

Comprobación de hipótesis específicas

Comparaciones *post hoc*

Comparaciones a posteriori

Comparaciones múltiples

Procedimientos de comparaciones de hipótesis específicas de la investigación

- Procedimiento DHS de Tukey
- Procedimiento de Dunnett
- Corrección de Bonferroni
- Procedimiento de Scheffé

Capítulo 11. Comprobación de hipótesis específicas. Pruebas
a posteriori 333

Por lo tanto, si el diseño tiene más de dos grupos ($A > 2$, tiene más de una diferencia de medias), será necesario continuar el análisis de los datos con otras técnicas que permitan descubrir entre qué grupos se encuentran las diferencias estadísticamente significativas como las 'pruebas de contraste de hipótesis específicas', también conocidas como pruebas *post hoc* o pruebas a posteriori (por ejemplo, Tukey, Dunnett, Bonferroni, Scheffé...).

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	96,000	2	48,000	24,000	,001
Intra-grupos	12,000	6	2,000		
Total	108,000	8			

Pero, cuando $A > 2$, la pregunta es: ¿dónde se encuentran las diferencias que son estadísticamente significativas ya que con el ANOVA se obtuvo que sí había tales diferencias? ¿Es entre las medias de $a_1 - a_2$? ¿Es entre las medias de $a_1 - a_3$? O quizás ¿es entre las medias de $a_3 - a_2$?

Las diferencias entre las medias siempre se valoran en términos de valores absolutos ya que es la misma diferencia $a_1 - a_2$ que $a_2 - a_1$, solamente cambia el signo.

Pág. 344

Tabla 13. Diferencias de medias para un diseño $A = 3$.

Medias	a_1 Media = 10	a_2 Media = 6
Media $a_2 = 6$	4	-
Media $a_3 = 14$	4	8

Pág. 348

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	96,000	2	48,000	24,000	,001
Intra-grupos	12,000	6	2,000		
Total	108,000	8			

Pero, cuando $A > 2$, la pregunta es: ¿dónde se encuentran las diferencias que son estadísticamente significativas ya que con el ANOVA se obtuvo que sí había tales diferencias? ¿Es entre las medias de $a_1 - a_2$? ¿Es entre las medias de $a_1 - a_3$? O quizás ¿es entre las medias de $a_3 - a_2$?

Las diferencias entre las medias siempre se valoran en términos de valores absolutos ya que es la misma diferencia $a_1 - a_2$ que $a_2 - a_1$, solamente cambia el signo.

Tabla 13. Diferencias de medias para un diseño $A = 3$.

Pág. 344

Medias	a_1 Media = 10	a_2 Media = 6
Media $a_2 = 6$	4	-
Media $a_3 = 14$	4	8

Pág. 348

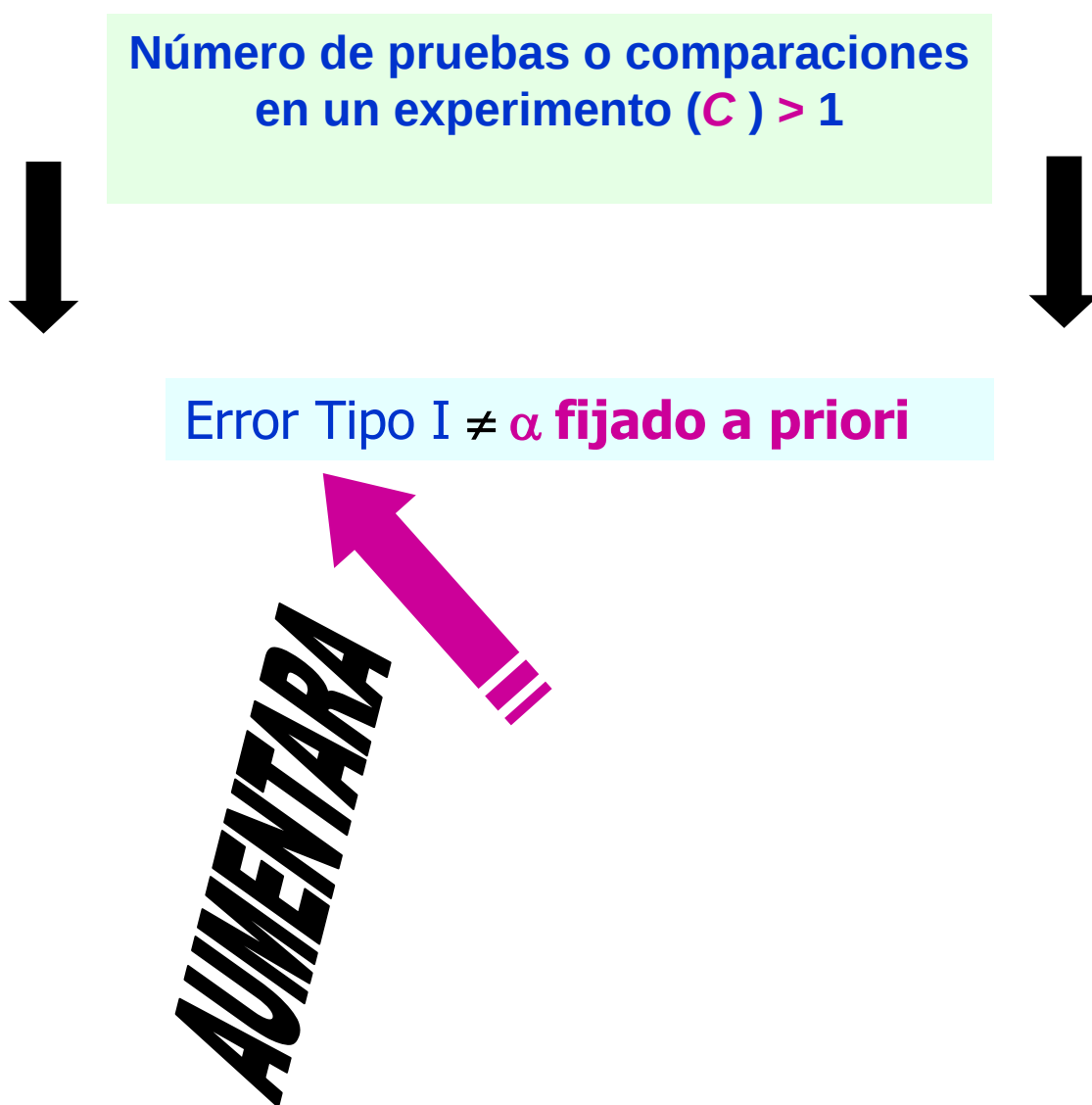
Con un diseño $A = 3$ existen 3 diferencias de medias entre los 3 grupos. Para conocer cuántas diferencias de medias simples existente en un diseño se puede aplicar la siguiente fórmula:

$$c = \frac{a(a-1)}{2}$$

$$c = \frac{ab(ab-1)}{2}$$

Pág. 348

¿**Problema?**: cuando se realizan pruebas a posteriori el **error tipo I** (rechazar H_0 cuando sea verdadera) puede ser **muy superior** al valor nominal indicado por el valor de alfa utilizado en el estudio.



Tasa de error de Tipo I

La Tasa de Error de Tipo I o alfa por experimento (α_{PE}) es:

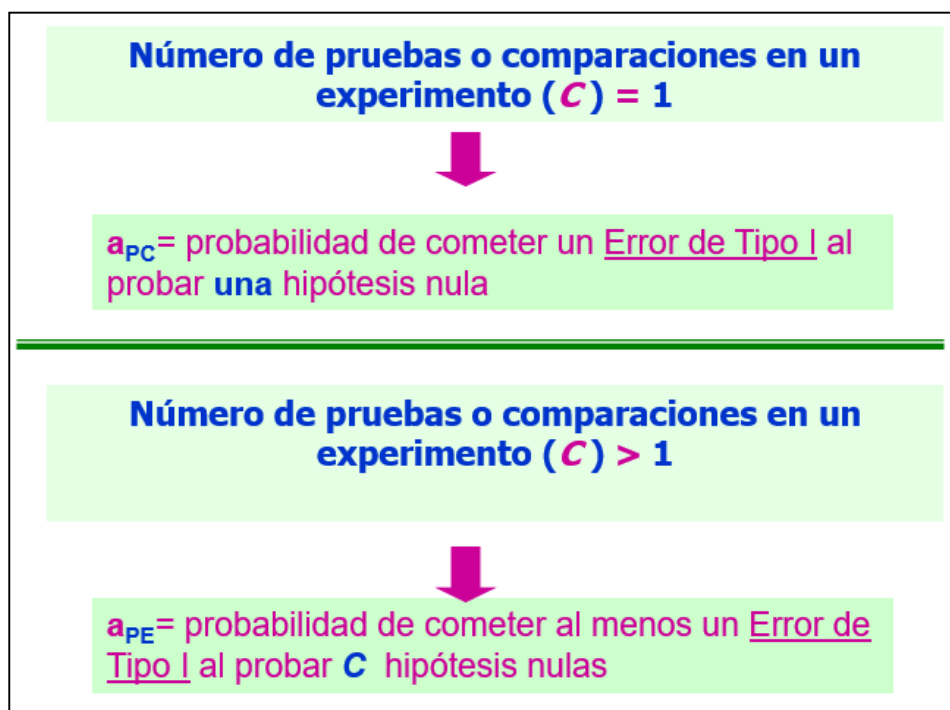
$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^C$$

Donde alfa por comparación, α_{PC} , es la probabilidad de cometer un error de Tipo I en 'una comparación' (comparación individual) y el alfa por experimento, α_{PE} , es la probabilidad de cometer al menos un error de Tipo I en un 'conjunto de comparaciones' (C) (comparaciones en total).

Pág. 349

Por lo tanto, el alfa por comparación es la probabilidad de rechazar incorrectamente una hipótesis nula cierta cuando se lleva a cabo una determinada comparación. Y la tasa de error de Tipo I por experimento es la probabilidad de cometer error de Tipo I cuando se analizan un conjunto de contrastes.

Pág. 350



Tasa de error de Tipo I

La Tasa de Error de Tipo I o alfa por experimento (α_{PE}) es:

$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^C$$

Donde alfa por comparación, α_{PC} , es la probabilidad de cometer un error de Tipo I en 'una comparación' (comparación individual) y el alfa por experimento, α_{PE} , es la probabilidad de cometer al menos un error de Tipo I en un 'conjunto de comparaciones' (C) (comparaciones en total).

Pág. 349

Por lo tanto, el alfa por comparación es la probabilidad de rechazar incorrectamente una hipótesis nula cierta cuando se lleva a cabo una determinada comparación. Y la tasa de error de Tipo I por experimento es la probabilidad de cometer error de Tipo I cuando se analizan un conjunto de contrastes.

Supongamos que en un estudio se van a realizar 4 comparaciones (pruebas de contraste). Si todas las hipótesis nulas fueran ciertas y $\alpha_{PC} = .05$ entonces la probabilidad de cometer al menos un Error de Tipo I al efectuar el conjunto de contrastes es:

$$\alpha_{PE} = 1 - (1 - .05)^4 = .1855$$

Ese valor de .1855 se aparta mucho del valor de alfa fijado a priori por el investigador o investigadora en .05 como error de Tipo I. Es decir, se tendría una probabilidad de error de Tipo I de .1855, que es totalmente inadmisibile. La probabilidad de rechazar una hipótesis nula siendo cierta es de .1855 (18.5%). ¿Qué ha pasado? La prueba de contraste de hipótesis se ha hecho más liberal (es decir, se rechaza la hipótesis nula con mayor facilidad), aumentando de este modo el error de Tipo I. ¿Cómo se puede corregir ese sesgo que amenaza a la validez de

Pág. 350

α_{PE} = probabilidad de cometer al menos un Error de Tipo I al probar **C** hipótesis nulas

α_{PE}

TASA DE ERROR DE TIPO I

TASA DE ERROR DE TIPO I POR EXPERIMENTO (PE)

DEPENDE

Valor α_{PC}

Número de comparaciones (**C**)

$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^C$$

Ejemplo: **3 comparaciones:**

Si todas las hipótesis nulas fueran **ciertas** y $\alpha_{PC} = 0.05$ entonces la probabilidad de cometer al menos un Error de Tipo I es:

$$\alpha_{PE} = 1 - (1 - 0.05)^3 = 0.1426$$

TASA DE ERROR DE TIPO I

TASA DE ERROR DE TIPO I POR EXPERIMENTO (PE)

CONTROLAR



¿CÓMO?

**Procedimientos de comparaciones
de hipótesis específicas**

EL CONTROL DE LA TASA DE ERROR TIPO I



Consecuencia: reducir el α_{PC} para poder controlar el α_{PE}

La prueba se hace **más conservadora**

El procedimiento más adecuado será:

1º Controle correctamente la tasa de Error de Tipo I

2º Cuando potencia estadística es máxima (menor Error Tipo II)

Por lo tanto, para tomar la decisión el investigador o investigadora debe considerar los siguientes aspectos:

1. El número de comparaciones (C) que la hipótesis plantea: pueden ser comparaciones exhaustivas (*a posteriori*) o pueden ser comparaciones planificadas (*a priori*).
2. Si las hipótesis experimentales son simples (plantea diferencias *entre pares de medias*) o complejas (plantea alguna diferencia entre medias que implica *un promedio de medias*).

Pág. 351

Figura 71. SPSS: Analizar → Comparar medias → ANOVA de un factor → *Post hoc*.

ANOVA de un factor: Comparaciones múltiples post hoc

Asumiendo varianzas iguales

☐ DMS ☐ S-N-K ☐ Waller-Duncan

☐ Bonferroni ☐ Tukey Tasa de errores tipo I (tipo II): 100

☐ Sidak ☐ Tukey-b ☐ Dunnett

☐ Scheffe ☐ Duncan Categoría de control: Último

☐ R-E-G-W F ☐ GT2 de Hochberg

☐ R-E-G-W Q ☐ Gabriel

Prueba

☒ Bilateral ☐ < Control ☐ > Control

No asumiendo varianzas iguales

☐ T2 de Tamhane ☐ T3 de Dunnett ☐ Games-Howell ☐ C de Dunnett

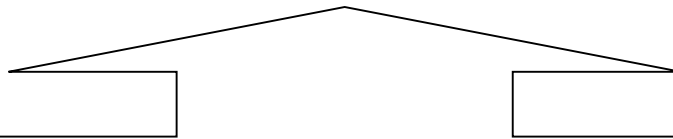
Nivel de significación: 0,05

Continuar Cancelar Ayuda

Pág. 341¹⁰

Procedimientos de comparaciones de hipótesis específicas de la investigación

- Procedimiento DHS de Tukey
- Procedimiento de Dunnett
- Corrección de Bonferroni
- Procedimiento de Scheffé



**Todos los procedimientos controla la Tasa
de Error de Tipo I**

Procedimiento DHS de Tukey

Es el más potente:

cuando se realizan **todas las comparaciones** posibles **dos a dos** y además son **simples**. Más utilizado

$$|\bar{Y}_g - \bar{Y}_h| \geq \frac{q(\alpha, a, g_{\text{error}})}{\sqrt{2}} \sqrt{MC_{\text{error}} \sum_{j=1}^a \frac{c_j^2}{n_j}}$$

Procedimiento de Dunnett

Es el más potente:

cuando se trata de comparar la media de **un grupo frente al Resto** y además son comparaciones **simples a priori**.

Y **C = a - 1** comparaciones

$$|\bar{Y}_g - \bar{Y}_h| \geq D(\alpha, a, g_{\text{error}}) \sqrt{MC_{\text{error}} \sum_{j=1}^a \frac{c_j^2}{n_j}}$$

Corrección de Bonferroni

Siempre que la hipótesis formule el número de comparaciones, aunque si **C** es grande entonces la prueba es poco potente.

Con comparaciones **simples** y **complejas**

$$\alpha_{PC} = \frac{\alpha_{PE}}{C}$$

$$|\bar{Y}_g - \bar{Y}_h| \geq \sqrt{F_{(a/C, 1, g_{\text{error}})}} \sqrt{MC_{\text{error}} \sum_{j=1}^a \frac{c_j^2}{n_j}}$$

Procedimiento de Scheffé

Es válido en cualquier circunstancia. Con comparaciones **simples** y **complejas**. Normalmente es la prueba **menos potente**

$$|\bar{Y}_g - \bar{Y}_h| \geq \sqrt{(a-1)F(\alpha, a-1, g_{\text{error}})} \sqrt{MC_{\text{error}} \sum_{j=1}^a \frac{c_j^2}{n_j}}$$

Asumen varianzas iguales

Procedimiento de Tukey

HSD de Tukey (diferencia honestamente significativa)

Utiliza el estadístico del rango estudentizado (q) para realizar todas las comparaciones simples por parejas entre los grupos. Establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por parejas.

Más potente: comparaciones simples y todas las comparaciones.

Procedimiento de Dunnett

Prueba t de comparaciones múltiples por parejas que compara un conjunto de tratamientos respecto a una única medida de control. La última categoría es la categoría de control predeterminada.

Más potente: comparaciones simples y $a-1$ comparaciones (comparación con un grupo control).

Corrección de Bonferroni

Utiliza las pruebas t para realizar comparaciones por parejas entre las medias de los grupos, pero controla la tasa de error global estableciendo que la tasa de error de cada prueba sea igual a la **tasa de error por experimento dividida entre el número total de contrastes**. Así, se corrige el nivel de significación observado en función de las múltiples comparaciones. Por ejemplo, si se realizan tres contrastes simultáneos y se desea mantener el nivel de significancia en .05, el alfa elegido para la prueba debe ser $.05/3 = .017$.

Procedimiento de Scheffé

Realiza comparaciones múltiples por parejas para todas las parejas de combinaciones de las medias posibles. Utiliza la distribución muestral F . Es la prueba más rigurosa (conservadora), cuando se hacen muchas comparaciones, y, por lo tanto, la probabilidad del error tipo I es menor (rechazar H_0 cuando es verdadera), pero aumenta el error de Tipo II (mantener H_0 cuando es falsa).

No asumen varianzas iguales

Las pruebas de comparaciones múltiples que no suponen varianzas iguales son:

- T2 de Tamhane
- T3 de Dunnett
- Games-Howell
- C de Dunnett

*Para analizar la **homogeneidad de varianzas (heterocedasticidad)** se puede utilizar la prueba de Levene. Si $p > \alpha$, se asume la homogeneidad de las varianzas.