Comparaciones Múltiples Y False Discovery Rate y su relación con Ciencia de Datos

Maestría en Cómputo Estadístico

Autores: Enrique Santibáñez & Edgar Baquero

2 de diciembre de 2020

Centro de Investigacón en Matemáticas, A.C.

Contenido

Introducción

Introducción

Papel de la inferencia estadística

- Cuestionarse acerca de parámetros que estimamos.

Introducción

Papel de la inferencia estadística

- Cuestionarse acerca de parámetros que estimamos.
- Lo hacemos a través de *Pruebas de Hipótesis*.

Introducción

Papel de la inferencia estadística

- Cuestionarse acerca de parámetros que estimamos.
- Lo hacemos a través de *Pruebas de Hipótesis*.
- Generalmente queremos comparar más de dos parámetros.

Introducción

Papel de la inferencia estadística

- Cuestionarse acerca de parámetros que estimamos.
- Lo hacemos a través de *Pruebas de Hipótesis*.
- Generalmente queremos comparar más de dos parámetros.
- Este proceso se conoce como Comparación Múltiple.

Aspectos preliminares

Pruebas de hipótesis simples

Definición

Se conoce como hipótesis simple, a una prueba en la que interviene sólo una única hipótesis nula H_0 y su complemento, la hipótesis alternativa H_1.

Example (Un quiz chévere...)

- *Fat Tony* va a ser juzgado como mafioso, y el Juez *Roy Snyder* lo hará:

Pruebas de hipótesis simples

Definición

Se conoce como hipótesis simple, a una prueba en la que interviene sólo una única hipótesis nula H_0 y su complemento, la hipótesis alternativa H_1.

Example (Un quiz chévere...)

- *Fat Tony* va a ser juzgado como mafioso, y el Juez *Roy Snyder* lo hará:
- 

Definición

Se conoce como hipótesis simple, a una prueba en la que interviene sólo una única hipótesis nula H_0 y su complemento, la hipótesis alternativa H_1.

Example (Un quiz chévere...)

- *Fat Tony* va a ser juzgado como mafioso, y el Juez *Roy Snyder* lo hará:
- 
-  

Pruebas de hipótesis simples

Example (Juicio oral - Continuación)

Roy Snyder plantea:

- H_0 : *Fat Tony* es inocente.
- H_1 : *Fat Tony* es culpable.

Los posibles resultados del juicio pueden mostrarse en la siguiente tabla:

	H_0 Cierta (inocencia)	H_0 Falsa (Culpable)
H_0 rechazada (Condenado)	Error tipo I	Decisión correcta
H_0 no rechazada (Libre)	Decisión correcta	Error tipo II

Observación

En el anterior ejemplo, pareciera que no tuviéramos relación entre el error tipo I y el error tipo II. Aún así, es posible probar que:

- $P(\text{error tipo I}) \to 0 \quad \Longrightarrow \quad P(\text{error tipo II}) \to 1$
- $P(\text{error tipo II}) \to 0 \quad \Longrightarrow \quad P(\text{error tipo I}) \to 1$

- Casi siempre damos prioridad a disminuir el error tipo I.

Pruebas de hipótesis simples

Observación

En el anterior ejemplo, pareciera que no tuviéramos relación entre el error tipo I y el error tipo II. Aún así, es posible probar que:

- $P(\textit{error tipo I}\ ) \to 0 \quad \Longrightarrow \quad P(\textit{error tipo II}) \to 1$
- $P(\textit{error tipo II}) \to 0 \quad \Longrightarrow \quad P(\textit{error tipo I}\ ) \to 1$

- Casi siempre damos prioridad a disminuir el error tipo I.
- En los casos donde se requiere controlar el error tipo **II**, trabajamos con *pruebas uniformemente más potentes*.

Observación

En el anterior ejemplo, pareciera que no tuviéramos relación entre el error tipo I y el error tipo II. Aún así, es posible probar que:

- $P(\textit{error tipo I}) \to 0 \quad \Longrightarrow \quad P(\textit{error tipo II}) \to 1$
- $P(\textit{error tipo II}) \to 0 \quad \Longrightarrow \quad P(\textit{error tipo I}) \to 1$

- Casi siempre damos prioridad a disminuir el error tipo I.
- En los casos donde se requiere controlar el error tipo II, trabajamos con *pruebas uniformemente más potentes*.
- Dado que el error tipo I quiere llevarse a 0, establecemos una cota superior para la cual este puede ser encontrado. Dicha cota se conoce como *nivel de significancia* y se denota con la letra α

Pruebas de hipótesis simples - Definiciones

Una vez establecida la filosofía básica de una prueba de hipótesis, planteamos algunos conceptos:

Definición

- Hipótesis de trabajo: *La aseveración acerca del espacio parametral Θ que nos interesa probar, junto con su complemento o hipótesis alternativa. Usualmente es denotado de la siguiente forma:*

$$H_0 : \theta \in \Theta_0; \quad H_1 : \theta \in \Theta_1,$$

donde $\{\Theta_0, \Theta_1\}$ es una partición de Θ, el espacio de parámetros.

Pruebas de hipótesis simples - Definiciones

Definición (Continuación)

- Estadístico de prueba: *Valor calculado en función de la muestra observada, denotado por T.*
- *Región de Rechazo (C): Conjunto de valores que puede tomar el estadístico de prueba T, para los cuales se rechaza H_0.*
- *Potencia: robabilidad de no cometer error de tipo II. En el caso de una hipótesis que plantee igualdad:*

$$\beta(\theta^*) = P(T \in C | \theta = \theta^*) \tag{1}$$

Pruebas de hipótesis simples - Definiciones

Definición (Continuación)

- *Nivel de significancia (α): puede ser definido en términos de (1):*

$$\alpha \leq \sup_{\theta \in \Theta_0} \beta(\theta). \tag{2}$$

- p-valor: *probabilidad, asumiendo la hipótesis nula como cierta, de haber observado un valor del estadístico de prueba al menos tan extremo como el que se observó:*

$$p = P(T > T_{obs} | \theta \in \Theta_0).$$

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

- Plantear la hipótesis nula y la hipótesis alternativa.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I. Plantear la hipótesis nula y la hipótesis alternativa.

II. Seleccionar un nivel de significancia α. El umbral probabiliístico bajo el cual la hipótesis será rechazada.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I Plantear la hipótesis nula y la hipótesis alternativa.

II Seleccionar un nivel de significancia α. El umbral probabilıístico bajo el cual la hipótesis será rechazada.

III Elegir la estadística de prueba adecuada T.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I. Plantear la hipótesis nula y la hipótesis alternativa.

II. Seleccionar un nivel de significancia α. El umbral probabiliístico bajo el cual la hipótesis será rechazada.

III. Elegir la estadística de prueba adecuada T.

IV. Encontrar la distribución de T bajo la hipótesis nula.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I Plantear la hipótesis nula y la hipótesis alternativa.

II Seleccionar un nivel de significancia α. El umbral probabiliístico bajo el cual la hipótesis será rechazada.

III Elegir la estadística de prueba adecuada T.

IV Encontrar la distribución de T bajo la hipótesis nula.

V Calcular la región crítica o región de rechazo C.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I Plantear la hipótesis nula y la hipótesis alternativa.

II Seleccionar un nivel de significancia α. El umbral probabilıístico bajo el cual la hipótesis será rechazada.

III Elegir la estadística de prueba adecuada T.

IV Encontrar la distribución de T bajo la hipótesis nula.

V Calcular la región crítica o región de rechazo C.

VI Encontrar el valor observado del estadístico de prueba T_{obs} de la muestra o, alternativamente, el p-valor.

Pruebas de hipótesis simples - Resumen

En resumen, realizar una prueba de hipótesis consiste en:

I Plantear la hipótesis nula y la hipótesis alternativa.

II Seleccionar un nivel de significancia α. El umbral probabiliístico bajo el cual la hipótesis será rechazada.

III Elegir la estadística de prueba adecuada T.

IV Encontrar la distribución de T bajo la hipótesis nula.

V Calcular la región crítica o región de rechazo C.

VI Encontrar el valor observado del estadístico de prueba T_{obs} de la muestra o, alternativamente, el p-valor.

VII Decidir si rechazar o no H_0 con base en la región C especificada en el paso (VI), o, alternativamente, rechazar H_0 si el p-valor obtenido pequeño.

Comparaciones Múltiples

Comparaciones Múltiples - Procedimiento

- Es posible juzgar acerca de un determinado número $m > 1$ de hipótesis nulas $H_{01}, \ldots, H_{0m}$.

Comparaciones Múltiples - Procedimiento

- Es posible juzgar acerca de un determinado número $m > 1$ de hipótesis nulas $H_{01}, \ldots, H_{0m}$.
- Es pertinente la realización de un procedimiento que nos permita evaluar la veracidad de una hipótesis general.

Comparaciones Múltiples - Procedimiento

- Es posible juzgar acerca de un determinado número $m > 1$ de hipótesis nulas $H_{01}, \ldots, H_{0m}$.
- Es pertinente la realización de un procedimiento que nos permita evaluar la veracidad de una hipótesis general.
- *Dutoit* et al (2003), estandarizó el procedimiento

Comparaciones Múltiples - Procedimiento

- Es posible juzgar acerca de un determinado número $m > 1$ de hipótesis nulas $H_{01}, \ldots, H_{0m}$.
- Es pertinente la realización de un procedimiento que nos permita evaluar la veracidad de una hipótesis general.
- *Dutoit* et al (2003), estandarizó el procedimiento
- Elegir y calcular un estadístico de prueba T_j para cada hipótesis individual j, con $j = 1, \ldots, m$

- Es posible juzgar acerca de un determinado número $m > 1$ de hipótesis nulas $H_{01}, \ldots, H_{0m}$.
- Es pertinente la realización de un procedimiento que nos permita evaluar la veracidad de una hipótesis general.
- *Dutoit* et al (2003), estandarizó el procedimiento
- Elegir y calcular un estadístico de prueba T_j para cada hipótesis individual j, con $j = 1, \ldots, m$
- Aplicar un procedimiento de prueba de hipótesis múltiple para determinar cuáles hipótesis se han de rechazar de manera que se controle de alguna forma específica el error tipo I.

Sobre la extensión del caso simple

- Una prueba de manea simultánea al conjunto de hipótesis $\{H_{01}, \ldots, H_{0m}\}$ no es equivalente a realizar m pruebas individuales.

Sobre la extensión del caso simple

- Una prueba de manea simultánea al conjunto de hipótesis $\{H_{01}, \ldots, H_{0m}\}$ no es equivalente a realizar m pruebas individuales.
- Primero, se necesitarían $\binom{m}{2} = \frac{m(m-1)}{2}$ comparaciones individuales.

Sobre la extensión del caso simple

- Una prueba de manea simultánea al conjunto de hipótesis $\{H_{01}, \ldots, H_{0m}\}$ no es equivalente a realizar m pruebas individuales.
- Primero, se necesitarían $\binom{m}{2} = \frac{m(m-1)}{2}$ comparaciones individuales.
- Segundo, con mayor relevacia, por independencia entre las hipótesis.

Sobre la extensión del caso simple

- Una prueba de manea simultánea al conjunto de hipótesis $\{H_{01}, \ldots, H_{0m}\}$ no es equivalente a realizar m pruebas individuales.
- Primero, se necesitarían $\binom{m}{2} = \frac{m(m-1)}{2}$ comparaciones individuales.
- Segundo, con mayor relevacia, por independencia entre las hipótesis.
- Esto es, el rechazo de H_{0i} podría influir (positiva o negativamente) en las posibilidades del rechazo de H_{0j}

Example (Lanzamiento de monedas - parte 1)

- Queremos probar si una moneda es equilibrada.

Sobre la extensión del caso simple

Example (Lanzamiento de monedas - parte 1)

- Queremos probar si una moneda es equilibrada.
- 9 de 10 lanzamientos son cara.

Example (Lanzamiento de monedas - parte 1)

- Queremos probar si una moneda es equilibrada.
- 9 de 10 lanzamientos son cara.
- La probabilidad de que se observe un resultado al menos tan extremo como ese, sería de $(10 + 1)(1/2)^{10} = 0{,}0107$.

Example (Lanzamiento de monedas - parte 1)

- Queremos probar si una moneda es equilibrada.
- 9 de 10 lanzamientos son cara.
- La probabilidad de que se observe un resultado al menos tan extremo como ese, sería de $(10+1)(1/2)^{10} = 0{,}0107$.
- Concluimos que la moneda no es justa.

Sobre la extensión del caso simple

Example (Lanzamiento de monedas - parte 2)

- Repitamos el experimento con 100 monedas.

Example (Lanzamiento de monedas - parte 2)

- Repitamos el experimento con 100 monedas.
- Esperaríamos que fuera igual o más raro, al lanzar una moneda equilibrado, observar 9 caras en 10 lanzamientos.

Sobre la extensión del caso simple

Example (Lanzamiento de monedas - parte 2)

- Repitamos el experimento con 100 monedas.
- Esperaríamos que fuera igual o más raro, al lanzar una moneda equilibrado, observar 9 caras en 10 lanzamientos.
- ¡Pues no!

Sobre la extensión del caso simple

Example (Lanzamiento de monedas - parte 2)

- Repitamos el experimento con 100 monedas.
- Esperaríamos que fuera igual o más raro, al lanzar una moneda equilibrado, observar 9 caras en 10 lanzamientos.
- ¡Pues no!
- De hecho la probabilidad de que en 100 repeticiones al menos una moneda sea equilibrada está dada por $1-(1-0,0107)^{100}=0{,}6604$

Sobre la extensión del caso simple

Example (Lanzamiento de monedas - parte 2)

- Repitamos el experimento con 100 monedas.
- Esperaríamos que fuera igual o más raro, al lanzar una moneda equilibrado, observar 9 caras en 10 lanzamientos.
- ¡Pues no!
- De hecho la probabilidad de que en 100 repeticiones al menos una moneda sea equilibrada está dada por $1 - (1 - 0,0107)^{100} = 0{,}6604$
- ¡Sería un error aplicar el razonamiento anterior para probar que las 100 monedas son equilibradas!

Sobre la extensión del caso simple

Observación

- *El anterior ejemplo, nos muestra la delicadeza de la multiplicidad.*
- *La noción de error se complica de manera creciente.*
- *Si una prueba simple se hace a un* 5 % *de confianza, afirmamos que existe un* 5 % *de probabilidad de que la hipótesis nula sea rechazada incorrectamente.*
- *Si se realizan* $m = 100$ *pruebas de hipótesis simultáneamente, donde todas son ciertas, el número esperado de rechazos incorrectos es 5*
- *Si las pruebas son independientes, la probabilidad de rechazar al menos una hipótesis incorrectamente es de* $1 - (1 - 0,05)^{100} = 0,994$*!!!*

Tabla de comparaciones múltiples

	Hipótesis No Rechazadas	Hipótesis Rechazadas	Total
Hipótesis Verdaderas	U	V	m_0
Hipótesis Falsas	K	S	m_1
	$m - R$	R	M

- m es el total de hipótesis realizadas.
- m_0 es el número de hipótesis nulas verdaderas.
- $m - m_0$ es el número de verdaderas hipótesis alternativas.
- V es el número de falsos positivos (error tipo I) (también conocido como *falso descubrimiento*).
- S es el número de verdaderos positivos (conocido como *descubrimiento verdadero*).
- K es el número de falsos negativos (error tipo II).
- U es el número de verdaderos negativos.
- $R = V + S$ es el número de hipótesis nulas rechazadas (conocido como *descubrimientos*)

Sobre los errores

Nos encontramos con la necesidad de controlar el error tipo I en el caso múltiple. Para esto, introducimos 3 nociones de error:

- **Tasa de Error por Comparación (PCER)**: Consiste de el valor esperado de errores de tipo I dividido entre el número total de hipótesis:

$$\mathrm{PCER} = \frac{\mathrm{E}(V)}{m}$$

Sobre los errores

Nos encontramos con la necesidad de controlar el error tipo I en el caso múltiple. Para esto, introducimos 3 nociones de error:

- **Tasa de Error por Comparación (PCER)**: Consiste de el valor esperado de errores de tipo I dividido entre el número total de hipótesis:

$$\mathrm{PCER} = \frac{\mathrm{E}(V)}{m}$$

- **Tasa de Error Global (FWER)**: Es la probabilidad de cometer uno o más errores de tipo I:

$$\mathrm{FWER} = P(V \geq 1) \iff \mathrm{FWER} = P(V > 0) = 1 - P(V = 0)$$

Sobre los errores

Nos encontramos con la necesidad de controlar el error tipo I en el caso múltiple. Para esto, introducimos 3 nociones de error:

- **Tasa de Error por Comparación (PCER)**: Consiste de el valor esperado de errores de tipo I dividido entre el número total de hipótesis:

$$\mathrm{PCER} = \frac{\mathrm{E}(V)}{m}$$

- **Tasa de Error Global (FWER)**: Es la probabilidad de cometer uno o más errores de tipo I:

$$\mathrm{FWER} = P(V \geq 1) \Longleftrightarrow \mathrm{FWER} = P(V > 0) = 1 - P(V = 0)$$

- **Tasa de Falsos Descubrimientos (FDR)**. se define como el valor esperado de Q:

$$Q = \begin{cases} V/R, & R > 0 \\ 0, & R = 0 \end{cases}$$

Controlando el FWER

Control del FWER

Definición

Si suponemos que FWER$\leq \alpha$, decimos que la probabilidad de cometer un error tipo I está controlada por un nivel α. Adicionalmente:

- *Un proceso controla el FWER débilmente, si el control del FWER a un nivel α, es garantizado sólo cuando todas las hipótesis nulas son ciertas.*
- *Un procedimiento controla fuertemente, si el control del FWER a un nivel α independientemente de la configuración de hipótesis falsas o verdaderas.*

Procedimientos de Control del FWER - Bonferroni

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.

Procedimientos de Control del FWER - Bonferroni

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.

Procedimientos de Control del FWER - Bonferroni

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.
- Procedemos a rechazar la hipótesis H_{0i}, si $p_i \leq \frac{\alpha}{m}$

Procedimientos de Control del FWER - Bonferroni

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.
- Procedemos a rechazar la hipótesis H_{0i}, si $p_i \leq \frac{\alpha}{m}$
- El control puede ser probado a través de la *desigualdad de Boole:*

$$\mathrm{FWER} = P\left\{\bigcup_{i=1}^{m_0}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \sum_{i=1}^{m_0}\left\{P\left(p_i \leq \frac{\alpha}{m}\right)\right\} = m_0\frac{\alpha}{m} \leq m\frac{\alpha}{m} = \alpha.$$

Example

En el estudio de un taller, se obtuvo un conjunto de datos para determinar si la proporción de artículos defectuosos producidos por los trabajadores era la misma durante el dia, la tarde o la noche. Se encontraron los siguientes datos:

	TURNO			
Estado artículo	Día	Tarde	Noche	Total
Defectuosos	45	55	70	170
No defectuosos	905	890	870	2665
Total	950	945	940	2835

Usamos un nivel de significancia de 5 % para determinar si la proporción de artículos defectuosos es la misma para los tres turnos.:

Procedimientos de Control del FWER - Bonferroni

Example (Continuación)

- Planteamiento de hipótesis:

$$\mathrm{H}_0 : \pi_D = \pi_T = \pi_N$$
$$\mathrm{H}_1 : \text{las proporciones poblacionales no son todas iguales}$$

- Establecer el nivel de significancia: $\alpha = 5\,\% \rightarrow$ error tipo *I*
- El estadístico de prueba ji-cuadrada que se utiliza para este tipo de prueba de hipótesis, corresponde a la expresión:

$$\chi_p^2 = \sum \frac{(f_o - f_e)^2}{f_e},$$

Procedimientos de Control del FWER - Bonferroni

Example (Continuación de la continuación)

Así:

	TURNO (f_o)			
Estado artículo	Día	Tarde	Noche	Total
Defectuosos	45	55	70	170
No defectuosos	905	890	870	2665
Total	950	945	940	2835

	TURNO (f_e)			
Estado artículo	Día	Tarde	Noche	Total
Defectuosos	$(170 \cdot 950)/2835$ 56.9	$(170 \cdot 945)/2835$ 56.6	$(170 \cdot 940)/2835$ 56.36	170
No defectuosos	$(2665 \cdot 950)/2835$ 893.03	$(2665 \cdot 945)/2835$ 888.33	$(2665 \dot{} 940)\,/2835$ 883.63	2665
Total	950	945	940	2835

$$\chi_p^2 = \left[\frac{(45 - 56{,}9)^2}{56{,}9} + \cdots + \frac{(870 - 883{,}63)^2}{883{,}63}\right] = 6{,}29 \text{ y,}$$

$$\chi_{cr}^2 = \chi^2_{((2-1)(3-1),0{,}05)} = 5{,}98$$

Procedimientos de Control del FWER - Bonferroni

Example (Continuación de la continuación - again)

VI Regla de decisión: rechazamos H_0, si $\chi^2_p > \chi^2_{cr}$.

VII Conclusión: rechazamos H_0.

¿Y ahora? Veamos cuáles proporciones son distintas.

Procedimientos de Control del FWER - Bonferroni

Example (Continuación de la continuación - again x2)

- Prueba 1:

$$H_0 : \pi_D = \pi_T$$
$$H_1 : \pi_D \neq \pi_T$$

- Prueba 2:

$$H_0 : \pi_D = \pi_N$$
$$H_1 : \pi_D \neq \pi_N$$

- Prueba 3:

$$H_0 : \pi_T = \pi_N$$
$$H_1 : \pi_T \neq \pi_N$$

Example (Continuación de la continuación - final :D)

Luego al aplicar la corrección de Bonferroni, se obtiene $\alpha^* = 1{,}7\,\%$. Al usar un paquete estadístico (por ejemplo `R`), se obtuvo los siguientes *p*-valores para cada par:

Comparación	*p*-valor	Conclusión
Día vs. Tarde	0.8563	AH_0
Día vs. Noche	0.0032	RH_0
Tarde vs. Noche	0.0041	RH_0

Interpretación: la proporción de defectos es similar entre el turno del día y el de la tarde. El turno de la noche difiere significativamente en la proporción de defectos de los demás turnos.

Procedimientos de Control del FWER - Šidák

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.

Procedimientos de Control del FWER - Šidák

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.

Procedimientos de Control del FWER - Šidák

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.
- Procedemos a rechazar la hipótesis H_{0i}, si $p_i \leq \alpha_{SID} = 1 - (1 - \alpha)^{\frac{1}{m}}$

Procedimientos de Control del FWER - Šidák

- Sea $\{H_{01}, \ldots, H_{0m}\}$ una familia de hipótesis.
- Sea p_i el p-valor correspondiente a la hipótesis H_{0i}.
- Procedemos a rechazar la hipótesis H_{0i}, si $p_i \leq \alpha_{SID} = 1 - (1 - \alpha)^{\frac{1}{m}}$
- Por ejemplo, para $\alpha = 0.05$ y $m = 10$, el nivel de ajuste por Bonferroni es 0.005 mientras que el ajuste de Šidák es 0.005116 aproximadamente.

Procedimientos de Control del FWER - Holm-Bonferroni

- Supongamos que tenemos m p-valores, ordenados de menor a mayor: $p_{(1)}, \ldots, p_{(m)}$,y sus respectivas hipótesis: $H_{01}, \ldots, H_{0m}$

Procedimientos de Control del FWER - Holm-Bonferroni

- Supongamos que tenemos m p-valores, ordenados de menor a mayor: $p_{(1)}, \ldots, p_{(m)}$,y sus respectivas hipótesis: $H_{01}, \ldots, H_{0m}$
- Evaluamos si $p_{(1)} \leq \frac{\alpha}{m}$. Si **sí**, rechazamos H_{01} y continuamos con el siguiente paso. Si **no**, paramos.

Procedimientos de Control del FWER - Holm-Bonferroni

- Supongamos que tenemos m p-valores, ordenados de menor a mayor: $p_{(1)}, \ldots, p_{(m)}$,y sus respectivas hipótesis: $H_{01}, \ldots, H_{0m}$
- Evaluamos si $p_{(1)} \leq \frac{\alpha}{m}$. Si **sí**, rechazamos H_{01} y continuamos con el siguiente paso. Si **no**, paramos.
- Evaluamos si $p_{(2)} \leq \frac{\alpha}{m-1}$. Si **sí**, rechazamos H_{02} y continuamos con el siguiente paso. Si **no**, paramos.

Procedimientos de Control del FWER - Holm-Bonferroni

- Supongamos que tenemos m p-valores, ordenados de menor a mayor: $p_{(1)}, \ldots, p_{(m)}$,y sus respectivas hipótesis: $H_{01}, \ldots, H_{0m}$
- Evaluamos si $p_{(1)} \leq \frac{\alpha}{m}$. Si **sí**, rechazamos H_{01} y continuamos con el siguiente paso. Si **no**, paramos.
- Evaluamos si $p_{(2)} \leq \frac{\alpha}{m-1}$. Si **sí**, rechazamos H_{02} y continuamos con el siguiente paso. Si **no**, paramos.
- Así sucesivamente: para cada p-valor, verificamos si $p_{(k)} < \frac{\alpha}{m+1-k}$. Si **sí**, rechazamos H_{0k} y continuamos con p-valores más grandes. Si **no**, paramos.

False Discovery Rate

Historia e Importancia

- La tasa de falsos descubrimiento (FDR) fue propuesto por primera vez por Benjamini and Hochberg (1995), la cual se colocó como una de las piezas clave en la investigación relacionada con FDR.

Historia e Importancia

- La tasa de falsos descubrimiento (FDR) fue propuesto por primera vez por Benjamini and Hochberg (1995), la cual se colocó como una de las piezas clave en la investigación relacionada con FDR.
- Hasta antes de dicha publicación, la mayor parte de la inferencia relacionada con PHM se hacía fundamentalmente con base en métodos relacionados con el control de FWER, o bien técnicas similares derivadas de modificaciones a la misma.

Definición

Si definimos la variable aleatoria Q como:

$$Q = \begin{cases} V/R & R > 0 \\ 0 & R = 0 \end{cases} \tag{3}$$

Entonces, se tiene que

$$FDR = \mathbb{E}(Q) = \mathbb{E}\left(\frac{V}{V+S}\right) = \mathbb{E}\left(\frac{V}{R}\right).$$

Por como esta definido Q (3), tenemos que

$$\begin{aligned} FDR = \mathbb{E}\left(\frac{V}{R}\right) &= \mathbb{E}\left(\frac{V}{R}|R > 0\right)(R > 0) + \mathbb{E}\left(\frac{V}{R}|R = 0\right)(R = 0) \\ &= \mathbb{E}\left(\frac{V}{R}|R > 0\right)(R > 0). \end{aligned} \tag{4}$$

Relación de FDR y FWER.

1. Si todas las hipótesis nulas son verdaderas, entonces controlar la FDR es equivalente a controlar la FWER.

1. Si todas las hipótesis nulas son verdaderas, entonces controlar la FDR es equivalente a controlar la FWER.

   Demostración. Si todas las hipótesis nulas son verdaderas entonces $V = R$. Si $V = 0$ entonces $\frac{V}{R} = 0$ y si $V > 0$ entonces $\frac{V}{R} = 1$, por lo que (utilizando el resultado de (4))

$$
\begin{aligned}
FDR &= \mathbb{E}\left(\frac{V}{R}|V=0\right)(V=0) + \mathbb{E}\left(\frac{V}{R}|V>0\right)(V>0) \\
&= 0 \times (V=0) + 1 \times (V \geq 1) \\
&= (V \geq 1) \\
&= FWER.
\end{aligned}
$$

2. Si controlamos el FWER (es decir, lo mantenemos por debajo de algún valor q^*), entonces estamos controlando también FDR.

2. Si controlamos el FWER (es decir, lo mantenemos por debajo de algún valor q^*), entonces estamos controlando también FDR.

Demostración. Si no todas las hipótesis nulas son verdaderas, entonces $V < R$ y $\frac{V}{R} < 1$, y esto implica que $\mathbb{E}\left(\frac{V}{R}|V \geq 1\right) < 1$. Ocupando lo anterior tenemos

$$
\begin{aligned}
FDR &= \mathbb{E}\left(\frac{V}{R}|V = 0\right)(V = 0) + \mathbb{E}\left(\frac{V}{R}|V \geq 1\right)(V \geq 1) \\
&= 0 \times (V = 0) + \mathbb{E}\left(\frac{V}{R}|V \geq 1\right)(V \geq 1) \\
&< FWER.
\end{aligned}
$$

- Decimos por tanto que los procedimientos para el control FWER resultan ser más conservativos, en el sentido de que rechazan en promedio un menor número de hipótesis, que los procedimientos que controlan a FDR.

Procemiento de control FDR de Benjamini y Hochberg(BH)

Considere la pruebas $H_1, H_2, ..., H_m$, basado en los p–values correspondientes $P_1, P_2, ..., P_m$. Entonces primero ordenemos los p–values, es decir, $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}$ denotar por la hipótesis nula $H_{(i)}$ correspondiente a $P_{(i)}$. Defina el siguiente procedimiento de prueba múltiple de B–H:

$$\begin{aligned} \text{sea k la } i \text{ más grande para la cual } P_{(i)} \leq \frac{i}{m} q*; \\ \text{luego rechaza todo } H_{(i)} = 1, 2, ..., k. \end{aligned} \tag{5}$$

Teorema

Para estadísticos de prueba independientes y para cualquier configuración de hipótesis nulas falsas, el procedimiento anterior controla el FDR en $q*$.

Demostración. El teorema se deriva del siguiente lema.

Lema

Para cualquier $0 < m_0 < m$ p-values independientes correspondientes a hipótesis nulas verdaderas, y para cualquier valor que puedan tomar los p-values de $m_1 = m - m_0$ correspondientes a las hipótesis nulas falsas, el procedimiento de prueba múltiple definido por el procedimiento(1) anterior satisface la desigualdad

$$\mathbb{E}(Q|P_{m_0+1} = p_1, \dots, P_m = p_{m_1}) \leq \frac{m_0}{m} q^* \tag{6}$$

Ahora, suponga que $m_1 = m - m_0$ algunas de las hipótesis son falsas. Cualquiera que sea la distribución conjunta de $P''_1, \cdots, P''_{m_1}$, que corresponde a estas falsas hipótesis, integrando la desigualdad (6) anterior obtenemos

$$\mathbb{E}(Q) < \frac{m_0}{m} q* < q*,$$

y el FDR está controlado.

Ejemplo para controlar la FDR con el procedimiento B-H

Neuhaus (1992) quiere probar los efectos de una nueva administración de carga frontal de $rt - PA$ vs el efecto clasico de administrar APSAC en la reducción de la mortalidad en el infarto miocardio.
El control de FDR es deseable ya que no se quiere concluir que el nuevo tratamiento sea mejor si es simplemente equivalente al tratamiento anterior en todos los aspectos (eventos cardíacos)

Transiciones

- Los p–values individuales que reportan para cada prueba de hipótesis son

Transiciones

- Los p−values individuales que reportan para cada prueba de hipótesis son

$$0{,}0001, 0{,}0004, 0{,}0019, 0{,}0095, 0{,}0201, 0{,}0278, 0{,}0298,$$
$$0{,}0344, 0{,}0459, 0{,}3240, 0{,}4262, 0{,}5719, 0{,}6528, 0{,}7590, 1{,}000.$$

Transiciones

- Los p–values individuales que reportan para cada prueba de hipótesis son

$$0{,}0001, 0{,}0004, 0{,}0019, 0{,}0095, 0{,}0201, 0{,}0278, 0{,}0298,$$
$$0{,}0344, 0{,}0459, 0{,}3240, 0{,}4262, 0{,}5719, 0{,}6528, 0{,}7590, 1{,}000.$$

- Y los autores concluyen que:

Transiciones

- Los p–values individuales que reportan para cada prueba de hipótesis son

$$0{,}0001, 0{,}0004, 0{,}0019, 0{,}0095, 0{,}0201, 0{,}0278, 0{,}0298,$$
$$0{,}0344, 0{,}0459, 0{,}3240, 0{,}4262, 0{,}5719, 0{,}6528, 0{,}7590, 1{,}000.$$

- Y los autores concluyen que:
  "*En comparación con el tratamiento con APSAC, a pesar de más reoclusiones tempranas, el curso clínico con el tratamiento con rt-PA es más favorable con menos complicaciones hemorrágicas y una tasa de mortalidad hospitalaria sustancialmente más baja, presumiblemente debido a una mejor permeabilidad temprana de la arteria relacionada con el infarto*".

Observación

En el estudio no se especifíca cual fue el estadístico de prueba que utilizaron.

Realizemos las metodología de Bonferroni y $B - H$ para controlar las diferentes tasas. Es decir, para controlar la FWER considerando $\alpha = 0{,}05$ rechazamos si

$$p_i < 0{,}05/15 = 0{,}0033.$$

Y para controlar la FDR con $q^* = 0{,}05$ y sea k la i más grande para el cual $p_{(i)} \leq \frac{i}{m} q^*$, entonces rechzamos todas las $H_{(i)} = 1, 2, \cdots, k$.

H_{0i}	p–values	Umb. BH	Umb. Bonferr.	R BH	R Bonferr.
1	0.0001	0.0034	0.0033	TRUE	TRUE
2	0.0004	0.0067	0.0033	TRUE	TRUE
3	0.0019	0.0100	0.0033	TRUE	TRUE
4	0.0095	0.0134	0.0033	TRUE	FALSE
5	0.0201	0.0167	0.0033	FALSE	FALSE
6	0.0278	0.0200	0.0033	FALSE	FALSE
7	0.0298	0.0234	0.0033	FALSE	FALSE
14	0.7590	0.0467	0.0033	FALSE	FALSE
15	1.0000	0.0500	0.0033	FALSE	FALSE

Cuadro 1: Resultados de los p–values.

- Las primeras tres hipótesis corresponden a una reacción alérgica reducida y a dos aspectos diferentes del sangrado; no incluyen la comparación de mortalidad. Por tanto, la afirmación sobre una reducción significativa de la mortalidad no está justificada desde el punto de vista clásico. Pero controlando el FDR rechazamos la hipótesis 4.

- Las primeras tres hipótesis corresponden a una reacción alérgica reducida y a dos aspectos diferentes del sangrado; no incluyen la comparación de mortalidad. Por tanto, la afirmación sobre una reducción significativa de la mortalidad no está justificada desde el punto de vista clásico. Pero controlando el FDR rechazamos la hipótesis 4.
- En la figura (**??**) observamos los ümbrales"para rechazar la hipótesis nula para las distintas metodologías. Se observa claramente que las metodologías para controlar el *FWER* resultan más conservadoras ya que tiene un umbral muy pequeño para rechazo, en cambio si se controla el *FDR* el umbral es 0.5.

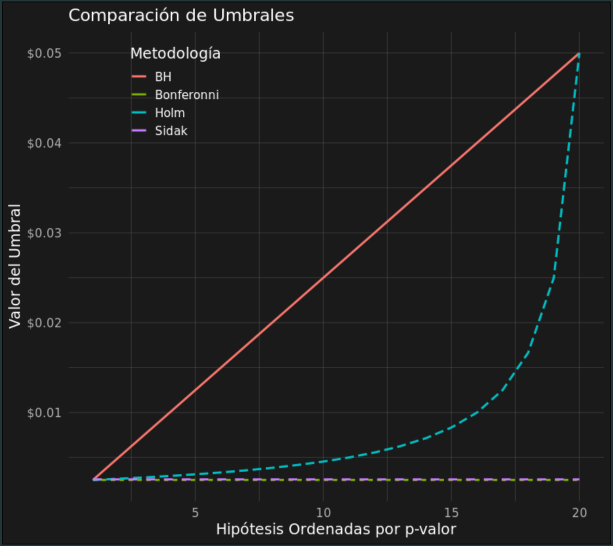


Figura 1: Comparación del valor del Umbral de rechazo

MCP - Ciencia de datos.

Aplicación del control FDR en neurociencia

- En este ejemplo, se realizó un análisis de las masas en imágenes de resonancia magnética (RM) extraídas de la iniciativa de datos abiertos de la serie de acceso abierto de estudios de imágenes (OASIS). Estos datos contienen imágenes ponderadas en T1 de 416 participantes con y sin demencia de entre 18 y 96 años, lo que permite investigar cómo la edad y las enfermedades relacionadas **con la edad influyen en la morfología cerebral**.

Aplicación del control FDR en neurociencia

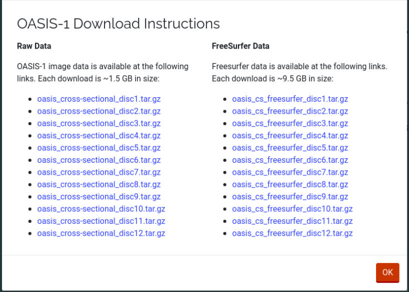
- En este ejemplo, se realizó un análisis de las masas en imágenes de resonancia magnética (RM) extraídas de la iniciativa de datos abiertos de la serie de acceso abierto de estudios de imágenes (OASIS). Estos datos contienen imágenes ponderadas en T1 de 416 participantes con y sin demencia de entre 18 y 96 años, lo que permite investigar cómo la edad y las enfermedades relacionadas **con la edad influyen en la morfología cerebral**.
- El objetivo del estudio fue determinar que partes del cerebro están relacionadas de su edad.

Metodología

- Para descargar el conjunto de datos de OASIS, visite *http : //www.oasis − brains.org/#data* y elija el lanzamiento OASIS-1, que contiene imágenes de resonancia magnética (MR) de 416 participantes de entre 18 y 96 años. El archivo de información del participante incluye variables demográficas básicas (edad, género, mano de obra, nivel educativo, estatus socioeconómico), variables clínicas y estimaciones de volumen cerebral.

Metodología

- Para descargar el conjunto de datos de OASIS, visite $http://www.oasis-brains.org/\#data$ y elija el lanzamiento OASIS-1, que contiene imágenes de resonancia magnética (MR) de 416 participantes de entre 18 y 96 años. El archivo de información del participante incluye variables demográficas básicas (edad, género, mano de obra, nivel educativo, estatus socioeconómico), variables clínicas y estimaciones de volumen cerebral.



Se considero calcular los coeficientes de correlación Spearman entre el grosor cortical y la edad en cada vóxel cortical, y se consideraron 163810 pruebas de la siguiente forma

$$H_{0i} : \rho_s = 0 \quad vs \quad H_{0i} : \rho_s \neq 0, \quad \forall\, i = 1, \cdots, 163810.$$

donde $\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$ y $d_i = rg(X_i) - rg(Y_i)$ es la diferencia de los rangos de cada observación. Se puede probar que si n es grande entonces

$$\rho_s \sqrt{\frac{n-2}{1-\rho_s^2}} \sim t_{n-2}$$

- Considerando lo anterior procedió a calcular los p–values de cada juego de hipótesis. Y posterior se le aplicaron las correcciones de Sidák para controlar el family–wise error rate (FWER), y el procedimiento BH para controlar la FDR.

- Considerando lo anterior procedió a calcular los p–values de cada juego de hipótesis. Y posterior se le aplicaron las correcciones de Sidák para controlar el family–wise error rate (FWER), y el procedimiento BH para controlar la FDR.
- Un total del 55 % de los vóxel mostró un efecto estadísticamente significativo del envejecimiento sobre el grosor cortical usando la corrección de Sidak. Por el contrario, cuando la corrección se basó en el procedimiento Benjamini-Hochberg, el número de vértices significativos fue del 82 %, lo que sugiere un efecto más generalizado. El nivel crítico sin corregir se fijó en $\alpha = 0{,}05$ en ambos análisis.

Comparación

Los resultados se muestran en la figura (**??**), lo cuál podemos concluir que algunos de los mayores efectos relacionados con la edad se observan en el surco central.

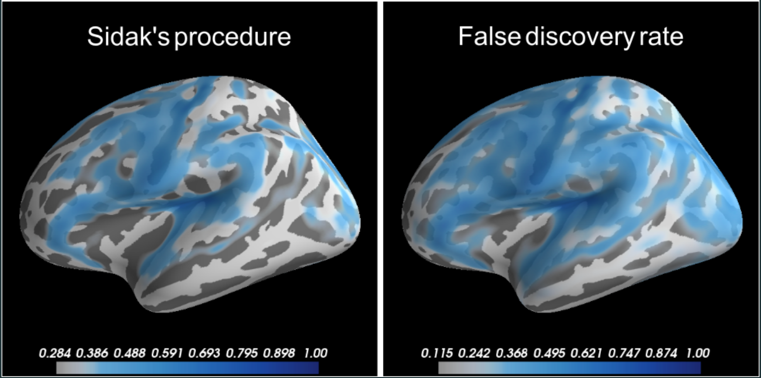


Figura 2: Diferencias cuando se controla el FWER y el FDR.

- Y. Benjamini and Y. Gavrilov, A Simple Forward Selection Procedure Based On False Discovery Rate Control(2009).

Artículos interesantes de ejemplos...

- Y. Benjamini and Y. Gavrilov, A Simple Forward Selection Procedure Based On False Discovery Rate Control(2009).
- Christopher J. Miller (1) et. al., Controling the False-Discovery Rate in Astrophysical Data Analysis(2001). '*El propósito de este artículo es presentar el procedimiento FDR a la comunidad astrofísica. Ilustramos el poder de FDR a través de varios ejemplos astronómicos, incluida la detección de características frente a una función unidimensional suave, por ejemplo, ver las ondulaciones de bariones en un espectro de potencia de fluctuaciones de materia y detección de píxeles de origen en datos de imágenes...*'

Conclusiones

- Se presentaron las bases teóricas sobre las pruebas de hipótesis múltiples desde el enfoque clásico controlando la *FWER* y la revolucionaria idea que propusieron Benjamini y Hochberh [1995] sobre controlar la *FDR*.

Conclusiones

- Se presentaron las bases teóricas sobre las pruebas de hipótesis múltiples desde el enfoque clásico controlando la *FWER* y la revolucionaria idea que propusieron Benjamini y Hochberh [1995] sobre controlar la *FDR*.
- La teórica presentada en este documento como se mencionó son las bases de *MCP*, en la actualidad ya existen una gran variedad de metodologías para controlar *FDR* y *FWER* la mayoría es son modificaciones del método de Bonferroní y *BH*.

Gracias ☻

Referencias i

Referencia *Javeriana*, 2020.