

# 18

## Comparaciones múltiples

Los diferentes métodos de hacer comparaciones múltiples se emplean *sólo* cuando el resultado del ANOVA resulta significativo. En tal caso, se sabe que existen diferencias entre las muestras, pero sin poder especificar entre cuales de ellas. Se necesita, entonces, alguna forma de poder compararlas entre sí, y alcanzar así el objetivo final del ANOVA. Por ejemplo, decidir si un placebo difiere realmente de los demás grupos, y todavía, si entre éstos algunos son diferentes de otros. Los métodos de *comparaciones múltiples* son la herramienta adecuada para este fin, pues permiten un análisis muy sutil de la significación encontrada. Para un *Modelo II* el trabajo es mucho más simple, solo se trata de cuantificar a la componente añadida de varianza. En cambio, para un *Modelo I* se debe analizar primero si las comparaciones fueron *planeadas* antes de realizar el experimento o después. Cuando se analizan los datos y se encuentran resultados sorprendidos, no previstos, a veces conviene un análisis posterior que se diseña en función de estos resultados, en cuyo caso se deben usar los métodos para comparaciones *no planeadas*. Existen numerosos modelos estadísticos para todas estas posibilidades, tanto para tamaños muestrales iguales como distintos en esta parte se presentan aquellos más recomendables por su potencia, sensibilidad y sencillez.

### 18.1 Componente añadida de varianza

Las variaciones detectadas con un Modelo II en el ANOVA se presentan en el cuadro respectivo, como proviniendo de dos fuentes, una debida a la variabilidad entre los grupos y la otra debida al error de medición. Cuando no se detecta significación en el ensayo, ocurre que los valores de los cuadrados medios involucrados son aproximadamente iguales; entonces su cociente será cercano a la unidad, esto es:

$$F = MS_E / MS_D \approx 1 \quad (\text{resultados no significativos})$$

Por otra parte, el valor esperado del estadígrafo F es:

$$E(F) = E(MS_E) / E(MS_D) = (\sigma^2 + n \sigma_A^2) / (\sigma^2)$$

Luego para que este cociente sea aproximadamente la unidad, la componente añadida de varianza  $\sigma^2_A$  debe ser nula. Por lo tanto, se dice que si el Cuadro de ANOVA resulta no significativo en un Modelo II, es porque la componente añadida no influye. En cambio, si el Cuadro resulta ser significativo, entonces la componente añadida no es despreciable, y se está interesado en el valor de su contribución a la variabilidad total, más que en su valor en sí. No interesa su magnitud, sino el porcentaje que tiene en la variabilidad total. El valor de interés se calcula con el cociente porcentual entre el valor de la componente añadida y el total. De la relación anterior resulta entonces:

$$E(MS_E) - E(MS_D) = (\sigma^2 + n \sigma^2_A) - (\sigma^2) = n \sigma^2_A$$

Donde n es el número de mediciones de cada grupo. O sea :

$$\sigma^2_A = [ E(MS_E) - E(MS_D) ] / n$$

Entonces la mejor estimación de la componente añadida de varianza es su valor muestral :

$$\sigma^2_A \approx DS^2_A = [ (MS_E - MS_D) ] / n$$

Pero en los hechos, más que la magnitud absoluta de la componente añadida  $DS^2_A$  interesa obtener el valor porcentual de su contribución al total (P%) y este valor se calcula con:

$$P\% = 100 (\sigma^2_A / \sigma^2) \approx 100 [ (DS^2_A) / (DS^2) ]$$

$$\text{Donde } DS^2 = MS_E + MS_D$$

Por ejemplo, para el Cuadro de ANOVA visto en el capítulo anterior (Paso 8 del ejemplo en el punto 17.3), suponiendo que los grupos se hubiesen escogido al azar, en lugar de cómo se hizo, se trataría de un Modelo II y así el valor de la contribución porcentual debida a la componente añadida resulta ser :

$$MS_E = 0,09317 \quad ; \quad MS_D = 0,00273 \quad ; \quad n = 5 \quad \text{O sea, } F = 34,1^* \quad (\text{resultado significativo})$$

Para el caso del Modelo II se puede obtener :

$$DS^2_A = (1/n) (MS_E - MS_D) = (1/5) (0,09317 - 0,00273) = 0,01809$$

$$DS^2 = MS_E + MS_D = 0,09317 + 0,00273 = 0,0959$$

Entonces resulta:

$$P\% = 100 (0,01809 / 0,0959) = 18,87 \%$$

Así el 18,87% de la variabilidad total del experimento, se debe a la componente añadida de varianza entre los grupos. El factor aleatorio incidente es muy importante. Para este problema, significa que el tipo de técnica empleada para medir el RGR debe ser tenido muy en cuenta para lograr estudiar la concordancia entre los diferentes laboratorios.

## 18.2 Comparaciones múltiples “ a priori ”

Cuando se trata de un modelo I de Anova, lo que más interesa es poder comparar las medias muestrales entre sí, una vez que se sabe que hay diferencia significativa entre ellas. Hay dos formas básicas de poder efectuar estas comparaciones :

*Comparaciones “a priori”*: Son aquellas comparaciones planificadas previamente, durante la etapa del diseño experimental. Es decir, las que el experimentador cree que va a encontrar diferencias significativas, *antes* de hacer el experimento.

*Comparaciones “a posteriori”*: Son aquellas comparaciones no planificadas de antemano. Surgen a partir de los datos experimentales, cuando el investigador descubre diferencias inesperadas y quiere testearlas.

Para explicar mejor el funcionamiento de este tipo de comparaciones se desarrolla un ejemplo de aplicación :

Un laboratorio investiga la composición de un medicamento nuevo, para tratar cierta enfermedad infecciosa. Decide testear 3 drogas A, B y C con la misma composición porcentual en la fórmula y un cóctel hecho con dos de ellas A y B con la mitad del porcentaje para cada una. Para ello, busca 50 pacientes escogidos al azar entre los que padecen la enfermedad con grados similares de avance y pertenecientes a un mismo estrato social y educacional. Lo que se mide es la cantidad de días que tardan en curarse completamente. Para los 5 casos a estudiar, escoge al azar 10 pacientes. Al primer grupo les suministra un placebo, al segundo la droga A, al tercero la B, al cuarto el cóctel (50% de A y 50% de B) y al quinto la droga C. Los resultados obtenidos se muestran en el cuadro siguiente:

**Cuadro 18.1 : Ejemplo de Anova de 1 factor. (Datos del Cuadro 17.6)**

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C	Totales
1	125	107	108	108	112	
2	117	108	111	109	116	
3	120	110	106	108	115	
4	125	109	108	111	113	
5	115	112	107	107	114	
6	121	110	106	106	112	
7	117	110	111	108	115	
8	117	107	110	107	115	
9	126	109	107	107	112	
10	118	111	108	109	117	
Total	1201	1093	1082	1080	1141	5597
Media	120,1	109,3	108,2	108	114,1	
$\Sigma x^2$	144383	119489	117104	116658	130217	627851
$(1/n)(\Sigma x)^2$	144240,1	119464,9	117072,4	116640	130188,1	627605,5

$$(1/N)(\Sigma \Sigma x)^2 = (5597)^2/50 = \boxed{626528,18}$$

Se trata de un diseño experimental de un Modelo I de Anova de un factor. Se calculan los totales por grupo, y la media de cada uno. El gran total es  $T = 5.597$  días. La suma de los cuadrados de cada grupo y el total de la suma de cuadrados  $\sum \sum x^2 = 627.851$ . Además, para cada grupo se calcula el cuadrado de su total dividido su tamaño  $(1/n)(\sum x)^2$ , su suma y, finalmente, el término de corrección  $T^2 / N = (1/n)(\sum \sum x)^2 = (5.597)^2 / 50 = 626.528,18$ . (Ver Cuadro anterior). Con todos estos datos se pueden calcular las sumas de cuadrados así:

$$SS_t = \sum \sum x^2 - T^2 / N = 627.851 - 626.528,18 = 1.322,82$$

$$SS_e = \sum (1/n)(\sum x)^2 - T^2 / N = 627.605,5 - 626.528,18 = 1.077,32$$

$$SS_d = SS_t - SS_e = 1.322,82 - 1.077,32 = 245,5$$

Con estos datos se arma la Tabla de Anova, para detectar si hay diferencias entre las medias muestrales, se compara el estadígrafo F contra los valores críticos de tabla.

**Cuadro 18.2 : Tabla de Anova del Cuadro 18.1.**

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre grupos	1.077,32	4	269,33	49,33***
Dentro grupos	245,5	45	5,46	
Total	1.322,82	49		

NOTA: A las sumas de cuadrados se las divide por sus grados de libertad respectivos y se obtienen los cuadrados medios. El estadígrafo F se obtiene haciendo el cociente de los dos cuadrados medios.

Para los tres niveles usuales:  $F_{0,05(4,45)} = 2,58$  ;  $F_{0,01(4,45)} = 3,77$  ;  $F_{0,001(4,45)} = 5,57$ . Como  $F = 49,33***$  se rechaza la hipótesis nula porque se tiene evidencia altamente significativa de diferencia entre las medias muestrales.

A esta altura de los cálculos se han descubierto diferencias altamente significativas entre los 5 grupos testeados. Pero no se puede establecer entre cuales grupos en una forma individual, que es justamente lo que más interesa. Por lo tanto, se necesitan modelos estadísticos que permitan comparar a los grupos entre sí. Esto es, los *modelos de comparaciones múltiples*. Para el problema que se está estudiando, el diseñador del experimento busca comparar varias cosas: (1) El grupo control contra los grupos a los cuales se le suministró el medicamento. (2) Luego quiere comparar las drogas puras contra el cóctel de dos de ellas y (3) Comparar las drogas entre sí. Esto lo diseñó antes de hacer las mediciones, por lo tanto corresponde usar un modelo “a priori”.

Hay dos formas de hacer estas comparaciones. La primera es hacer todas las comparaciones posibles entre dos grupos, usando el modelo t-Student aplicado al caso de dos muestras independientes, como se vio en el capítulo anterior. La segunda forma es descomponer la suma de cuadrados del tratamiento en comparaciones separadas, usando el modelo F-Fisher como continuación del Anova. Esta última, es la manera adecuada porque es la más sencilla y elegante de ambas. Matemáticamente ambos casos son equivalentes.

La regla general para realizar las comparaciones múltiples “a priori” es sencilla:

*Paso 1)* Para comparar  $k$  grupos de tamaño  $n_i$  se toma la suma del grupo elegido para comenzar las comparaciones  $T_1$ , se lo eleva al cuadrado y se lo divide por su tamaño muestral, resulta entonces:

$$S_1 = T_1^2 / n_1 = (1201)^2 / 10 = 144.240,1 \quad (\text{grupo placebo})$$

*Paso 2)* Se suman los totales de los grupos restantes  $T_r = \sum_2^k T_i = T_2 + T_3 + \dots + T_k$

$$T_r = (1093 + 1082 + 1080 + 1141) = 4396$$

*Paso 3)* Se eleva al cuadrado este último total y se lo divide por el tamaño muestral de todos esos grupos, calculando así el segundo término buscado:

$$S_2 = (T_r)^2 / (n_2 + n_3 + \dots + n_k) = (4396)^2 / 40 = 483.120,4$$

*Paso 4)* Se calcula el término de corrección, sumando todos los grupos componentes de la comparación, elevándola al cuadrado y dividiendo por el tamaño total de las muestras integrantes

$$T = \sum_1^k T_i = T_1 + T_2 + \dots + T_k = 5597 \quad S_3 = T^2 / (n_1 + n_2 + \dots + n_k) = T^2 / N$$

$$S_3 = (5597)^2 / 50 = 626.528,18$$

*Paso 5)* Se calcula la suma de cuadrados para la comparación buscada como:

$$SS (\text{placebo vs. medicamentos}) = S_1 + S_2 - S_3 = 832,32$$

Para este caso hay un grado de libertad, por lo tanto  $MS = SS / v = SS = 832,32$

*Paso 6)* Se calcula el estadígrafo  $F = MS (\text{placebo vs. medicamentos}) / MS \text{ dentro}$

$$F = 832,32 / 5,46 = 152,44$$

*Paso 7)* Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (1,45)} = 4,05 \quad F_{0,01; (1,45)} = 7,23 \quad F_{0,001; (1,45)} = 12,4$$

En conclusión, se rechaza la hipótesis nula con resultados altamente significativos  $F = 152,44^{***}$

Esto significa, que se tiene evidencia científica muy fuerte de las diferencias entre el grupo de control y los demás grupos, se puede concluir que el efecto del medicamento sirve para curar la enfermedad infecciosa (hay validación estadística). Este era el objetivo principal de la investigación, pero ahora, se puede continuar más allá, gracias a las bondades de este modelo. Por ejemplo, se puede investigar si hay diferencias entre el grupo al cual se le suministró el cóctel de droga y los demás que fueron tratados con las drogas puras. Para ello, se repite la misma técnica anterior, pero usándola con los datos de los grupos remanentes.

Paso 1) Se calcula la primera suma de cuadrados con los datos del grupo que recibió el cóctel de drogas:

$$S_1 = T_2^2 / n_2 = (1080)^2 / 10 = 116.640$$

Paso 2) Se suman los totales de los grupos restantes  $T_r = \sum_3^k T_i = T_3 + \dots + T_k$

$$T_r = (1093 + 1082 + 1141) = 3316$$

Paso 3) Se eleva al cuadrado este último total y se lo divide por el tamaño muestral de todos esos grupos, calculando así el segundo término buscado:

$$S_2 = (T_r)^2 / (n_3 + \dots + n_k) = (3316)^2 / 30 = 366.528,53$$

Paso 4) Se calcula el término de corrección, sumando todos los grupos remanentes de la comparación, elevándola al cuadrado y dividiendo por el tamaño total de las muestras integrantes

$$T' = \sum_2^k T_i = T_2 + T_3 + \dots + T_k = 4396 \quad S_3 = T'^2 / (n_2 + \dots + n_k) = T'^2 / N$$

$$S_3 = (4396)^2 / 40 = 483120,4$$

Paso 5) Se calcula la suma de cuadrados para la comparación buscada como:

$$SS (\text{cóctel vs. drogas puras}) = S_1 + S_2 - S_3 = 48,13$$

Para este caso hay un grado de libertad, por lo tanto  $MS = SS / v = SS = 48,13$

Paso 6) Se calcula el estadígrafo  $F = MS (\text{cóctel vs. drogas puras}) / MS \text{ dentro}$

$$F = 48,13 / 5,46 = 8,82$$

Paso 7) Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (1,45)} = 4,059 \quad F_{0,01; (1,45)} = 7,2525 \quad F_{0,001; (1,45)} = 12,45$$

En conclusión, se rechaza la hipótesis nula con resultados muy significativos  $F = 8,82^{**}$

Se tiene evidencia científica muy fuerte de las diferencias entre el grupo al que se le suministró el cóctel y los restantes grupos, se puede concluir que el efecto del medicamento mezclado difiere muy significativamente de los puros. Por último, se puede hacer una comparación más entre los diferentes tipos de drogas. Lo que se busca ahora, es detectar si hay diferencia entre ellas y para esto se repite la técnica anterior, usando los tres grupos remanentes.

Paso 1) Para comparar los 3 grupos remanentes:

$$S_1 = (1082)^2 / 10 = 117.072,4 \quad S_2 = (1093)^2 / 10 = 119.464,9 \quad S_3 = (1141)^2 / 10 = 130.188,1$$

Paso 2) Se calcula el término de corrección, sumando todos los grupos componentes de la comparación, elevándola al cuadrado y dividiendo por el número total de muestras integrantes:

$$T'' = \sum_3^k T_i = T_3 + T_4 + T_5 = 3316 \qquad S_4 = T''^2 / (n_3 + n_4 + n_5)$$

$$S_4 = (3316)^2 / 30 = 366.528,53$$

Paso 3) Se calcula la suma de cuadrados para la comparación buscada como:

$$SS \text{ (entre drogas puras)} = S_1 + S_2 + S_3 - S_4 = 196,87$$

Para este caso hay dos grados de libertad, por lo tanto  $MS = SS / v = SS / 2 = 98,435$

Paso 4) Se calcula el estadígrafo  $F = MS \text{ (entre drogas puras)} / MS \text{ dentro}$

$$F = 98,435 / 5,46 = 18,03$$

Paso 5) Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (2,45)} = 3,204 \qquad F_{0,01; (2,45)} = 5,110 \qquad F_{0,001; (2,45)} = 8,085$$

En conclusión, se rechaza la hipótesis nula con resultados altamente significativos  $F = 18,03^{***}$

Se tiene evidencia científica muy fuerte de las diferencias entre los diferentes tipos de drogas investigados. Todo esto se puede resumir en un Cuadro de Anova similar.

**Cuadro 18.3: Comparaciones múltiples (Datos del ejemplo en Cuadro 18.1).**

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
<b>Entre grupos</b>	<b>1077,32</b>	<b>4</b>	<b>269,33</b>	<b>49,33***</b>
Control vs. Drogas	832,32	1	832,32	152,44***
Cóctel vs. Drogas	48,13	1	48,13	8,82**
Drogas entre sí	196,87	2	98,44	18,03***
<b>Dentro grupos</b>	<b>245,5</b>	<b>45</b>	<b>5,46</b>	
<b>Total</b>	<b>1322,82</b>	<b>49</b>		

Observando la tabla anterior puede notarse, que la suma de cuadrados entre grupos (1077,32) se descompuso en tres términos, al igual que los grados de libertad respectivos. Lo que permite efectuar tres ensayos de hipótesis extras y sacar las conclusiones correspondientes. No se puede descomponer en más de cuatro comparaciones porque no alcanzarían los grados de libertad, para poder hacer una comparación múltiple de tipo *ortogonal*. Cuando se quieran hacer más comparaciones, hay que emplear modelos *no ortogonales* como el de Dunn-Sidak que el lector interesado puede encontrar en la bibliografía como Biometry de Sokal-Rohlf (Edición 1981).

## 18.3 Comparaciones múltiples “ a posteriori ”

Cuando se desean hacer comparaciones sobre la base de los resultados obtenidos, es decir, cuando las comparaciones no fueron planificadas de antemano por el experimentador, hay que usar los modelos “a posteriori”. En estos modelos se deben distinguir los dos casos posibles:

*Modelos “a posteriori” con tamaños muestrales iguales:*

- Modelo de Tukey (*T-method*).
- Modelo de Tukey corregido (*T'-method*) cuando las muestras son aproximadamente iguales.
- Modelo de Welsch (*Welsch-method*).
- Modelo de Dunn-Sidák.

*Modelos “a posteriori” con tamaños muestrales distintos:*

- Modelo de Hochberg (*GT2-method*).
- Modelo de Tukey – Kramer (*TK-method*).
- Modelo de Student – Neumann – Keuls (*SNK-method*).
- Modelo de Scheffé.
- Modelo de Gabriel (*SS-STP method*) para hacer todas las comparaciones posibles.

Todos estos métodos son llamados también modelos *no ortogonales* y pueden encontrarse en la bibliografía. Las diferencias entre uno y otro se basan en la potencia para discriminar diferencias significativas entre dos medias muestrales. Tienen ventajas y desventajas que los investigadores en Estadística siguen discutiendo, en un campo abierto hasta hoy. Para ser consecuente con el objetivo de esta obra, que busca la sencillez a fin de facilitar la comprensión del lector y la simplicidad de cálculos, se ha decidido desarrollar a dos de ellos. Uno para el caso de tamaños iguales de muestras (Tukey) y el otro para tamaños distintos. Pero en ambos, se prefiere mostrar las comparaciones gráficas sugeridas por Gabriel (1978 y 1980) en lugar de los tediosos cálculos para comparar en una tabla todos los casos posibles.

La idea básica comienza en el modelo de Student para comparar dos medias muestrales independientes:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{DS_1^2 / n_1 + DS_2^2 / n_2}} \quad \text{versus} \quad t_{\alpha; n-1}$$

Como se hace el supuesto de homoscedasticidad, resulta  $DS_1^2 = DS_2^2 = DS^2$

Si los tamaños de muestras son iguales entonces:

$$\sqrt{DS_1^2 / n_1 + DS_2^2 / n_2} = \sqrt{DS^2 / n + DS^2 / n} = DS \cdot \sqrt{2 / n}$$

Pero si los tamaños muestrales son diferentes, como el cuadrado medio intra-grupo  $MS_d$ , resulta ser una especie de promedio entre todas las varianzas muestrales, se puede obtener la expresión:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{2 \cdot MS_d / n}} \quad \text{versus} \quad t_{\alpha; n-a}$$

Si se supone que los valores esperados son iguales ( $\mu_1 = \mu_2$ ) y en el caso límite, cuando el valor del estadígrafo  $t$  es igual al de tablas la comparación anterior queda:

$$(\bar{x}_1 - \bar{x}_2)_{\text{límite}} = t_{\alpha; n-a} \sqrt{2 \cdot MS_d / n} = \text{LSD} \quad (\text{para un ensayo de dos colas})$$

Donde LSD es una especie de “*mínima diferencia significativa*”. Es un valor único, que una vez calculado, le permite al investigador revisar cualquier pareja de medias, para decidir si difieren significativamente entre sí. Por ejemplo, para el problema presentado en el Cuadro 18.1:

$$\text{LSD} = t_{0,05;45} \cdot \sqrt{2 \cdot (5,46) / 10} = 2,017 \cdot 1,045 = 2,107$$

Por lo tanto, cualquier pareja de medias que difieran en 2,107 son significativamente diferentes una de otra, con probabilidad  $p < 0,05$ . Aplicando estos conceptos al problema, se vuelven a comprobar las diferencias testeadas (usando el promedio entre las drogas puras).

Sin embargo, debe recalarse que estas comparaciones *son válidas, sólo si fueron planificadas de antemano*. Para poder incluir las comparaciones no planificadas, se puede buscar una fórmula equivalente a la anterior, tal como:

$$\text{MDS} = (\text{valor crítico de tablas}) \cdot (\text{error típico})$$

Donde MDS es la *mínima diferencia significativa*, usando una distribución estadística apropiada para el caso a decidir con un test y de donde obtener un valor crítico. Con un error típico de referencia para el estadígrafo elegido. Los diferentes modelos mencionados más arriba, buscan optimizar la distribución siguiendo varias líneas de pensamiento. Pero, el criterio más difundido es emplear el “*test del rango studentizado*” cuya tabla de valores críticos puede verse en la Tabla 18 del Anexo. Entonces, para esa distribución estadística queda:

$$\text{MDS} = Q_{\alpha (a; N-a)} \sqrt{MS_d / n}$$

Donde la tabla de rangos studentizados posee dos argumentos:  $a$  = número de grupos analizados y el número de grados de libertad ( $n-a$ ). Para un nivel de significación como  $\alpha = 0,05$  elegido, se puede calcular el valor del MDS para el ejemplo de la investigación de la Tabla 18.1:

$$\text{MDS} = Q_{0,05 (5 ; 45)} \sqrt{5,46 / 10} = 4,02 \sqrt{0,546} = 2,969$$

Para hacer las comparaciones “*a posteriori*” se mostrará el método de Tukey con la sugerencia gráfica desarrollada por Gabriel en el Cuadro 18.4 siguiente. Para no tener que hacer todos los pares de combinaciones posibles en una tabla engorrosa, Gabriel sugirió calcular los límites superior e inferior de cada media muestral estudiada, determinando así un intervalo de confianza dentro del cual, el verdadero valor puede estar en cualquier parte, con una confianza obtenida a partir del nivel de significación  $\alpha$  elegido por el investigador. Luego se grafican estos intervalos

y cuando no se superpongan entre sí, la diferencia entre las medias será significativa. De esta forma, con un simple vistazo, se pueden detectar las diferencias buscadas.

**Cuadro 18.4 : Comparaciones no planificadas para tamaños muestrales iguales.**

Usando los valores del problema de la Tabla 18.1 se procede como sigue:

*Paso 1 :* Se calcula el valor de la mínima diferencia significativa MSD con :

$$MDS = Q_{\alpha(a; N-a)} \sqrt{MS_d / n} \quad \text{En este caso es:}$$

$$MDS = Q_{0,05(5; 45)} \sqrt{0,546} \quad \text{Interpolando entre 40 y 60 para los grados de libertad resulta:}$$

$$MDS = 2,9691$$

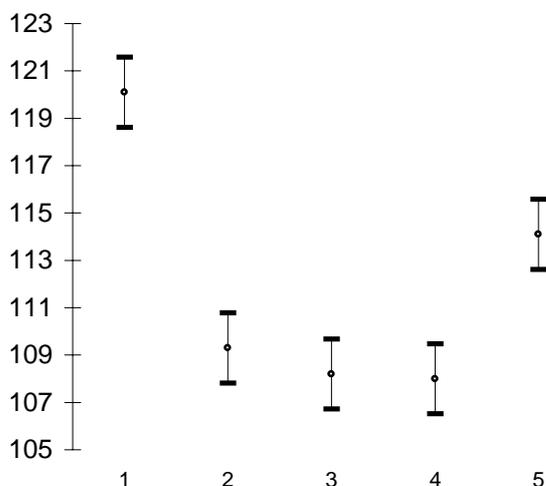
*Paso 2 :* Se calculan los límites superior e inferior para cada media muestral, sumando y restando la mitad de MSD a cada media muestral respectivamente, con las relaciones:

$$LSi = \bar{x}_i + MSD/2 \quad \text{y} \quad Lli = \bar{x}_i - MSD/2 \quad \text{Notar que: 95\%CI (Lli ; LSi)}$$

	<b>Placebo</b>	<b>Droga A</b>	<b>Droga B</b>	<b>Mezcla A+B</b>	<b>Droga C</b>
<b>LSi</b>	121,58	110,78	109,68	109,48	115,58
<b>Medias</b>	120,1	109,3	108,2	108	114,1
<b>Lli</b>	118,62	107,82	106,72	106,52	112,62

*Paso 3 :* Se representan estos valores como se muestra en el Gráfico 18.1 siguiente. Ahí se puede ver claramente que el intervalo del Grupo 1 (placebo) no se superpone con los demás y se concluye que difiere significativamente de los otros. Queda demostrado que el medicamento reduce el tiempo de cura. Además, el grupo 5 (droga C) difiere de los demás, o sea las drogas A y B. Se concluye que el medicamento con droga C tiene efectividad, pero no tanta como las otras drogas. Como los intervalos entre las drogas A, B y su cóctel se superponen, se puede deducir que no hay diferencia entre ellos. Por lo tanto, la sospecha de que el cóctel era mejor que cada una por separado no se pudo confirmar.

**Gráfico 18.1 : Comparaciones entre medias muestrales de la tabla anterior.**



## 18.4 Tamaños muestrales distintos

Cuando se desean hacer comparaciones con tamaños muestrales distintos, el método sigue siendo el mismo en el caso de las comparaciones planificadas, porque al descomponer el Anova en partes, se tiene en cuenta el tamaño muestral individual, como se mostró antes. Pero, para el caso de comparaciones “a posteriori” el método desarrollado en el punto anterior deja de ser válido. De todos los modelos propuestos para trabajar en este caso, de tener diferentes tamaños muestrales se elige el método GT2 simplificado. Gabriel en 1978 propuso hacer más sencillos los engorrosos cálculos del método de Hochberg (GT2), presentado en 1974, con una manera gráfica aproximada. Al igual que en el punto anterior, se pueden visualizar fácilmente las diferencias significativas, observando los intervalos que no se superponen. De hecho, los cálculos son muy similares, excepto que se usa otra distribución estadística, denominada *Máximo módulo studentizado de dos colas*, que se muestra en la Tabla 19 del Anexo.

En este caso, la Mínima Diferencia Significativa es del tipo:

$$MDS_{ij} = (\text{valor crítico de tablas}) \cdot (\text{error típico}) = M_{\alpha(k; v)} \sqrt{DS_i^2 + DS_j^2}$$

Donde  $M_{\alpha(k; v)}$  se obtiene de la Tabla 19 con  $k = (a/2)(a-1)$  y  $v = N - a$   
 Por su parte el error típico se puede relacionar con el  $MS_d$  resultando:

$$MDS_{ij} = M_{\alpha(k; v)} \sqrt{(MS_d / n_i + MS_d / n_j)}$$

Con estos datos se arma una tabla de todas las comparaciones posibles para poder sacar las conclusiones necesarias. Sin embargo, como esto sería muy engorroso, se prefiere desarrollar el método simplificado por Gabriel, en un ejemplo a continuación:

En una investigación sobre gastos farmacéuticos semanales, relacionado con los enfermos internados, en 8 hospitales de una provincia; se eligieron 128 casos, tratando de mantener la proporción de sexos y con similares características de internación y con edades similares. Los datos recogidos expresados en cientos de pesos, se muestran en el Cuadro 18.5 :

**Cuadro 18.5: Gastos farmacéuticos en internados (por \$100).**

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7	Hospital 8
Medias	352.4	282.2	312.4	295.4	276.2	252.4	234.6	263.8
$n_i$	20	18	18	12	16	14	18	12
$DS_i^2$	45.8	40.4	20.5	30.4	28.8	30.4	20,8	20.4

*Paso 1)* Se calcula la varianza ponderada usando las muestrales con :

$$MS_d = \frac{\sum_1^8 (n_i - 1) \cdot DS_i^2}{\sum_1^8 (n_i - 1)} = \frac{(20-1)45,8 + (18-1)40,4 + \dots + (12-1)20,4}{(19+17+17+11+15+13+17+11)} = 3.645,1 / 120 = 30,4$$

Los grados de libertad asociados son  $v = N - a = (128 - 8) = 120$

Paso 2) Buscando en tablas con  $k = (a/2)(a-1) = (8/2)(8-1) = 28$  y  $v = 120$  resulta:

$$M_{\alpha(k;v)} = M_{0,05(28;120)} = 3,934$$

Paso 3) Se calculan los límites superior e inferior para cada media muestral, para armar los intervalos de comparación, en forma similar a la vista. Cuando dos intervalos no se superpongan significará que hay diferencia significativa entre las dos muestras.

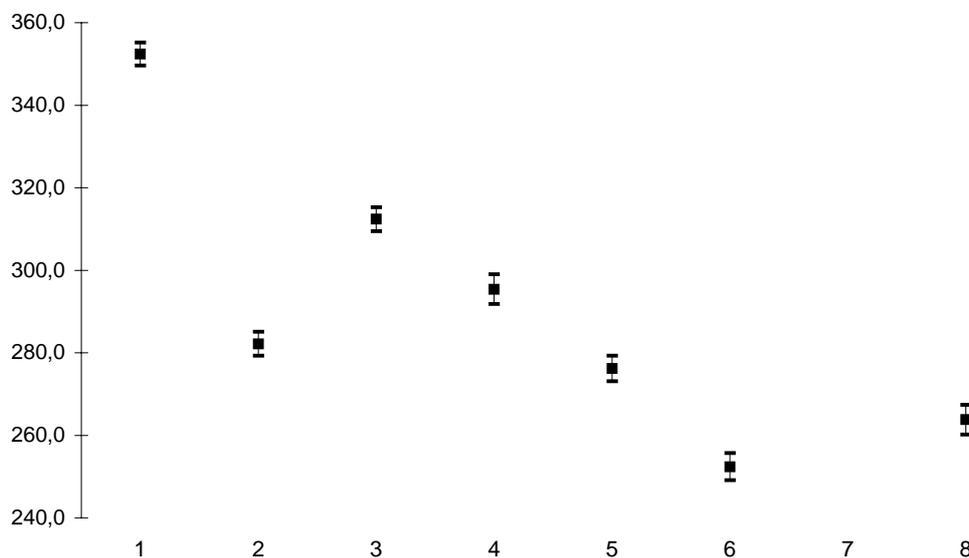
$$LSi = \bar{x}_i + E_i \quad \text{y} \quad LLi = \bar{x}_i - E_i \quad \text{Con}$$

$$E_i = (1/\sqrt{2}) \cdot M_{\alpha(k;v)} \cdot \sqrt{MS_d / ni} = 0,7071 \cdot 3,934 \cdot \sqrt{30,4} \sqrt{1/ni} = 15,34 \sqrt{1/ni}$$

Con esta relación se pueden ahora calcular, los intervalos de confianza para una de las medias muestrales, como se muestra a continuación con los 95% CI (LLi ; LSi):

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7	Hospital 8
Ei	3,43	3,62	3,62	4,43	3,84	4,10	3,62	4,43
LSi	355,83	285,82	316,02	299,83	280,04	256,50	238,22	268,23
Medias	352,40	282,20	312,40	295,40	276,20	252,40	234,60	263,80
LLi	348,97	278,58	308,78	290,97	272,36	248,30	230,98	259,37

Paso 4) Se grafican los límites de los intervalos, para detectar superposiciones:



Si bien la mayoría son diferentes, se pueden destacar dos cosas, mientras en el Hospital 1 los gastos son máximos, en el Hospital 6 son mínimos. Con esa información, el investigador encontró uno de los motivos principales de las diferencias: En el primer caso los médicos no recibían según el Vademécum, mientras que en el Hospital 6 sí. Lo cual le permitió reducir los gastos en forma sencilla y rápida.

## 18.5 Diseño básico de experimentos

Cuando se desea hacer un experimento se debe identificar claramente la magnitud clínica en los objetivos del mismo. Si es más de una magnitud se deben emplear los análisis de regresión y correlación que se verán más adelante o el análisis multivariado que está más allá de los objetivos de la presente obra. Pero ahora, ya se está en condiciones de discutir la parte básica.

Las “reglas de oro” a tener en cuenta son:

- *Cuanto mayor sea el número de muestras, más confiables serán las conclusiones.*
- *Siempre que se pueda, se deben tomar muestras de igual tamaño.*
- *Se debe tratar de usar modelos paramétricos en lugar de no paramétricos ( el más potente).*
- *Cuando haya que investigar a más de dos grupos, diseñar con modelos de Anova y planificar de antemano las comparaciones entre muestras a realizar.*

La primera regla apunta a mejorar el nivel de confianza y a potenciar el modelo elegido usado para la toma de decisiones. La segunda busca simplificar los cálculos porque las fórmulas empleadas en los casos de diferente tamaño se complican demasiado. La tercera se debe al hecho que muchas veces, en un mismo experimento, se pueden emplear varios modelos estadísticos y siempre los paramétricos son mucho más potentes que sus equivalentes no paramétricos.

La cuarta regla indica que hay que tratar de usar los modelos de Anova cada vez que se pueda. En efecto, la capacidad que tienen de descomponer la varianza total del experimento en términos aditivos es decisiva. Toda medición estará afectada de una cierta indeterminación llamada error y otro tanto ocurre con los grupos de mediciones del experimento. Nunca puede saberse a ciencia cierta las causas que originan la variabilidad en los resultados experimentales. Todo lo que puede hacerse es acotarla a términos manejables en la práctica; por ejemplo para medir la longitud de una mesa basta una regla común, no es necesario hacerlo en décimas de milímetro con otra regla especial. Sin embargo, a veces, el investigador necesita saber cuanta variabilidad genera el efecto de un cierto factor. La manera tradicional de hacerlo, era dejar fijos o constantes todos los demás factores que él sospecha como causas de variabilidad, y en forma controlada hace variar a uno solo de ellos para estudiarlo.

Por ejemplo, si se desea conocer la influencia del factor humano en las mediciones, se usa siempre el mismo equipo, los mismos reactivos, el mismo método, una única muestra (fraccionada en alícuotas de ser necesario), a la misma temperatura y humedad ambiental, etc. Y con mediciones repetidas de la misma muestra, se hace operar a un individuo, luego a otro y así hasta el último. De esta forma, si todo queda casi constante, la única fuente de variabilidad será la de usar diferentes personas. Antes, se les hacía hacer una medición a cada uno y la varianza total era una forma de cuantificar la variabilidad del factor humano más la inherente al sistema de medición que es inevitable. Luego, si una sola persona repetía la misma cantidad de mediciones, la varianza total obtenida se la podía suponer como la del error casual. Si se comparaban ambas se-

ries de mediciones se podía tener una idea de las diferencias en exactitud con el modelo Student, y las diferencias en precisión comparando las varianzas con el modelo de Fisher.

Pero desde la aparición del Anova toda esta operatoria fue dejada un lado. Por ejemplo, en el mismo experimento ahora se hace realizar a cada uno de los individuos la misma cantidad de mediciones junto a la persona de control, y con el modelo de Anova de un factor se puede descomponer la varianza total en dos partes: una debida al factor humano (MSe) y la otra, remanente (MSd) debida al error. Con un Modelo II se puede estimar la relación porcentual entre ambas, pero con un Modelo I se puede todavía comparar al control contra las otras personas y además a éstas entre sí. Todas estas ventajas, se logran con el mismo gasto y con la misma duración del experimento que a la manera tradicional. El secreto está en un buen diseño del experimento y en la planificación de las comparaciones, antes de realizarlo.

Aún más, se puede comparar más de un factor a la vez usando otros modelos de Anova que se verán más adelante. En el ejemplo anterior, además de estudiar la influencia del factor humano, se puede analizar la influencia de otros factores como: tipo de instrumento, marca de reactivos, técnicas diferentes, etc. Y para ello se pueden usar Modelos de Anova Multifactoriales con o sin replicación.

## 18.6 Comparaciones no paramétricas

Cuando se encuentren diferencias significativas, con el modelo de Kruskal-Wallis, como en el ejemplo del capítulo anterior, se debe continuar el estudio con las comparaciones múltiples para el caso de Estadística No Paramétrica. En este punto se mostrará el modelo STP para el caso de muestras de igual tamaño, basado en el estadístico U de Mann-Whitney. Si las muestras son de diferente tamaño, se puede emplear el test de la U para dos muestras y trabajar sobre todos los pares posibles. Otra vez se tiene una ventaja con muestras de igual tamaño, porque se tiene un modelo que simplifica los cálculos, de por sí muy engorrosos.

Se explicará este modelo con el ejemplo anteriormente mencionado.

*Paso 1)* Se ordenan los datos por grupo en orden creciente, como se muestra abajo:

**Cuadro 18.6: Los datos del Cuadro 17.6 ordenados en forma creciente.**

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C
1	115	107	106	106	112
2	117	107	106	107	112
3	117	108	107	107	112
4	117	109	107	107	113
5	118	109	108	108	114
6	120	110	108	108	115
7	121	110	108	108	115
8	125	110	110	109	115
9	125	111	111	109	116
10	126	112	111	111	117
Total	1201	1093	1082	1080	1141

*Paso 2)* Para cada par de muestras hay que calcular el valor  $U_p$  de la manera siguiente: hay que contar para cada dato de una muestra, el número de datos de la otra muestra que tienen un valor más bajo, se le asigna un punto a cada uno, y el número de datos empatados con él a los que se le asigna  $\frac{1}{2}$  punto. Por ejemplo, si se comparan las muestras placebo y droga A, se puede notar que el valor más chico del grupo placebo, es mayor cualquiera de los de A. Por lo tanto si se toma el primer valor de Placebo: 115, hay 10 valores de A menores que él y le corresponden 10 puntos. Luego se toma el segundo valor 117 y ocurre lo mismo, por lo que son 10 puntos más y así sucesivamente para los restantes, se tendrá un total de 10. 10 = 100 puntos. Otro tanto ocurre si se compara el placebo contra la droga B y contra el cóctel de drogas A+B. Esto es,

$$C(\text{placebo-droga A}) = 10 + 10 + \dots + 10 = 100 \text{ puntos} = C(\text{placebo-droga B}) = C(\text{placebo-cóctel})$$

Pero con la droga C, la cosa cambia, ahora si se toma el valor más chico del placebo: 115 se puede ver que hay 5 valores más chicos y 3 valores iguales, o sea, es un total de  $5 + 1,5$  puntos para ese primer caso. Para el segundo dato :117 se tienen 9 valores más chicos y un empate, o sea un total de  $9,5$  puntos. Para el tercer y cuarto dato ocurre lo mismo. Para los restantes datos todos los valores son menores y le tocan 10 puntos. Esto es,

$$C(\text{placebo-droga C}) = 6,5 + 9,5 + 9,5 + 9,5 + 10 + 10 + 10 + 10 + 10 + 10 = 95 \text{ puntos}$$

Análogamente, tomando ahora el grupo de la droga A, se lo va comparando contra los otros tres grupos como sigue:

$$C(\text{droga A-droga B}) = 3 + 3 + 5,5 + 7 + 7 + 7,5 + 7,5 + 7,5 + 9 + 10 = 67 \text{ puntos}$$

$$C(\text{droga A-droga A+B}) = 2,5 + 2,5 + 5,5 + 8 + 8 + 9 + 9 + 9 + 9,5 + 10 = 73 \text{ puntos}$$

$$C(\text{droga A-droga C}) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1,5 = 1,5 \text{ puntos}$$

$$C(\text{droga B-droga A+B}) = 0,5 + 0,5 + 2,5 + 2,5 + 5,5 + 5,5 + 5,5 + 9 + 9,5 + 9,5 = 50,5 \text{ puntos}$$

$$C(\text{droga B-droga C}) = 0 \text{ (todos los datos de droga C son mayores)}$$

$$C(\text{droga A+B-droga C}) = 0$$

*Paso 3)* Ahora se determina el valor  $U_p$  correspondiente para cada comparación C realizada en el paso anterior, eligiendo el mayor de los dos casos posibles:  $U_p = C$  ó  $U_p = n^2 - C$  y se vuelca en una tabla como la siguiente:

**Cuadro 18.7 : Puntajes del estadígrafo  $U_p$  .**

$U_p$	Placebo	Droga A	Droga B	Cóctel A+B	Droga C
Placebo	/				
Droga A	100**	/			
Droga B	100**	67	/		
Cóctel A+B	100**	73	50,5	/	
Droga C	95**	98,5**	100**	100**	/

El valor crítico para  $U_p$  es :

$$U_{\alpha(a;n)} = (n^2 / 2) + n \cdot Q_{\alpha(k;\infty)} \sqrt{\frac{(2n+1)}{24}}$$

Donde se usa la tabla del rango studentizado para  $k = 5$  y un número infinito de grados de libertad, donde es  $Q_{0,05(5; \infty)} = 3,858$  y  $Q_{0,01(5; \infty)} = 4,603$ ;  $n = 10$ . Luego:

$$U_{0,05(5; 10)} = (100 / 2) + 10 \cdot 3,858 \sqrt{\frac{(2(10)+1)}{24}} = 50 + 36,089 = 86,089 \approx 86,1$$

$$U_{0,01(5; 10)} = (100 / 2) + 10 \cdot 4,603 \sqrt{\frac{(2(10)+1)}{24}} = 50 + 43,057 = 93,057 \approx 93,1$$

Se comparan estos valores con los de la tabla anterior y las diferencias muy significativas se las indica con un doble asterisco. Mirando la tabla en sentido vertical se observa que en la primera columna, el placebo difiere significativamente de los otros cuatro grupos. La droga A solo difiere de la droga C. La droga B difiere del cóctel y de la droga C. El cóctel difiere de la droga C. Mirando en sentido horizontal se puede ver que la droga C difiere del resto, el cóctel del placebo y B, la B del placebo lo mismo que la A.

Otra forma de ver las diferencias en su conjunto es con:

$$\mu_{\text{placebo}} \gg \mu_{\text{droga C}} \gg \mu_{\text{droga A}} \approx \mu_{\text{droga B}} \approx \mu_{\text{cóctel A+B}}$$

Donde el símbolo  $>$  representa diferencias significativas y el símbolo  $\gg$  se usa para las diferencias muy significativas. Como se puede ver, hay mucha coincidencia con los resultados del modelo paramétrico visto más arriba. Sin embargo, esto no siempre es así, pues por lo general este modelo no discrimina tan bien, las diferencias entre medias muestrales como lo hace su equivalente modelo paramétrico.

## 18.7 Bioequivalencia

Los llamados tests de *Bioequivalencia* (o pruebas de equivalencia) han tenido un desarrollo reciente en Bioestadística. La diferencia de este tipo de test con los vistos anteriormente, es que aquí interesa ver si el efecto de un tratamiento clínico cualquiera, es superior en promedio al del placebo. O sea, en lugar de plantear una hipótesis nula habitual de que no hay diferencia entre las medias de los grupos, para poder así rechazarla cuando haya evidencia suficiente, la estrategia en los test de equivalencia consiste en suponer que la diferencia entre las medias es igual o mayor un valor de la magnitud clínica que se considera inferior clínicamente hablando.

De esta forma, la hipótesis alternativa queda: *La diferencia de medias es menor que una cantidad clínicamente relevante, y por lo tanto los tratamientos pueden ser considerados equivalentes*. Notar que, en caso de rechazar la hipótesis nula, habrá validación estadística de Bioequivalencia entre los métodos. Por ejemplo, se tiene un tratamiento viejo y aparece uno nuevo con ciertas ventajas como de menor costo, más rápido, menos invasivo, etc. Entonces el problema reside en ver si ambos métodos son equivalentes desde un punto de vista clínico, para poder efectuar el reemplazo del viejo con el nuevo.

*Ejemplo:* Se desea medir el índice cardíaco (CI), que es la relación entre el flujo sanguíneo expresado en litros por minuto, y la superficie de cuerpo del paciente. Esta magnitud continua se mide con un método invasivo llamado: termo dilución (TD), que consiste en colocar un catéter en el corazón para realizar la medición. Se propone un nuevo método llamado la bioimpedancia (BI), que consiste en adherir un instrumento externo en la piel del cuerpo del paciente, el cual mide una diferencia eléctrica. La gran ventaja es que este método es no invasivo. Se busca decidir si hay Bioequivalencia entre ambos métodos para poder efectuar el reemplazo. El valor referencial es de 2,75 litros/minuto/m<sup>2</sup>. Cuando el paciente tiene un valor muy por encima o por debajo, significa que hay un problema con el flujo del torrente sanguíneo. Esta y otras pruebas ayudan al médico a decidir el tratamiento adecuado para cada caso. El método TD es más confiable que el método BI, por lo tanto se adopta el siguiente criterio:

*Si el nuevo método BI, cae dentro del 20% del valor del viejo método (TD), entonces ambos métodos pueden ser considerados equivalentes.*

Se sabe que el valor de corte es 2,75 litros/minuto/m<sup>2</sup>. El 20% de este valor es  $\delta = 0,55$ , entonces se puede plantear la siguiente hipótesis nula:

$H_0 : | \mu_{TD} - \mu_{BI} | \geq 0,55$  - El método BI cae fuera del intervalo admisible

$H_1 : | \mu_{TD} - \mu_{BI} | < 0,55$  - El método BI cae dentro de la dispersión admisible del 20%

Se miden un total de  $N = 68$  individuos con el BI y resulta un promedio de 2,68 litros/minuto/m<sup>2</sup>, con un desvío de 0,26 litros/minuto/m<sup>2</sup>, entonces el error estándar de estimación será:

$$DS (BI) = 0,26 / (68)^{1/2} = 0,0265 \text{ litros/minuto/m}^2,$$

Y el test para probar la  $H_0$  es de una sola cola para  $v = 68 - 1 = 67$  grados de libertad. De tablas se obtiene el valor Student con  $t_{0,95; 67} = 1,67$ , entonces es:

$$t = (2,75 - 2,68) / 0,0265 = 2,64 > 1,67 \text{ por lo tanto se rechaza la hipótesis nula.}$$

Esto significa que se tiene prueba estadística para validar la  $H_1$ , y se pueden considerar equivalentes a ambos métodos. O bien, hay Bioequivalencia entre el método TD y el BI. Notar que otra manera de hacer lo mismo es calcular:

$$| \mu_{TD} - \mu_{BI} | + t_{0,95; 67} DS / \sqrt{N} = 0,07 + 1,67 \cdot 0,0265 = 0,114 < 0,55$$

Lo que muestra que la diferencia encontrada es menor que la máxima admisible de 0,55 y por lo tanto hay Bioequivalencia.

Cuando la pregunta sea si  $\mu_{TD} > \mu_{BI}$  entonces usando las hipótesis anteriores se trata de un test de una sola cola, en cambio si la pregunta es  $\mu_{TD} ? \mu_{BI}$  se tratará de un test de dos colas y dependiendo de ello se usarán distintos valores de tablas. Sin embargo, la forma más sencilla es usando el intervalo de confianza, abierto para una cola y cerrado para dos. En el caso de intervalo abierto solo interesa el valor límite como se mostró más arriba: si es menor hay equivalencia y si es mayor no la hay.

## 18.8 Problemas propuestos

- 1) Tiene sentido hacer comparaciones múltiples, cuando el Anova es significativo. **V F**
- 2) Las comparaciones múltiples se pueden usar en el Modelo I y II de Anova. **V F**
- 3) Los modelos de comparaciones múltiples solo sirven para comparaciones planeadas. **V F**
- 4) En comparaciones no planificadas solo se pueden usar modelos no Paramétricos. **V F**
- 5) La componente añadida de varianza se usa en los Modelos II de Anova. **V F**
- 6) La componente añadida es un porcentaje de la variabilidad total del experimento. **V F**
- 7) Conviene planificar las comparaciones antes de hacer el experimento. **V F**
- 8) El modelo de Tukey-Kramer se usa cuando hay tamaños muestrales diferentes. **V F**
- 9) Los modelos de Gabriel se usan en un gráfico para simplificar las comparaciones. **V F**
- 10) Con el modelo Student se pueden comparar hasta dos medias muestrales. **V F**
- 11) La MDS es el producto entre un valor de tablas y el error típico del modelo. **V F**
- 12) El MDS para muestras de diferente tamaño no es una constante. **V F**
- 13) Para aumentar la confiabilidad hay que aumentar el número de muestras. **V F**
- 14) Conviene usar los modelos Paramétricos siempre que se pueda. **V F**
- 15) Para simplificar los cálculos conviene tomar muestras de igual tamaño. **V F**
- 16) Conviene diseñar el experimento usando la estrategia del Anova. **V F**
- 17) No se puede medir la incidencia del factor humano en un experimento. **V F**
- 18) El modelo de Kruskal-Wallis es el Anova de 1 factor de tipo no paramétrico. **V F**
- 19) No se pueden hacer comparaciones múltiples no paramétricas. **V F**
- 20) Explicar el concepto de Bioequivalencia

21) En un experimento sobre la acción de un medicamento contra el Colesterol, se midió la reducción del mismo al cabo de una semana de iniciado el tratamiento. Determinar si hay diferencia significativa entre los cuatro grupos testeados:

Nº	40-44 años	45-49 años	50-54 años	55 años y más
1	189	175	186	100
2	238	205	190	128
3	250	198	170	110
4	190	200	160	139
5	210	180	195	154

- a) En caso de que corresponda hacer comparaciones múltiples si se planificó comparar los mayores de 54 años contra todos los demás, y luego los más jóvenes contra los dos intermedios.
- b) Buscar todas las comparaciones posibles con un modelo paramétrico y otro no paramétrico.

22) En un experimento de Bacharat (1940) sobre el factor de difusión de una sustancia, la cual si está presente en el momento de la inoculación incrementa la mancha que esta produce en la piel de un conejo. Se eligieron al azar 6 localidades geográficas diferentes y a 6 conejos de cada uno. La idea era estudiar si la ubicación geográfica tenía influencia. Se midió el tamaño de la marca en  $\text{cm}^2$ . Los datos obtenidos fueron los siguientes:

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre conejos	12.8333	5	2.56673	4,39
Dentro de conejos	17.5266	30	0.5832	
Total	30.2599	35		

Hallar la componente añadida de varianza.

23) Averiguar si las medias son diferentes en las tasas de citocromo-oxidasa:

-24 hs. después de una inyección de metaxyclo  $n = 5$  ; media = 24,8 ; varianza = 0,81

-Control  $n = 3$  ; media = 19,7 ; varianza = 1,96

(aplicada al macho de la especie *Periplaneta*).

24) Aplicar los modelos de comparaciones múltiples a los problemas planteados al final del capítulo anterior, cuando los resultados sean significativos. Discutir acerca del significado de los resultados encontrados para cada caso.

25) Se ha medido un suero control de cierta magnitud clínica con cinco técnicas diferentes, provenientes cada una de un laboratorio diferente. Se trata de establecer si los laboratorios miden lo mismo, o si hay diferencia entre ellos. El suero control tiene  $\mu = 135$  mg/dl con un desvío estándar  $DS = 2$  mg/dl y ha sido enviado por el laboratorio de referencia. Se lo fracciona en 50 alícuotas y se lo reparte equitativamente entre los 5 laboratorios que participan de la experiencia. Ninguno de ellos sabe que se trata del mismo suero, ni tampoco el valor del mismo. Solo se les informa que tienen que medir 1 muestra ciega de suero cada vez que se les envía. Se eligen al azar 10 días diferentes del mes, para efectuar el envío. No interesa ver la diferencia entre los días del mes, sino las diferencias que se observan de las mediciones repetidas. Luego de realizada la experiencia los valores informados son:

Nº	Labor. 1	Labor. 2	Labor. 3	Labor. 4	Labor. 5
1	140	133	128	136	145
2	142	132	127	135	142
3	138	134	129	137	143
4	139	136	132	134	141
5	141	137	130	135	142
6	137	135	133	136	140
7	136	138	131	134	141
8	139	134	129	136	143
9	140	136	132	135	144
10	141	135	133	134	140

Se pide:

- Detectar si hay diferencia entre los laboratorios con un Cuadro de Anova
- Mediante el método de Tukey compara los laboratorios entre sí.
- Usando el método gráfico de Gabriel presentar todas las comparaciones.
- Con el valor patrón decidir cuales laboratorios están calibrados en exactitud.

- e) Con un desvío máximo aceptable del 5%, controlar la precisión de cada laboratorio.
- f) Introduciendo el valor patrón en el gráfico de Gabriel ver la calibración en exactitud y comparar estas conclusiones con las del punto (d).
- g) Discutir sobre el sentido de las diferencias halladas y como deberían ser comunicadas a cada laboratorio involucrado.

26) Decidir si hay diferencia entre las drogas testeadas de acuerdo a los resultados siguientes

Nº	Placebo	Droga 1	Droga 2	Droga 3
1	450	350	250	280
2	460	360	240	290
3	455	370	245	295
4	470	360	250	285
5	465	380	255	280

Usando cualquiera de las técnicas vistas en este capítulo para efectuar comparaciones múltiples:

- a) Las comparaciones planeadas fueron:
  - Placebo versus drogas
  - Droga 1 versus las otras dos
  - Droga 2 versus droga 3

- a) Usar los métodos de Gabriel para comparaciones no planeadas.

28) En un estudio de control de parásitos, cada rata fue inyectada con 500 larvas del parásito llamado *Nippostrongylus muris*. Diez días después fueron sacrificadas y se contó el número de gusanos adultos. Se quiere dilucidar la siguiente cuestión: ¿Hay diferencia en la resistencia a la invasión parasitaria por grupos de ratas suministradas por diferentes proveedores? Se analizaron 4 grupos diferentes, formados por 5 ratas cada uno. Los resultados fueron:

Nº	Grupo 1	Grupo 2	Grupo 3	Grupo 4
1	279	378	172	381
2	338	275	335	346
3	334	412	335	340
4	198	265	282	471
5	303	286	250	318