

12A. COMPARACIONES MÚLTIPLES DE MEDIAS.

La comparación de las tres medias de *Glucosa*, en Imc Normal (107,7 mg/dl), Sobrepeso (154,2 mg/dl) y Obesidad (174,8 mg/dl) analizada en Unidad 11 resulta significativa ($p < 0,001$; Tabla 11A.3). La conclusión de este análisis es demasiado general: algunas de las tres medias son significativamente diferentes entre sí.

Se puede precisar más la conclusión de este análisis realizando **comparaciones múltiples** dos a dos entre los grupos para averiguar qué niveles de masa corporal difieren en la media de *Glucosa* y cuáles son las comparaciones concretas responsables de la significación. Las comparaciones múltiples pueden ser:

- **Todas las comparaciones múltiples posibles.** El número total de posibles comparaciones múltiples dos a dos de k grupos es $k(k-1)/2$. En nuestro ejemplo $k = 3$ niveles de masa corporal, por lo que se pueden realizar un total de $3(3-1)/2 = 3$ comparaciones.
- **Comparaciones múltiples respecto a una categoría de referencia.** En general se elige la categoría que representa la ausencia del carácter, la nula, basal o de referencia. El número de comparaciones es $k-1$, siendo k el número de grupos, que es mucho menor que $k(k-1)/2$. En este ejemplo serían $k-1 = 2$ comparaciones.

En función de los conocimientos teóricos sobre el tema concreto de investigación, el momento en que se formulan y la actitud con que se analizan, las comparaciones múltiples se pueden plantear de dos formas:

- **Comparaciones múltiples “a priori”** mediante una actitud confirmatoria de las hipótesis establecidas previamente en base a conocimientos teóricos sobre el asunto de la investigación. En ocasiones se desea contrastar diversas hipótesis específicas acerca del efecto de la exposición sobre la respuesta. Estas hipótesis se deben plantear en la fase de diseño del estudio, antes de obtener los datos, “a priori”. Una vez finalizado el estudio se realiza su análisis mediante una actitud confirmatoria de las hipótesis previamente planteadas por parte del investigador. En el ejemplo, se puede plantear “a priori” valorar el efecto sobre la *Glucosa* de tener Sobrepeso u Obesidad respecto a Imc normal.
- **Comparaciones múltiples “a posteriori”** o “post hoc” mediante una actitud exploratoria de todas las posibles comparaciones. En general no hay suficiente conocimiento sobre el tema de investigación como para plantear hipótesis “a priori” y se realizan todas las posibles comparaciones múltiples “a posteriori” con fin de generar nuevas hipótesis.

Corrección del riesgo alfa

Con las comparaciones múltiples de medias sucede lo mismo que con las comparaciones múltiples de proporciones analizadas en la Unidad 7. El riesgo alfa global aumenta a medida que aumenta el número de comparaciones como se muestra en la Tabla 7A.6 y debe corregirse en cada comparación por algún procedimiento.

Tipos de pruebas de comparaciones múltiples de medias

Existen numerosas pruebas que permiten realizar comparaciones múltiples de medias de manera que el riesgo alfa del conjunto de comparaciones no supere un valor fijado previamente, habitualmente 0,05. De todas ellas se comentaran las pruebas t de Student o DM y la de Bonferroni. Ambas pruebas son los extremos en cuanto a corrección del riesgo alfa: DMS no le corrige nada y Bonferroni le corrige en exceso. Son de interés histórico y académico pero no se utilizan en la práctica habitual. Recomendamos utilizar las pruebas de Scheffé, Tukey o Dunnett según se muestra en la Figura 12A.1. La prueba a utilizar debe estar predeterminada antes de analizar los datos y sólo se debe emplear una. No es correcto realizar varias pruebas y utilizar la que más interese.

Figura 12A.1. Pruebas estadísticas para comparaciones múltiples de medias.

| Tipo de comparaciones múltiples | | |
|--|-----------------|--|
| Todas las posibles comparaciones múltiples | | Respecto a una categoría de referencia |
| Tamaño de grupos similar, Normalidad e Igualdad de varianzas | | Prueba de Dunnett |
| No | Sí | |
| Prueba de Scheffé | Prueba de Tukey | |

Videotutorial 12A1 Comparaciones3.avi

Con el archivo *Factores.sav* se muestra como se obtienen las comparaciones múltiples de la respuesta *Glucosa* entre las categorías de la variable predictora *Nivel de masa corporal Normal/Sobrepeso/Obesidad* por los procedimientos de *DMS*, *Bonferroni*, *Tukey*, *Scheffé* y *Dunnett* con el cuadro ANOVA de un factor.

▪ Prueba t de Student o DMS

En la Tabla 12A.1 se muestran la significación de la *prueba t de Student* de las tres comparaciones múltiples de *Glucosa* en los niveles de obesidad. Esta prueba también se llama DMS (Diferencia Mínima Significativa) o LSD (Least Significant Difference). Compara dos a dos las medias de los grupos utilizando la t de Student para grupos independientes y varianzas homogéneas. No corrige el riesgo alfa, por lo que aumenta la probabilidad de cometer error aleatorio tipo I. Apenas utilizada actualmente pero interesante desde el punto de vista histórico y didáctico. Asumiendo un riesgo α de 0,05 para cada comparación se concluye que las comparaciones significativas son Imc Normal – Obeso ($p = 0,001$) e Imc Normal – Sobrepeso ($p = 0,013$).

Tabla 12A.1. Comparaciones múltiples de *Nivel de masa corporal – Glucosa* con la prueba t de Student.

| Comparación | p | Diferencia de medias (mg/dl) | IC 95% (mg/dl) |
|------------------------|-------|------------------------------|----------------|
| Imc Normal – Sobrepeso | 0,013 | 47,2 | 10,0 – 84,4 |
| Imc Normal – Obeso | 0,001 | 67,1 | 29,0 – 105,2 |
| Sobrepeso – Obeso | 0,163 | 19,9 | -8,2 – 48,0 |

▪ Prueba de Bonferroni

En la Tabla 12A.2 se muestran la significación de la *prueba de Bonferroni*, también llamada prueba de Student-Fisher para comparaciones múltiples con la corrección de Bonferroni. Mantiene el riesgo alfa global en 0,05 por lo que el riesgo alfa de cada comparación se obtiene dividiendo 0,05 entre el número de comparaciones. Es similar a la prueba DSM pero multiplicando la p por el número de comparaciones con lo que se aumenta la probabilidad de cometer error aleatorio tipo II. Siguen siendo significativas las dos comparaciones anteriores. Es poco potente, muy conservadora y poco utilizada.

Tabla 12A.2. Comparaciones múltiples de *Nivel de masa corporal – Glucosa* con la prueba de Bonferroni.

| Comparación | p | Diferencia de medias (mg/dl) | IC 95% (mg/dl) |
|------------------------|-------|------------------------------|----------------|
| Imc Normal – Sobrepeso | 0,040 | 47,2 | 1,6 – 92,9 |
| Imc Normal – Obeso | 0,002 | 67,1 | 20,4 – 113,9 |
| Sobrepeso – Obeso | 0,490 | 19,9 | -14,6 – 54,4 |

Al igual que la anterior tiene valor histórico y didáctico. Ambas pruebas son los dos extremos: DSM no corrige nada el riesgo alfa (aumenta el error aleatorio tipo I) y Bonferroni lo corrige demasiado siendo poco potente (aumenta el error aleatorio tipo II). Se aconseja utilizar las pruebas de comparaciones múltiples de potencia intermedia (Figura 12A.1) que se describen a continuación.

▪ Prueba de Tukey

Se prefiere si se cumple la normalidad, la igualdad de varianzas y el tamaño de los grupos es similar. Corrige el riesgo alfa teniendo en cuenta todas las posibles comparaciones que se pueden realizar. En la Tabla 12A.3 se aprecia que siguen siendo significativas las comparaciones Imc Normal – Obeso ($p = 0,002$) e Imc Normal – Sobrepeso ($p = 0,035$).

Tabla 12A.3. Comparaciones múltiples de *Nivel de masa corporal – Glucosa* con la prueba de Tukey.

| Comparación | p | Diferencia de medias (mg/dl) | IC 95% (mg/dl) |
|------------------------|-------|------------------------------|----------------|
| Imc Normal – Sobrepeso | 0,035 | 47,2 | 2,6 – 91,8 |
| Imc Normal – Obeso | 0,002 | 67,1 | 21,5 – 112,8 |
| Sobrepeso – Obeso | 0,343 | 19,9 | -13,8 – 53,6 |

▪ Prueba de Scheffé

Es aplicable tanto si el número de elementos por grupo es igual o diferente, por lo que se puede utilizar en caso de tamaños de grupos diferentes. Es robusta, tanto si no se cumple la igualdad de varianzas en todos los grupos como si no se cumple la normalidad, por lo que se puede utilizar en caso de vulneración de los supuestos de normalidad o igualdad de varianzas. En la Tabla 12A.4 se observa que son significativas las mismas comparaciones que con las pruebas anteriores.

Tabla 12A.4. Comparaciones múltiples de *Nivel de masa corporal – Glucosa* con la prueba de Scheffé.

| Comparación | p | Diferencia de medias (mg/dl) | IC 95% (mg/dl) |
|------------------------|-------|------------------------------|----------------|
| Imc Normal – Sobrepeso | 0,046 | 47,2 | 0,7 – 93,8 |
| Imc Normal – Obeso | 0,003 | 67,1 | 19,4 – 114,8 |
| Sobrepeso – Obeso | 0,377 | 19,9 | -15,3 – 55,1 |

▪ Prueba de *Dunnett*

Está diseñada para realizar un número determinado de comparaciones. Se utiliza si solo se pretende analizar las comparaciones respecto a una categoría de referencia, habitualmente la categoría basal o de referencia, por ejemplo varios tratamientos respecto al grupo tratado con placebo. En cuanto a corrección del riesgo alfa es la más eficiente, puesto que tiene en cuenta la circunstancia de que hay menos comparaciones. En la Tabla 12A.5 se observa que sobrepeso y obesidad son significativas respecto a Imc normal.

Tabla 12A.5. Comparaciones múltiples de *Nivel de masa corporal – Glucosa* con la prueba de *Dunnett*.

| Comparación | p | Diferencia de medias (mg/dl) | IC 95% (mg/dl) |
|------------------------|-------|------------------------------|----------------|
| Imc Normal – Sobrepeso | 0,023 | 47,2 | 5,9 – 88,6 |
| Imc Normal – Obeso | 0,001 | 67,1 | 24,8 – 109,5 |

Pruebas no paramétricas de comparaciones múltiples de medias

- En teoría en caso de no cumplirse el supuesto de normalidad las comparaciones múltiples de medias se deberían realizar con pruebas no paramétricas pero en general no están incluidas en la mayoría de programas estadísticos. Son similares a las pruebas de comparaciones múltiples de medias paramétricas, por ejemplo la de Scheffé si son de tamaño diferente, Tukey si los grupos son de igual tamaño o *Dunnett* para comparaciones respecto a una categoría de referencia, pero comparando rangos medios en vez de valores medios en la fórmula del estadístico.
- Una manera práctica de hacerlo sería hacer las comparaciones múltiples dos a dos con la prueba no paramétrica U de Mann Whitney y posteriormente corregir las p por los procedimientos de Bonferroni o de Holm de manera similar a como se hizo en la Unidad 7 con las comparaciones múltiples de proporciones.
- Pero en la práctica dado que la prueba de Scheffé es bastante robusta, es la que se emplea en caso de vulneraciones de los supuestos de normalidad e igualdad de varianzas.