribución de este estadístico, que depende de los parámetros k y ν , ha sido tabulada y los valores críticos se presentan en el Apéndice C en la Tabla VI.

Método de Tukey (Honestly-significant-difference)

Este procedimiento se va a desarrollar considerando en primer lugar el caso del modelo unifactorial equilibrado. En este modelo vamos a construir intervalos de confianza con coeficiente de confianza conjunto $1 - \alpha$ para todas las posibles comparaciones por parejas asociadas a los I niveles, es decir las $\binom{I}{2}$ comparaciones por parejas.

El nivel de confianza conjunto $1 - \alpha$ indica que de cada 100 muestras en $(1 - \alpha) \times 100$ de ellas, cada uno de los intervalos contiene a su correspondiente diferencia de medias. Por tanto el nivel de confianza de cada uno de los intervalos será al menos $1 - \alpha$.

Para construir dichos intervalos consideramos las desviaciones

$$(\bar{y}_1 - \mu_1), (\bar{y}_2 - \mu_2), \dots, (\bar{y}_I - \mu_I) \quad . \quad (3.14)$$

Estas desviaciones son variables aleatorias independientes normalmente distribuidas con media 0 y varianza σ^2/n . Además \widehat{S}_R^2/n es un estimador de σ^2/n que es independiente de dichas desviaciones. Entonces, de la definición de rango estudentizado, se sigue que

$$\frac{\max(\bar{y}_i. - \mu_i) - \min(\bar{y}_i. - \mu_i)}{\left(\frac{\widehat{S}_R^2}{n}\right)^{1/2}} \rightsquigarrow q_{I, N-I} \quad , \quad (3.15)$$

donde

- $N - I$ es el número de grados de libertad asociado a \widehat{S}_R^2
- $\max(\bar{y}_i. - \mu_i)$ es la desviación mayor
- $\min(\bar{y}_i. - \mu_i)$ es la desviación menor.

Entonces se verifica que

$$\Pr \left[\frac{\max(\bar{y}_i. - \mu_i) - \min(\bar{y}_i. - \mu_i)}{\left(\frac{\widehat{S}_R^2}{n}\right)^{1/2}} \leq q_{\alpha; I, N-I} \right] = 1 - \alpha \quad . \quad (3.16)$$

Consideramos, a continuación, la siguiente desigualdad generada para todas las parejas de medias de tratamientos i y j

$$|(\bar{y}_i. - \mu_i) - (\bar{y}_j. - \mu_j)| \leq \max(\bar{y}_i. - \mu_i) - \min(\bar{y}_j. - \mu_j) \quad . \quad (3.17)$$

Puesto que la desigualdad (3.17) se verifica para todas las parejas μ_i y μ_j , se sigue de (3.16) que

$$\Pr \left[\left| \frac{(\bar{y}_i. - \mu_i) - (\bar{y}_j. - \mu_j)}{\left(\frac{\widehat{S}_R^2}{n}\right)^{1/2}} \right| \leq q_{\alpha; I, N-I} \right] = 1 - \alpha \quad , \quad (3.18)$$

incluye todos los pares $I(I-1)/2$ de comparaciones entre los I niveles del factor.

Reordenando la desigualdad (3.18) obtenemos los límites de confianza, al nivel de confianza conjunto $1 - \alpha$, para todas las $\binom{I}{2}$ comparaciones por parejas de las I medias, dados por

$$(\bar{y}_{i.} - \bar{y}_{j.}) - HSD \leq \mu_i - \mu_j \leq (\bar{y}_{i.} - \bar{y}_{j.}) + HSD \quad , \quad (3.19)$$

donde

$$HSD = q_{\alpha; I, N-I} \sqrt{\frac{\widehat{S}_R^2}{n}} \quad .$$

Si el intervalo (3.19) no contiene el 0 se concluye que las medias μ_i y μ_j difieren significativamente entre sí.

El coeficiente de confianza $1 - \alpha$ indica que de cada 100 muestras en $(1 - \alpha) \times 100$ de ellas se determinará correctamente cuales de las $\binom{I}{2}$ comparaciones por parejas de medias son significativas.

Por tanto, el método de Tukey resuelve el contraste

$$H_0 : \mu_i = \mu_j \quad vs \quad H_1 : \mu_i \neq \mu_j \quad ,$$

de la siguiente manera

$$\text{Si } |\bar{y}_{i.} - \bar{y}_{j.}| \leq HSD \implies \text{Aceptar } H_0$$

$$\text{Si } |\bar{y}_{i.} - \bar{y}_{j.}| > HSD \implies \text{Rechazar } H_0$$

En el caso de que el modelo no sea equilibrado se emplea el ajuste de Tukey-Kramer, que consiste en reemplazar n por la media armónica, n_h , de los tamaños de los grupos, es decir

$$n_h = \frac{2}{\sum_{i=1}^I \frac{1}{n_i}} \quad . \quad (3.20)$$

Para ilustrar el uso de este procedimiento consideremos el ejemplo de referencia. En este caso hay que usar $q_{0,05;5,21} = 4,22$ como base para determinar el valor crítico y, en consecuencia, ningún par de medias se declara significativamente diferente, salvo que $|\bar{y}_{i.} - \bar{y}_{j.}|$ exceda a

$$HSD = (4,22) \sqrt{\frac{\widehat{S}_R^2}{n_h}} = (4,22) \sqrt{\frac{4,67}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad .$$

Con esta prueba se obtienen los mismos resultados que con el procedimiento de Bonferroni.

El segundo procedimiento de comparaciones múltiples considerado en esta sección es el test de rango múltiple de Duncan. Es un procedimiento iterativo en el cual se intenta comprobar la igualdad de medias basándose en las más extremas. Es similar al procedimiento HSD de Tukey excepto que el nivel de significación varía de unas comparaciones a otras.

Test de rango múltiple de Duncan

El contraste de Duncan utiliza, como el HSD de Tukey, la distribución del recorrido estudentizado. Se diferencia de ese test en que su aplicación es secuencial, en el sentido de no utilizar un único valor crítico para todas las diferencias de medias, como el de Tukey, sino un valor crítico que depende del número de medias comprendido entre las dos medias que se comparan, habiendo ordenado previamente las medias en orden creciente.

Consideremos, en primer lugar, el modelo equilibrado y después generalizaremos para el caso no-equilibrado.

Se acepta que no hay diferencia significativa entre la media mayor y la media menor de p medias, \bar{y}_i e \bar{y}_j , si se verifica

$$|\bar{y}_i - \bar{y}_j| \leq R_p, \quad (3.21)$$

y serán consideradas iguales también todas las medias comprendidas entre ellas. En la expresión (3.21), el valor de R_p es

$$R_p = q_{\alpha_p; p, N-I} \sqrt{\frac{\hat{S}_R^2}{n}} \quad p = 2, 3, \dots, I \quad (3.22)$$

donde

- $q_{\alpha_p; p, N-I}$ es el punto crítico del rango estudentizado basado en la comparación de la media mayor y la menor de p medias. Los valores críticos $q_{\alpha_p; p, \nu}$, para $p = 2, 3, \dots, I$, se presentan en la Tabla VII del Apéndice C para los niveles de significación de comparaciones individuales $\alpha = 0,01$ y $\alpha = 0,05$.
- \hat{S}_R^2 es la varianza residual con $N - I$ grados de libertad.
- α_p es el nivel de significación conjunto relativo a p medias consecutivas; es decir, es la probabilidad de rechazar erróneamente al menos una de las $p - 1$ comparaciones independientes asociadas a las medias consideradas. Dicho nivel de significación está relacionado con el nivel de significación α , de una comparación individual, a través de la ecuación

$$\alpha_p = 1 - (1 - \alpha)^{p-1}. \quad (3.23)$$

Para la aplicación del test de rango múltiple de Duncan, una vez que las medias estén en orden ascendente, se calculan las diferencias entre las medias, comenzando por el valor más pequeño frente al más alto de las $p = I$ medias de los tratamientos, comparando esta diferencia con el valor R_I en la ecuación (3.22) con un nivel de significación α_I . Si esas dos medias no se consideran significativamente diferentes, entonces el contraste se termina y se concluye que ninguna de las medias son significativamente diferentes entre sí al nivel de significación α_I . Esto es equivalente a no rechazar $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$. Si las dos medias extremas son significativamente diferentes, el contraste continúa.

En el siguiente paso se calcula la diferencia entre el valor más pequeño y el segundo valor más grande y esta diferencia se compara con R_{I-1} . Si este contraste no es estadísticamente significativo, la prueba cesa en esta comparación y sólo las dos medias extremas se consideran significativamente diferentes. Si este contraste es estadísticamente significativo, la prueba continúa hasta encontrar la primera pareja de medias que no sea significativamente distinta. A continuación, se calcula la diferencia entre la segunda media más pequeña y la más grande y se compara con R_{I-1} . Este proceso continúa hasta que se han considerado las diferencias entre todas las $I(I-1)/2$ posibles parejas.

Para modelos no-equilibrados, la expresión de R_p es

$$R_p = q_{\alpha_p; p, N-I} \sqrt{\frac{\hat{S}_R^2}{n_h}} \quad p = 2, 3, \dots, I \quad , \quad (3.24)$$

donde n_h es la media armónica dada por la expresión (3.20).

Comentario 3.1

La probabilidad de rechazar erróneamente al menos una hipótesis nula es decir, la probabilidad de detectar incorrectamente como significativa la diferencia entre dos medias de un grupo de tamaño p , es el nivel de significación conjunto α_p relacionado con el nivel de significación α por medio de la expresión (3.23).

Por ejemplo, si $\alpha = 0,05$, entonces

- *Si comparamos parejas de medias adyacentes, el nivel de significación conjunto es*

$$\alpha_2 = 1 - (1 - 0,05)^{2-1} = 0,05 \quad ,$$

que lógicamente coincide con el nivel de significación α .

- *Si comparamos parejas de medias separadas por una media, el nivel de significación conjunto es*

$$\alpha_3 = 1 - (1 - 0,05)^{3-1} = 0,10 \quad .$$

- Si comparamos parejas de medias separadas por dos medias, el nivel de significación conjunto es

$$\alpha_4 = 1 - (1 - 0,05)^{4-1} = 0,142 \quad .$$

Observamos que al aumentar p , aumenta el nivel de significación conjunto.

Para ilustrar la prueba de intervalos múltiples de Duncan consideramos el ejemplo de referencia. Recuerdese que $\hat{S}_R^2 = 4,67$ con 21 grados de libertad. Ordenando las medias de los tratamientos en orden ascendente, se tiene

$$\bar{y}_5 = 45 \quad \bar{y}_4 = 47 \quad \bar{y}_3 = 48 \quad \bar{y}_1 = 50 \quad \bar{y}_2 = 57 \quad .$$

Usando la Tabla de amplitudes significativas de Duncan, Tabla VII del Apéndice C, para 21 grados de libertad y $\alpha = 0,05$, se obtienen los puntos críticos del rango estudentizado:

$$\begin{aligned} q_2 &= q(0,05; 2, 21) = 2,94 & q_3 &= q(0,05; 3, 21) = 3,09 \\ q_4 &= q(0,05; 4, 21) = 3,17 & q_5 &= q(0,05; 5, 21) = 3,24 \quad . \end{aligned}$$

Al hacer las comparaciones se obtiene:

$$5 \text{ vs } 2 \quad q_5 \sqrt{2,33 \left(\frac{1}{6} + \frac{1}{5} \right)} = 2,994, \quad |\bar{y}_{5.} - \bar{y}_{2.}| = 12 > 2,994(*)$$

$$5 \text{ vs } 1 \quad q_4 \sqrt{2,33 \left(\frac{1}{6} + \frac{1}{6} \right)} = 2,793, \quad |\bar{y}_{5.} - \bar{y}_{1.}| = 5 > 2,793(*)$$

$$5 \text{ vs } 3 \quad q_3 \sqrt{2,33 \left(\frac{1}{6} + \frac{1}{5} \right)} = 2,856, \quad |\bar{y}_{5.} - \bar{y}_{3.}| = 3 > 2,856(*)$$

$$5 \text{ vs } 4 \quad q_2 \sqrt{2,33 \left(\frac{1}{6} + \frac{1}{4} \right)} = 2,896, \quad |\bar{y}_{5.} - \bar{y}_{4.}| = 2 < 2,896$$

$$4 \text{ vs } 2 \quad q_4 \sqrt{2,33 \left(\frac{1}{4} + \frac{1}{5} \right)} = 3,245, \quad |\bar{y}_{4.} - \bar{y}_{2.}| = 10 > 3,245(*)$$

$$4 \text{ vs } 1 \quad q_3 \sqrt{2,33 \left(\frac{1}{4} + \frac{1}{6} \right)} = 3,044, \quad |\bar{y}_{4.} - \bar{y}_{1.}| = 3 < 3,044$$

$$4 \text{ vs } 3 \quad q_2 \sqrt{2,33 \left(\frac{1}{4} + \frac{1}{5} \right)} = 3,010, \quad |\bar{y}_{4.} - \bar{y}_{3.}| = 1 < 3,010$$

$$3 \text{ vs } 2 \quad q_3 \sqrt{2,33 \left(\frac{1}{5} + \frac{1}{5} \right)} = 2,983, \quad |\bar{y}_{3.} - \bar{y}_{2.}| = 9 > 2,983(*)$$

$$3 \text{ vs } 1 \quad q_2 \sqrt{2,33 \left(\frac{1}{5} + \frac{1}{6} \right)} = 2,717, \quad |\bar{y}_{3.} - \bar{y}_{1.}| = 2 < 2,717$$

donde * denota que la correspondiente diferencia de medias es significativa.

Test de Newman-Keuls

Este contraste fué desarrollado por Newman en 1939 y ampliado por Keuls en 1952, se suele denominar contraste de Newman-Keuls. Al igual que el contraste de Duncan, es un procedimiento iterativo y, desde el punto de vista operacional, es similar a dicho método.

En el procedimiento de Newman-Keuls, los valores con los que se comparan las sucesivas diferencias de medias vienen dados por

$$K_p = q_{\alpha;p,N-I} \sqrt{\frac{\hat{S}_R^2}{n_h}} \quad p = 2, 3, \dots, I, \quad (3.25)$$

donde

- * $q_{\alpha;p,N-I}$ es el punto crítico del rango estudentizado, definido en (3.13). En la Tabla VI del Apéndice C se recogen los valores críticos de dicha distribución.
- * \hat{S}_R^2 es la varianza residual con $N - I$ grados de libertad.
- * n_h es la media armónica, dada en la ecuación (3.20), que se utiliza cuando los tamaños de las muestras son desiguales. En el caso del modelo equilibrado el valor de la media armónica n_h coincide con el tamaño n de las muestras.

En este método las medias más extremas de un grupo de p medias, incluidas ambas, se comparan con K_p exactamente de la misma forma que en la prueba de intervalos múltiples de Duncan.

Comentarios 3.2

- C1)** *El contraste de Newman-Keuls es más conservador que el de Duncan en el sentido de que el error de tipo I es menor.*
- C2)** *En el contraste de Newman-Keuls el nivel de significación es α , en cambio en el contraste de Duncan es α_p , cuyo valor cambia dependiendo del número de medias comprendidas entre las que se comparan. Por lo tanto, la potencia de la prueba de Newman-Keuls es menor que la del procedimiento de Duncan porque generalmente α es menor que α_p . Podemos comparar los valores de las Tablas VI y VII del Apéndice C para comprobar que el procedimiento de Newman-Keuls conduce a una prueba menos eficaz que la de los intervalos múltiples de Duncan. Se observa que para $p > 2$, siempre*

$$q_{\alpha;p,N-I} > q_{\alpha_p;p,N-I} \quad . \quad (3.26)$$

Así, por ejemplo, considerando los grados de libertad correspondientes al ejemplo de referencia y tomando 0.05 como nivel de significación α , tenemos la siguiente tabla

p	2	3	4	5
$q_{\alpha;p,\nu}$	2.94	3.09	3.17	3.24
$q_{\alpha;p,\nu}$	2.94	3.57	3.94	4.22

donde se comprueba que, efectivamente, para $p > 2$ se verifica la desigualdad dada en (3.26).

En otras palabras, es “más difícil” declarar que dos medias son significativamente diferentes al utilizar la prueba de Newman-Keuls que cuando se usa el procedimiento de Duncan.

Para ilustrar este procedimiento se utiliza el ejemplo de referencia y se comprueba que los resultados coinciden con los obtenidos por los métodos de Bonferroni y Tukey.

Hasta ahora hemos tratado únicamente comparaciones por parejas de niveles de un factor; es decir, hemos considerado hipótesis del tipo

$$\begin{aligned} H_0 : & \mu_j = \mu_k \\ H_1 : & \mu_j \neq \mu_k \quad . \end{aligned}$$

De forma equivalente, estas hipótesis se pueden reescribir como

$$\begin{aligned} H_0 : & \mu_j - \mu_k = 0 \\ H_1 : & \mu_j - \mu_k \neq 0 \quad , \end{aligned}$$

que son casos particulares de hipótesis de la forma

$$\begin{aligned} H_0 : & \sum_{i=1}^I a_i \mu_i = 0 \\ H_1 : & \sum_{i=1}^I a_i \mu_i \neq 0 \quad , \end{aligned}$$

donde todos los a_i son nulos excepto $a_j = 1$ y $a_k = -1$. Además, se verifica la propiedad de que $\sum_i a_i = 0$. Todo esto nos da pie a introducir el concepto de *contraste*.

3.2.4. Contrastes

Se denomina *contraste* a toda combinación lineal C , de los parámetros del modelo de análisis de la varianza, de la forma

$$C = a_1\mu_1 + a_2\mu_2 + \cdots + a_I\mu_I = \sum_{i=1}^I a_i\mu_i \quad , \quad (3.27)$$

tal que los coeficientes verifiquen

$$\sum_{i=1}^I a_i = 0 \quad . \quad (3.28)$$

Ejemplo de contrastes distintos a las comparaciones por parejas son, entre otros,

$$\mu_1 - \frac{\mu_2 + \mu_3}{2} \quad (3.29)$$

$$\mu_1 + \mu_4 - (\mu_2 + \mu_3) \quad . \quad (3.30)$$

- * Si se acepta que el primer contraste es cero se afirma que la media del nivel 1 del factor es igual al promedio de las medias de los niveles 2 y 3.
- * Si se acepta que el segundo contraste es cero se afirma que las medias de los niveles 1 y 4, consideradas en grupo, son iguales a las medias de los niveles 2 y 3, también consideradas en grupo.

La combinación lineal de la forma (3.27) también se puede expresar en función de los efectos del factor

$$C = \sum_{i=1}^I a_i \mu_i = \sum_{i=1}^I a_i (\mu + \tau_i) = \mu \sum_{i=1}^I a_i + \sum_{i=1}^I a_i \tau_i = \sum_{i=1}^I a_i \tau_i \quad , \quad (3.31)$$

ya que $\sum_{i=1}^I a_i = 0$.

Dado que un contraste es una función de los parámetros del modelo, que son desconocidos, su valor se podrá estimar utilizando los estimadores de los parámetros que intervienen. Concretamente, se puede demostrar que el estimador óptimo de un contraste viene dado por la misma combinación lineal de los estimadores de los parámetros que intervienen; es decir

$$\hat{C} = \sum_{i=1}^I a_i \hat{\mu}_i = \sum_{i=1}^I a_i \bar{y}_i \quad ,$$

que también se puede escribir como la misma combinación lineal de los estimadores de los τ_i ; es decir,

$$\hat{C} = \sum_{i=1}^I a_i \hat{\tau}_i \quad . \quad (3.32)$$

En efecto

$$\hat{C} = \sum_{i=1}^I a_i \hat{\mu}_i = \sum_{i=1}^I a_i (\hat{\mu}_i - \hat{\mu} + \hat{\mu}) = \sum_{i=1}^I a_i (\hat{\mu}_i - \hat{\mu}) + \sum_{i=1}^I a_i \hat{\mu} = \sum_{i=1}^I a_i \hat{\tau}_i \quad . \quad (3.33)$$

Si consideramos el contraste (3.29), la combinación lineal $\bar{y}_1 - (\bar{y}_2 + \bar{y}_3)/2$, es el estimador de ese contraste. Para evitar coeficientes fraccionarios es común reescribir esa combinación como

$$2\bar{y}_1 - (\bar{y}_2 + \bar{y}_3) \quad . \quad (3.34)$$

De la misma forma, un estimador del contraste dado en (3.30) es

$$(\bar{y}_1 + \bar{y}_4) - (\bar{y}_2 + \bar{y}_3) \quad . \quad (3.35)$$

Como disponemos de un estimador para cada contraste y conocemos la distribución en el muestreo de los estimadores de los parámetros del modelo, podemos conocer también la distribución del estimador de un contraste y realizar inferencia estadística sobre el mismo. Para ello, vamos a mostrar algunas propiedades de dichos estimadores.

Propiedades de los estimadores de los contrastes

1ª) \hat{C} es un estimador insesgado de C . En efecto

$$E\{\hat{C}\} = E\left[\sum_{i=1}^I a_i \hat{\mu}_i\right] = E\left[\sum_{i=1}^I a_i \bar{y}_{i.}\right] = \sum_{i=1}^I a_i E[\bar{y}_{i.}] = \sum_{i=1}^I a_i \mu_i \quad , \quad (3.36)$$

puesto que $\bar{y}_{i.}$ son variables aleatorias con media μ_i .

2ª) La varianza de \hat{C} es $\sigma^2 \sum_i \frac{a_i^2}{n_i}$. En efecto

$$\sigma^2\{\hat{C}\} = \text{Var}\left\{\sum_{i=1}^I a_i \bar{y}_{i.}\right\} = \sum_{i=1}^I a_i^2 \text{Var}\{\bar{y}_{i.}\} = \sum_{i=1}^I a_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_{i=1}^I \frac{a_i^2}{n_i} \quad (3.37)$$

puesto que $\bar{y}_{i.}$ son variables aleatorias independientes con varianza σ^2/n_i .

Por tanto, un estimador de la varianza de \hat{C} es

$$S^2\{\hat{C}\} = \hat{S}_R^2 \sum_{i=1}^I \frac{a_i^2}{n_i} \quad , \quad (3.38)$$

donde \hat{S}_R^2 , es la varianza residual.

3^a) \widehat{C} se distribuye según una Normal al ser una combinación lineal de variables aleatorias Normales independientes. Además, por las propiedades 1^a y 2^a, su media es C y su desviación típica es $\sigma\{\widehat{C}\}$. En otras palabras,

$$\widehat{C} \rightsquigarrow N(C, \sigma\{\widehat{C}\})$$

Por tanto, al considerar la estimación de $\sigma\{\widehat{C}\}$ en función de \widehat{S}_R^2 , se verifica que el cociente

$$\frac{\widehat{C} - C}{S\{\widehat{C}\}} \rightsquigarrow t_{N-I} \quad . \quad (3.39)$$

Consecuentemente, podemos efectuar la estimación por intervalos de un contraste C al nivel de confianza $1 - \alpha$, que vienen dados por la expresión

$$\widehat{C} \pm t_{\alpha/2; N-I} S\{\widehat{C}\} \quad . \quad (3.40)$$

En el ejemplo de referencia y considerando el contraste dado en (3.29),

$$C = 2\mu_1 - (\mu_2 + \mu_3) \quad ,$$

se tiene que:

1^a) La estimación puntual de C es

$$\widehat{C} = 2\bar{y}_1 - (\bar{y}_2 + \bar{y}_3) = 2(50) - (57 + 48) = -5 \quad .$$

2^a) La estimación de la varianza muestral de \widehat{C} es

$$S^2\{\widehat{C}\} = \widehat{S}_R^2 \sum_{i=1}^I \frac{a_i^2}{n_i} = 4,67(1,067) = 4,982 \quad ,$$

donde $a_1 = 2$, $a_2 = a_3 = -1$ y

$$\sum_{i=1}^I \frac{a_i^2}{n_i} = \frac{4}{6} + \frac{1}{5} + \frac{1}{5} = \frac{16}{15} = 1,067 \quad .$$

Los límites de confianza para C , al nivel confianza del 95 %, son

$$\widehat{C} \pm t_{\alpha/2; N-I} \sqrt{S^2\{\widehat{C}\}} = -5 \pm 2,08(2,23) = -5 \pm 4,63 \quad ,$$

por lo tanto el intervalo de confianza correspondiente es $(-9,67 - 0,37)$.

En los procedimientos de comparaciones múltiples de los efectos del factor interesa manejar únicamente contrastes independientes. Dicha independencia se garantiza al considerar la *ortogonalidad* entre contrastes, que pasamos a definir a continuación.

Contrastes ortogonales

Decimos que dos contrastes³ $C_1 = \sum_i a_i \bar{y}_i$ y $C_2 = \sum_i b_i \bar{y}_i$ son ortogonales si verifican que

$$\sum_{i=1}^I a_i b_i = 0 \quad . \quad (3.41)$$

De la misma forma, se dice que tres o más contrastes son *mutuamente ortogonales* si todas las posibles parejas de contrastes son ortogonales.

Los coeficientes de contrastes ortogonales se pueden elegir de muchas formas para un conjunto dado de tratamientos. Generalmente la naturaleza del experimento sugiere las comparaciones que pueden ser más interesantes.

Para ilustrar los contrastes ortogonales, consideremos el ejemplo de referencia y recordemos que los tamaños de muestras correspondientes a cada uno de los 5 tipos de fertilizantes son: 6, 5, 5, 4, y 6. Supongamos que los contrastes más apropiados son:

- 1) $C_1 = \mu_1 - \mu_3$
- 2) $C_2 = -\mu_1 - \mu_3 + \mu_4 + \mu_5$
- 3) $C_3 = \mu_4 - \mu_5$
- 4) $C_4 = -\mu_1 + 4\mu_2 - \mu_3 - \mu_4 - \mu_5 \quad .$

Los coeficientes en estas comparaciones expresadas en forma tabular serían

Tabla 2-1.

	Respuesta media del nivel del factor				
Contraste	μ_1	μ_2	μ_3	μ_4	μ_5
C_1	1	0	-1	0	0
C_2	-1	0	-1	1	1
C_3	0	0	0	1	-1
C_4	-1	4	-1	-1	-1

Obsérvese que la suma de los coeficientes de cada fila es cero indicando que cada C_i es un contraste. Además, es inmediato comprobar que los productos dos a dos de los

³Aunque $\hat{C} = \sum_i a_i \bar{y}_i$ es un estimador del contraste $C = \sum_i a_i \mu_i$, nos referiremos a él como contraste y lo denotaremos por C .

coeficientes de los contrastes suman cero indicando que los contrastes son mutuamente ortogonales. En efecto, comprobemos, por ejemplo, que los contrastes C_1 y C_2 son ortogonales

$$(1)(-1) + (0)(0) + (-1)(-1) + (0)(1) + (0)(1) = 0 \quad .$$

Veamos a continuación las expresiones para las sumas de cuadrados asociadas a un contraste cualquiera, tanto en el caso equilibrado como en el caso no-equilibrado.

En el diseño no-equilibrado se puede demostrar que al contraste $C = \sum_i a_i \mu_i$ se le puede asociar una suma de cuadrados dada por

$$SSC = \frac{\left(\sum_{i=1}^I a_i \bar{y}_i \right)^2}{\sum_{i=1}^I \frac{a_i^2}{n_i}} \quad , \quad (3.42)$$

y por lo tanto, la expresión de la suma de cuadrados en el modelo equilibrado es

$$SSC = \frac{n \left(\sum_{i=1}^I a_i \bar{y}_i \right)^2}{\sum_{i=1}^I a_i^2} \quad , \quad (3.43)$$

teniendo estas sumas de cuadrados un grado de libertad.

Por tanto, para comprobar las hipótesis del contraste

$$\begin{aligned} H_0 : \quad & \sum_{i=1}^I a_i \mu_i = 0 \\ H_1 : \quad & \sum_{i=1}^I a_i \mu_i \neq 0 \quad , \end{aligned} \quad (3.44)$$

basta comparar la suma de cuadrados asociada al contraste con el cuadrado medio del error. Entonces, el cociente SSC/\hat{S}_R^2 es el estadístico de contraste que, bajo la hipótesis nula, sigue una distribución $F_{1, N-I}$.

A continuación vamos a realizar las sumas de cuadrados asociadas a los contrastes para el ejemplo de referencia. Supongamos que efectuamos los contrastes ortogonales expuestos

en la Tabla 2-1. Dichos contrastes junto con sus estimaciones puntuales y su suma de cuadrados asociada se muestran en la siguiente tabla

Tabla 2-2

Hipótesis	Estimaciones	S.Cuadrados
$H_0: \mu_1 = \mu_3$	$C_1 = 2$	$SSC_1 = 10,89$
$H_0: \mu_1 + \mu_3 = \mu_4 + \mu_5$	$C_2 = -6$	$SSC_2 = 45,97$
$H_0: \mu_4 = \mu_5$	$C_3 = 2$	$SSC_3 = 9,61$
$H_0: 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	$C_4 = 38$	$SSC_4 = 362,35$

Donde

$$C_1 = (1)(\bar{y}_{1.}) + (-1)(\bar{y}_{3.}) = 50 - 48 = 2$$

$$C_2 = (-1)(\bar{y}_{1.} + \bar{y}_{3.}) + (1)(\bar{y}_{4.} + \bar{y}_{5.}) = -(50 + 48) + (47 + 45) = -6$$

$$C_3 = (1)(\bar{y}_{4.}) + (-1)(\bar{y}_{5.}) = 47 - 45 = 2$$

$$C_4 = (-1)(\bar{y}_{1.} + \bar{y}_{3.} + \bar{y}_{4.} + \bar{y}_{5.}) + 4(1)\bar{y}_{2.} = -50 - 48 - 47 - 45 + 4(57) = 38$$

$$SSC_1 = \frac{(2)^2}{\frac{1^2}{6} + \frac{(-1)^2}{5}} = 10,89$$

$$SSC_2 = \frac{(-6)^2}{(-1)^2 \left(\frac{1}{6} + \frac{1}{5} \right) + (1)^2 \left(\frac{1}{4} + \frac{1}{6} \right)} = 45,97$$

$$SSC_3 = \frac{(2)^2}{\frac{1^2}{4} + \frac{(-1)^2}{6}} = 9,61$$

$$SSC_4 = \frac{(38)^2}{(-1)^2 \left(\frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{6} \right) + \frac{(4)^2}{5}} = 362,35$$

Teniendo en cuenta que el valor de la varianza residual es 4.67, con 21 grados de libertad, obtenemos los siguientes valores para los estadísticos de contraste $F(C_1)$, $F(C_2)$, $F(C_3)$ y $F(C_4)$,

$$F(C_1)_{(exp)} = \frac{SSC_1}{\hat{S}_R^2} = \frac{10,89}{4,67} = 2,33$$

$$F(C_2)_{(exp)} = \frac{SSC_2}{\hat{S}_R^2} = \frac{45,97}{4,67} = 9,84$$

$$F(C_3)_{(exp)} = \frac{SSC_3}{\hat{S}_R^2} = \frac{9,61}{4,67} = 2,05$$

$$F(C_4)_{(exp)} = \frac{SSC_4}{\hat{S}_R^2} = \frac{362,35}{4,67} = 77,59 \quad .$$

Si realizamos el contraste al nivel de significación del 5 % y puesto que $F_{0,05;1,21} = 4,32$, se concluye que $\mu_1 = \mu_3$, $\mu_4 = \mu_5$ y en los otros dos contrastes se rechaza la hipótesis nula.

Generalmente las sumas de cuadrados de los contrastes y las pruebas correspondientes se incorporan a la Tabla ANOVA, de la siguiente forma:

Tabla 2-3.

F. V.	S.C.	G.L.	M. C.	F_{exp}	F_α
Entre grupos	439.88	4	109.97	23.56	2.84
$C_1 : \mu_1 = \mu_3$	10.89	1	10.89	2.33	4.32
$C_2 : \mu_1 + \mu_3 = \mu_4 + \mu_5$	45.97	1	45.97	9.84	4.32
$C_3 : \mu_4 = \mu_5$	9.61	1	9.61	2.05	4.32
$C_4 : 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	362.35	1	362.35	77.59	4.32
Interna	98.00	21	4.67		
TOTAL	537.88	25			

Comentarios sobre el nivel de significación

Cuando se realizan comparaciones múltiples una cuestión importante es la elección del nivel de significación de los tests o el nivel de confianza para la estimación por intervalos. Puesto que varios tests o varias estimaciones por intervalos se hacen sobre el mismo conjunto de datos, es necesario distinguir entre dos probabilidades de error. Describimos esas *tasas de error de Tipo I* para contrastes de hipótesis, pero igualmente se haría para intervalos de confianza.

- *Tasa de error individual o para una pareja*, α , es la probabilidad de rechazar erróneamente la hipótesis nula de una única comparación entre una pareja de efectos

o medias. Esencialmente indica el nivel de significación asociado a un solo test estadístico de la forma,

$$\begin{aligned} H_0 : \mu_i &= \mu_j \\ H_1 : \mu_i &\neq \mu_j \end{aligned} \quad (3.45)$$

- *Tasa de error global o del experimento*, α^* , es la probabilidad de rechazar erróneamente por lo menos una hipótesis nula al realizar las $I - 1$ comparaciones independientes relativas a I medias sobre el mismo conjunto de datos. Esta tasa de error puede ser mucho mayor que la tasa de error individual.

Ambas tasas de error están relacionadas por las siguientes expresiones

$$\alpha^* = 1 - (1 - \alpha)^{I-1} \quad , \quad (3.46)$$

$$\alpha = 1 - (1 - \alpha^*)^{1/(I-1)} \quad . \quad (3.47)$$

Estas fórmulas no son estrictamente válidas cuando hay implicados contrastes no-ortogonales. En estas situaciones, sin embargo, pueden utilizarse dichas ecuaciones obteniéndose resultados aproximados. En tales casos, la tasa global puede ser mayor que la indicada por la ecuación (3.46).

Para la ilustración numérica, tomemos como nivel de significación individual 0.05 y como valor de I , 5. Si las 5 medias poblacionales son iguales, la probabilidad de rechazar incorrectamente una o más de las 4 comparaciones ortogonales es $\alpha^* = 1 - (1 - 0,05)^4 = 0,1854$. Es decir, la tasa de error global es casi 4 veces mayor que la tasa individual. Por otra parte, si tomamos α^* igual a 0.05, entonces $\alpha = 1 - (1 - 0,05)^{0,25} = 0,01274$. Observamos que una tasa de error del experimento igual a 0.05 es mucho más exigente que una tasa de error individual de 0.05.

Aunque el experimentador se interesa en controlar todas las tasas de error y niveles de confianza, es importante subrayar que el objetivo principal de los procedimientos de comparaciones múltiples es informarse lo más posible sobre las poblaciones, procesos o fenómenos que intervienen en el experimento. Fijar unas tasas de error pequeñas no es el único propósito de un experimento, es solamente una medida del grado de exigencia o rigor de la metodología estadística empleada.

En muchos experimentos el investigador no sabe a priori los contrastes que le interesa realizar, siendo después de un análisis preliminar de los datos cuando se descubren las comparaciones de interés.

3.2.5. Método de Scheffé para comparaciones múltiples

Scheffé (1953) propuso un método para realizar cualquier contraste entre medias de tratamientos. Dicho procedimiento no requiere que el modelo sea equilibrado.

Sea una familia de contrastes de la forma

$$C = \sum_i a_i \mu_i \quad , \quad (3.48)$$

el objetivo de este procedimiento es decidir, para cada uno de estos contrastes, entre las hipótesis

$$\begin{aligned} H_0 : C &= 0 \\ H_1 : C &\neq 0 \quad . \end{aligned} \quad (3.49)$$

El método de Scheffé está basado en la construcción de intervalos de confianza para todos los posibles contrastes de la forma (3.49). Estos intervalos tienen un nivel de confianza simultáneo $1 - \alpha$, es decir, la probabilidad de que todos los intervalos sean correctos simultáneamente es igual a $1 - \alpha$. Scheffé demostró que dichos intervalos de confianza tienen la siguiente expresión

$$\hat{C} \pm S\{\hat{C}\} \sqrt{(I-1)F_{\alpha; I-1, N-I}} \quad , \quad (3.50)$$

donde las cantidades que intervienen son:

- \hat{C} es el estimador insesgado de C .
- $S\{\hat{C}\}$ es el estimador de la desviación típica del contraste

$$S\{\hat{C}\} = \sqrt{\hat{S}_R^2 \sum_{i=1}^I \frac{a_i^2}{n_i}} \quad , \quad (3.51)$$

siendo \hat{S}_R^2 la varianza residual con $N - I$ grados de libertad.

Por consiguiente, se rechaza la hipótesis H_0 sobre un contraste C si el intervalo de confianza para C

$$\hat{C} - S\{\hat{C}\} \sqrt{(I-1)F_{\alpha; I-1, N-I}} \quad , \quad \hat{C} + S\{\hat{C}\} \sqrt{(I-1)F_{\alpha; I-1, N-I}} \quad (3.52)$$

no incluye al cero, es decir, si

$$|\hat{C}| > S\{\hat{C}\} \sqrt{(I-1)F_{\alpha; I-1, N-I}} \quad . \quad (3.53)$$

Para ilustrar el procedimiento, consideramos el ejemplo de referencia y suponemos que los contrastes de interés son

$$C_2 = \mu_1 + \mu_3 - \mu_4 - \mu_5$$

y

$$C_3 = \mu_4 - \mu_5$$

Las estimaciones numéricas de estos contrastes son

$$\hat{C}_2 = \bar{y}_{1.} + \bar{y}_{3.} - \bar{y}_{4.} - \bar{y}_{5.} = 50 + 48 - 47 - 45 = 6$$

$$\hat{C}_3 = \bar{y}_{4.} - \bar{y}_{5.} = 47 - 45 = 2$$

y sus errores típicos, determinados usando la ecuación (3.51), son

$$S\{\hat{C}_2\} = \sqrt{\hat{S}_R^2 \sum_{i=1}^5 \frac{a_i^2}{n_i}} = \sqrt{4,67 \left(\frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{6} \right)} = 1,912$$

y

$$S\{\hat{C}_3\} = \sqrt{\hat{S}_R^2 \sum_{i=1}^5 \frac{a_i^2}{n_i}} = \sqrt{4,67 \left(\frac{1}{4} + \frac{1}{6} \right)} = 1,394 \quad .$$

Por lo tanto, los intervalos de confianza son, respectivamente

$$I_2 = \left(\hat{C}_2 - S\{\hat{C}_2\} \sqrt{(I-1)F_{0,05;I-1,N-I}}, \hat{C}_2 + S\{\hat{C}_2\} \sqrt{(I-1)F_{0,05;I-1,N-I}} \right) =$$

$$(6 - 1,912\sqrt{4F_{0,05;4,21}}, 6 + 1,912\sqrt{4F_{0,05;4,21}}) = (-0,44, 12,44) \quad .$$

$$I_3 = \left(\hat{C}_3 - S\{\hat{C}_3\} \sqrt{(I-1)F_{0,05;I-1,N-I}}, \hat{C}_3 + S\{\hat{C}_3\} \sqrt{(I-1)F_{0,05;I-1,N-I}} \right) =$$

$$(2 - 1,394\sqrt{4F_{0,05;4,21}}, 2 + 1,394\sqrt{4F_{0,05;4,21}}) = (-2,69, 6,69) \quad .$$

Puesto que los intervalos I_2 e I_3 contienen al 0, se concluye que tanto $C_2 = \mu_1 + \mu_3 - \mu_4 - \mu_5$, como $C_3 = \mu_4 - \mu_5$ no son significativamente distintos de cero. En otras palabras, existe evidencia para concluir, por una parte que las medias de los tratamientos uno y tres, tomadas como grupo, no difieren significativamente de las medias de los tratamientos cuatro y cinco, también tomadas como grupo y por otra parte que no existe diferencia significativa entre los tratamientos cuatro y cinco.

3.3. Comparaciones entre los métodos

En diversos textos se encuentran comentarios sobre las comparaciones entre los distintos métodos; entre otras referencias, se recomienda consultar Neter et al. (1990) y Mason et al. (1989). A continuación, presentamos aquí un resumen de ambos estudios.

- * Cuando sólo se hacen comparaciones por parejas, el método de Tukey conduce a límites de confianza más estrechos que el método de Scheffé, por lo cual el método de Tukey encontrará más diferencias significativas, siendo en este caso el método preferido. En cambio cuando los contrastes son más complicados que la diferencia de medias, es el método de Scheffé el que tiene límites de confianza más estrechos.
- * El método de Bonferroni es preferido al de Scheffé cuando el número de contrastes es igual o menor que el número de niveles del factor.
- * Cuando el número de comparaciones por parejas es muy grande no se debe utilizar el método de Bonferroni, ya que el nivel de significación de cada comparación puede llegar a ser demasiado pequeño para considerarse de utilidad. En estas situaciones los tests de rangos múltiples como Tukey, Newman-Keuls y Duncan ofrecen una solución de compromiso entre la tasa de error global deseada y una tasa de error individual demasiado pequeña y por tanto inaceptable. Estos métodos son preferidos en el sentido de producir intervalos de confianza más estrechos.
- * El método LSD de Fisher es el que proporciona más diferencias significativas; a continuación, le siguen los métodos de Duncan y Tukey. Eligiremos uno u otro dependiendo del riesgo que estemos dispuestos a correr al aceptar más o menos diferencias significativas. Es decir, a aceptar como significativas diferencias que no lo sean (situación no conservadora), o a aceptar menos diferencias significativas de las que realmente existan (situación conservadora). Cramer y Swanson (1973), realizaron estudios de simulación por el método de Montecarlo de los que concluyeron que el procedimiento *LSD* es una prueba muy eficiente para detectar diferencias verdaderas entre las medias si se aplica después que la prueba *F* del análisis de la varianza resultó significativa al 5 %. También concluyeron que el procedimiento de intervalos múltiples de Duncan es un buen método para detectar diferencias reales.
- * El procedimiento de Tukey tiene un error tipo I menor que los correspondientes errores de los tests de Newman-Keuls y de Duncan; es decir, es un test más conservador. En consecuencia, el procedimiento de Tukey tiene menos potencia que los procedimientos de Newman-Keuls o de Duncan.
- * Si se desea controlar la tasa de error individual, los métodos LSD de Fisher y de Duncan resultan apropiados. Facilitan más protección de los errores de Tipo I que

los otros métodos y son menos conservadores que los procedimientos basados en la elección de la tasa de global.

- * Si se desea controlar la tasa de error global, los métodos más útiles incluyen el test de Bonferroni y procedimiento de Tukey. Ambas técnicas tienen fuertes defensores. El test de Bonferroni tiene la ventaja de utilizar un estadístico t , siendo su principal desventaja, cuando se realizan un gran número de comparaciones, que el nivel de significación individual se hace demasiado pequeño.

Existen otros procedimientos para hacer comparaciones múltiples. Algunos de ellos están designados para casos especiales, tales como para comparar tratamientos experimentales con un tratamiento control.

3.4. Comparaciones de tratamientos con un Control

Estamos ante el caso de un experimento en el que uno de los tratamientos es un *control* y el experimentador puede estar interesado en comparar los restantes tratamientos con este *control*. Por tanto, sólo deben realizarse $I - 1$ comparaciones. Un procedimiento que realiza dichas comparaciones fué desarrollado por Dunnett (1964).

Supongamos que el control es el tratamiento I . Entonces se desean probar las hipótesis

$$\begin{aligned} H_0 : & \mu_i = \mu_I \\ H_1 : & \mu_i \neq \mu_I \end{aligned} \quad i = 1, 2, \dots, I - 1 \quad . \quad (3.54)$$

El procedimiento de Dunnett es una modificación de la prueba t . Para cada hipótesis se calculan las diferencias que se observan entre las medias muestrales

$$| \bar{y}_i. - \bar{y}_I. | \quad i = 1, 2, \dots, I - 1 \quad .$$

La pareja de medias μ_i y μ_I se consideran diferentes, a un nivel de significación conjunto α asociado a las $I - 1$ comparaciones, si

$$| \bar{y}_i. - \bar{y}_I. | > d_{\alpha; I-1, N-I} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_i} + \frac{1}{n_I} \right)} \quad , \quad (3.55)$$

en donde la cantidad $d_{\alpha; I-1, N-I}$ se obtiene a partir de la Tabla VIII del Apéndice C.

Para ilustrar la prueba de Dunnett, consideramos el ejemplo de referencia y aunque en el desarrollo teórico hemos dicho que el control es el último tratamiento, la posición que

ocupe el control es arbitraria, por ello, en este ejemplo vamos a considerar que es la primera posición, el tratamiento 1. En dicho ejemplo, $I=5$, $\nu=21$ y para un nivel de significación del 5% el valor de $d_{0,05;4,21}$ es 2.64. Por tanto, las diferencias críticas y observadas son, respectivamente,

$$2 \text{ vs } 1 : d_{0,05;4,21} \sqrt{4,67 \left(\frac{1}{5} + \frac{1}{6} \right)} = 3,454 \quad | \bar{y}_2. - \bar{y}_1. | = 7*$$

$$3 \text{ vs } 1 : d_{0,05;4,21} \sqrt{4,67 \left(\frac{1}{5} + \frac{1}{6} \right)} = 3,454 \quad | \bar{y}_3. - \bar{y}_1. | = 2$$

$$4 \text{ vs } 1 : d_{0,05;4,21} \sqrt{4,67 \left(\frac{1}{4} + \frac{1}{6} \right)} = 3,682 \quad | \bar{y}_4. - \bar{y}_1. | = 3$$

$$5 \text{ vs } 1 : d_{0,05;4,16} \sqrt{4,67 \left(\frac{1}{6} + \frac{1}{6} \right)} = 3,293 \quad | \bar{y}_5. - \bar{y}_1. | = 5*$$

Observamos que las parejas de medias 2 vs 1 y 5 vs 1 resultan significativamente diferentes.

Al comparar tratamientos con un control es conveniente tomar más observaciones para el tratamiento de control que para los otros tratamientos. Como norma general, se debe verificar que la razón n_I/n_i sea aproximadamente igual a la raíz cuadrada del número total de tratamientos. Es decir,

$$\frac{n_I}{n_i} \approx \sqrt{I} \quad .$$

Bibliografía utilizada

- * **García Leal, J. & Lara Porras, A.M.** (1998). *“Diseño Estadístico de Experimentos. Análisis de la Varianza.”* Grupo Editorial Universitario.
- * **Lara Porras, A.M.** (2000). *“Diseño Estadístico de Experimentos, Análisis de la Varianza y Temas Relacionados: Tratamiento Informático mediante SPSS”* Proyecto Sur de Ediciones.