Otras alternativas para comparaciones múltiples: prueba de Scheffé y False Discovery Rate.

December 7, 2022

Gutiérrez Salazar Carlos Cuaúhtemoc

López Salomón María Guadalupe

CIMAT

Objetivos y Motivación

Las pruebas múltiples pueden ser usadas como una parte fundamental de procedimientos estadísticos como: selección de variables, familia de modelos matemáticos que intentan explicar la relación entre rasgos latentes, características y rasgos unobservables y sus manifestaciones y resultados observados (Item response theory), modelos de ecuaciones estructurales, árboles de decisión, etc.

Son diversas las aplicaciones que han surgido recientemente planteando nuevos tipos de problemas de multiplicidad y estimulando así, el desarrollo en pruebas de hipótesis múltiple.

Se pretende revisar diferentes técnicas de comparación múltiple para identificar qué tratamientos son estadísticamente diferentes, así como la tasa de descubrimiento falso (FDR) como método para conceptualizar la tasa de errores tipo I en las pruebas de hipótesis nulas al realizar comparaciones múltiples.

Introducción

Las comparaciones múltiples, o el problema de pruebas múltiples ocurre cuando se considera un conjunto de inferencias estadísticas simultáneamente o se infiere un subconjunto de parámetros seleccionados basados en los valores observados. Entre más observaciones se realicen, más probable es que se produzcan inferencias erróneas. Diversas técnicas estadísticas se han desarrollado para dirigirse a este problema, típicamente al requerir o exigir un umbral de significancia más estricto para las comparaciones individuales, de manera que se compense el número de inferencias que se realizan. La corrección de Bonferroni es el método para contrarrestar el problema de comparaciones múltiples; sin embargo, es un método conservativo que proporciona un mayor riesgo de fallo para rechazar una falsa hipótesis nula más que otros métodos, mientras ignoras información potencialmente valiosa, tal como la distribución de los p-valores a través de todas las comparaciones (que, si la hipótesis nula es correcta para todas las comparaciones, se espera que tome la forma de una distribución uniforme). Cabe destacar, que la corrección de Bonferroni controla la tasa de error familiar (relacionada a que la hipótesis nula sea verdadera para todas las comparaciones simultáneamente y alternativamente que la hipótesis nula sea falsa para al menos una prueba. Enfoques alternativos, tales como el procedimiento de Benjamini-Hochberg, son explícitamente diseñados para controlar la tasa de error cuando se detectan "descubrimientos individuales", y son por lo tanto más apropiados en muchos escenarios, tales como la detección de genes diferencialmente expresados en bioinformática. Este método fue nombrado después de Carlo Bonferroni.

Por otra parte, con un enfoque menos conservador y posiblemente más apropiado para identificar los pocos efectos importantes de los muchos efectos triviales probados tenemos el concepto de FDR (*false discovery rate)*, este concepto fue formalmente descrito por Yoav Benjamini y Yosef Hochberg en 1955 y ha sido particularmente influyente, dado que fue la primera alternativa al FWER para ganar amplia aceptación en muchos campos científicos (especialmente en ciencias de la vida, desde genética hasta bioquímica, oncología y ciencias de las plantas. En 2005, el paper de Benjamini y Hochberg de 1995 fue identificado como uno de los 25 papers estadísticos más citados. Previamente a la introducción de 1995 del concepto FDR, varios precursores de ideas han sido considerados en la literatura de estadística. En 1979, Holm propuso el procedimiento de Holm, un algoritmo paso a paso para controlar el FWER que es al menos tan poderoso como el bien conocido ajuste de Bonferroni. Este amplio algoritmo ordena los p-valores y secuencialmente rechaza las hipótesis empezando desde el más pequeño de los p-valores.

El trabajo de Benjamini y Hochberg (1995) tuvo su orígen en dos papers relacionados con las

pruebas múltiples: el primero por Schweder y Ppjotvoll (1982) en donde se sugirió graficar los p-valores clasificados y evaluar el número de hipótesis nulas verdaderas (m_0) a través de una línea ajustada a los ojos comenzando de los p-valores más grandes. Los p-valores que se desvían de esta línea recta deberían entonces corresponder a la falsa hipótesis nula. Esta idea fue posteriormente desarrollada en un algoritmo e incorporaron la estimación de (m_0) en procedimientos tales como Bonferroni, Holm o Hochberg. Esta idea está cercanamente relacionada a la interpretación gráfica del procedimiento BH. El segundo paper por Branko Soric (1989), introdujo la terminología de "descubrimiento" en el contexto de las pruebas de hipótesis múltiples. Soric usó el número esperado de falsos descubrimientos divididos por el número de descubrimientos $[V]/R$ como una advertencia que "por una larga parte de descubrimientos estadísticos podrían estar equivocados". Esto condujo a Benjamini y Hochberg a la idea de que una tasa de error similar, en lugar de ser simplemente una advertencia, lo cual puede servir como un objetivo digno de controlar. El procedimiento BH fue probado en controlar el FDR para pruebas independientes en 1995 por Benjamini y Hochberg.

Contenido

Antecedentes

Muchos experimentos científicos se ven sujetos a pruebas estadísticas y análisis que involucran la evaluación simultánea de más de una pregunta. Por ejemplo, si para una prueba médica, se puede comparar más de un grupo de tratamiento con un grupo de control y valorar sobre diferentes variables medidas como resultado de las pruebas, sean sus signos vitales a lo largo de un tiempo (temperatura corporal, la frecuencia del pulso, frecuencia de la respiración, presión arterial), o información sobre sus órganos, como el corazón, los riñones y el hígado a través de una prueba de sangre; estas pruebas múltiples dan pie a los problemas de multiplicidad que ocurren si queremos hacer múltiples preguntas a través de inferencias simultáneas. Muy probablemente la multiplicidad es sumamente importante cuando se requiere evidencia contundente y tomar buenas decisiones. En problemas de prueba de hipótesis que involucran hipótesis nulas únicas las pruebas estadísticas son frecuentemente elegidas para controlar la tasa de error de Tipo I de rechazar incorrectamente H en un nivel de significancia α pre-establecido. Si existe un gran número de preguntas experimentales y no hay ajuste de multiplicidad, se caerá en el error de Tipo I casi de manera segura concluyendo por un efecto que parezca significativo incluso cuando no exista alguno. La multiplicidad es un criterio necesario cuando se requiere reproducibilidad de resultados.

De manera general, cuando probamos m hipótesis nulas usando diferentes estadísticos de prueba, la probabilidad de cometer al menos un error de Tipo I es $1-(1-\alpha)^m$. Introduciremos algunos conceptos generales como procedimientos prueba *single-step* y *stepwise,* así como *p-values* ajustados y desajustados, entre otros, con el objetivo de revisar posteriormente algunos procedimientos de comparación múltiple.

Tasas de error y conceptos generales.

Para cualquier problema de prueba, existen tres tipo de errores. Una decisión falsa positiva ocurre cuando declaramos un efecto cuando ninguno existe. Similarmente, una decisión falsa negativa ocurre cuando fallamos en declarar un efecto que existe realmente.

En los problemas de pruebas de hipótesis este tipo de errores se denotan como: *Tipo I, Tipo II* respectivamente.

TABLE 1
Number of errors committed when testing m null hypotheses

	Declared non-significant	*Declared significant*	*Total*
True null hypotheses	**U**	**V**	m_0
Non-true null hypotheses	**T**	**S**	$m - m_0$
	$m - \mathbf{R}$	**R**	m

Figure 1: Tipo I y II de errores en pruebas de hipótesis multiples

De la figura 1 notamos que las hipótesis declaradas como *no-significativas* son las hipótesis no rechazadas mientras que las *declaradas significativas* son las rechazadas.

Revisaremos algunas definiciones de tasas de error comunmente usadas para los procedimientos de comparación múltiple.

Tasa de error por comparación: es la proporción esperada de errores Tipo I entre m decisiones. Si cada una de las m hipótesis se prueba por separado en un nivel de significancia α pre-especificado, luego $PCER = \alpha m_0/m \leq \alpha$.

En diversas aplicaciones, sin embargo, controlar la tasa de error por comparación es inadecuado por lo que en su lugar se consideran las hipótesis $H_1, ..., H_m$ como una familia conjunta donde un solo error de Tipo I conduce a una decisión incorrecta. Lo anterior motiva a definir la tasa de error *familywise* o *familywise error rate* (FWER):

$$FWER = \mathbb{P}(V > 0) = \mathbb{P}(cometer\ \ al\ \ menos\ \ un\ \ error\ \ Tipo\ \ I)$$

La ecuación anterior se define como *"familywise error rate"* o tasa de error de familia, o tambien llamado *"experimentwise error rate"*. De manera visual tenemos que:

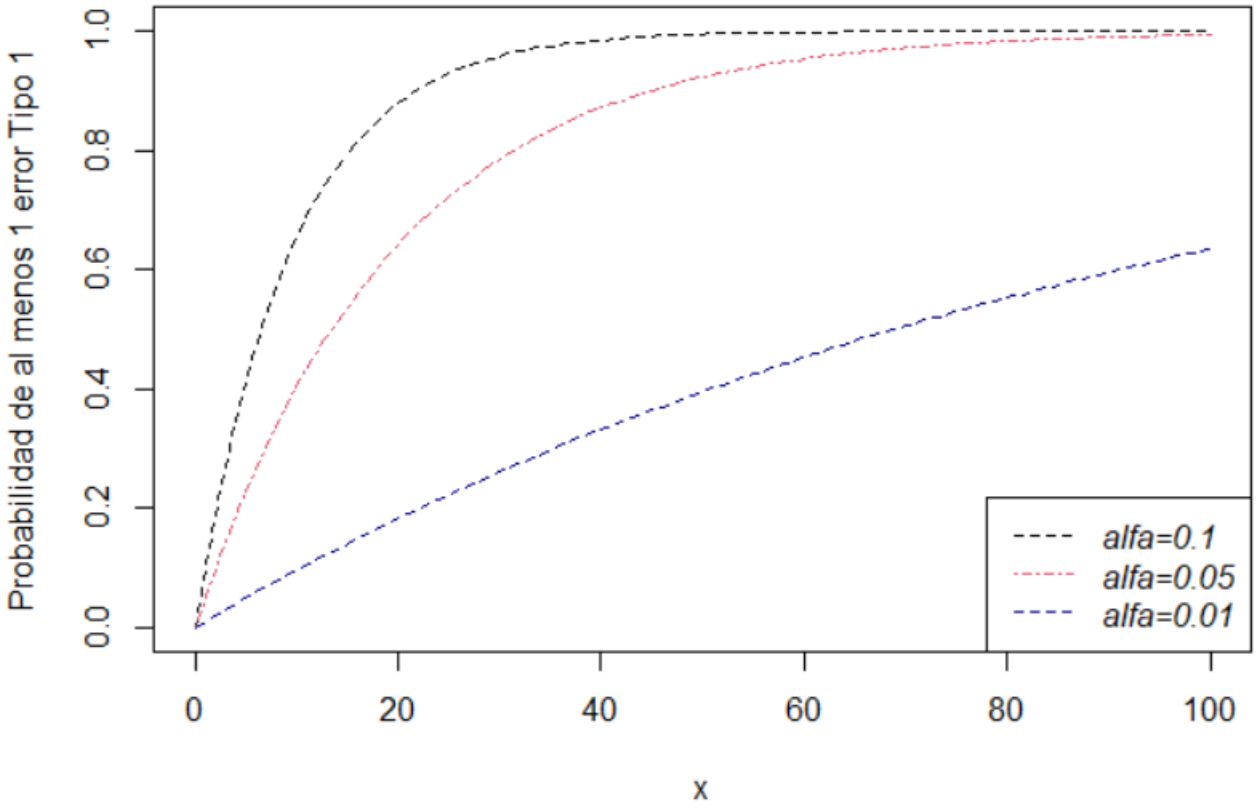

Figure 2: Probabilidad de al menos un error de tipo I

FWER es la tasa de error más común usada en pruebas múltiples, particularmente cuando el número de hipótesis m es muy grande y/o cuando evidencia contundente no se requiere (como en el caso de estudios de proyección de alta dimensión en biología molecular o descubrimiento de drogas) el uso de control de la tasa de error FWER no es apropiada pues resulta muy estricta .

Por otra parte, los métodos que controlan FWER aseguran que la probabilidad de cometer cualquier error de Tipo I está limitada por α.

Un acercamiento alternativo es relacionar el número de falsos positivos V al número total de rechazos R. Sea $Q = V/R$, si $R > 0$ y $Q = 0$ en otro caso.

FDR

La tasa de descubrimiento falso (FDR por sus siglas en inglés) es un método para conceptualizar la tasa de errores de Tipo 1 en las pruebas de hipótesis nulas cuando se realizan comparaciones múltiples. Los procedimientos de control de FDR están diseñados para controlar el FDR, que es la proporción esperada de "descubrimientos" (hipótesis nulas rechazadas) que son falsas (rechazos incorrectos de la nula). De manera equivalente, el FDR representa la relación esperada entre el número de calificaciones positivas falsas (error tipo 1) y el número total de calificaciones positivas (rechazos del nulo).

$$FDR = \mathbb{E}(\mathrm{Q})$$

$$= \mathbb{E}(\frac{\mathrm{V}}{\mathrm{R}} \mid \mathrm{R} > 0)\mathbb{P}(R > 0) + 0 \cdot \mathbb{P}(R = 0)$$

$$= \mathbb{E}(\frac{\mathrm{V}}{\mathrm{R}} \mid \mathrm{R} > 0)\mathbb{P}(R > 0)$$

El número total de rechazos del nulo incluye tanto el número de falsos positivos (FP) como el de verdaderos positivos (TP).

Los procedimientos de control de FDR proporcionan **un control menos estricto de los errores de Tipo 1** en comparación con los procedimientos de control de la tasa de error familiar (*familywise error rate, FWER*) como la corrección de Bonferroni, que controlan la probabilidad de al menos un error de Tipo 1. Por lo tanto, los procedimientos de control de FDR tienen mayor poder, a costa de un mayor número de errores de Tipo 1.

Al hacer pruebas de hipótesis con una sola hipótesis nula, los estadísticos de prueba se escogen directamente para controlar el error Tipo 1 e incorrectamente rechazar la hipótesis nula con una significancia pre-establecida α.

Si queremos probar sobre múltiples hipótesis (m), de manera simultánea, es decir con pruebas “de un solo paso” (diferentes de las “stepwise” o de múltiples pasos), la probabilidad de declarar efectos significativos falsos incrementa con m. Por ejemplo, si tenemos 2 hipótesis nulas, ($m = 2$), entonces cada nivel de confianza se propaga sobre el siguiente, resultando en un valor mayor que el nivel de significancia inicial.

Esto se traduce, en pocas palabras a que si hay un número grande de pruebas y no hay ajustes por la multiplicidad, se cometerá un error Tipo 1 de forma casi segura.

Si por ejemplo, en un experimento, después de una prueba de varianza el estadístico F es rechaz-

ado, podemos comparar todos los "tratamientos" en pares y así hallar para cuáles pares encontramos diferencias.

Hacer caso omiso de la multiplicidad nos conduce a un efecto conocido como Data dredging (data snooping o p-hacking).

Figure 3: Data dredging

Buscar patrones en los datos es legítimo, sin embargo, aplicar una prueba estadística de significancia, o prueba de hipótesis, a los mismos datos de los que surge un patrón es incorrecto. Una forma de construir hipótesis y evitar incurrir en el 'data dredging' es realizar pruebas "post-hoc".

Los métodos particularmente útiles en el análisis de varianza y en la construcción de bandas de confianza simultáneas para regresiones que involucran funciones de base son el método de Scheffé y, si el investigador sólo tiene en mente comparaciones por pares, el método de Tukey. El uso de la tasa de descubrimiento falso de Benjamini-Hochberg es un enfoque más sofisticado que se ha convertido en un método popular para el control de múltiples pruebas de hipótesis.

Las revistas académicas cambian cada vez más al formato de informe registrado, que tiene como objetivo contrarrestar problemas muy serios como el data dredging y HARKing (hipotetizar después de que se hallaron los resultados), que han hecho que la investigación de prueba de teoría sea muy poco confiable.

Procedimientos single-step and stepwise.

Una posibilidad de clasificar los procedimientos de comparación múltiple es dividirlos en: *single-step* y *stepwise.*

- Procedimientos *single-step:* se caracterizan por el hecho de que el rechazo o no-rechazo de la

hipótesis nula no toma en cuenta la decisión de cualquier otra hipótesis, de manera que el orden en el cual se prueban las hipótesis no es importante y uno podría pensar en las múltiples inferencias que se realizan en un solo paso. Un ejemplo del procedimiento single-step es la prueba Bonferroni.

- Procedimiento *stepwise:* el rechazo o no rechazo de la hipótesis nula puede depender en la decisión de las otras hipótesis. El igualmente conocido, Holm, es un procedimiento stepwise y es una extensión de la prueba Bonferroni. Los procedimientos stepwise se dividen en: *step-down* y *step-up.* Ambos procedimientos asumen una secuencia de hipótesis ordenadas con datos dependientes En el caso step-down se comienza probando las primeras hipótesis H_1 ordenadas y se va bajando a través de la secuencia mientras se rechazan las hipótesis. El procedimiento se detiene en el primero no-rechazo y $H_1, ..., H_{i-1}$ son rechazadas. Un ejemplo de este procedimiento es la prueba Holm. Para el procedimiento step-up se comienza probando H_m y se va subiendo a través de la secuencia mientras se van reteniendo las hipótesis. El procedimiento se detiene en el primer rechazo y $H_1, ..., H_i$ son rechazadas.

En general los procedimientos single-step son menos poderoso que sus extensiones stepwise en el sentido de que cualquier hipótesis rechazada por el anterior también será rechazada por el último pero no viceversa. La ventaja poderosa de los procedimientos stepwise, sin embargo, radia en el costo de dificultades incrementadas al construir intervalos de confianza simultáneos compatibles para el parámetro de interés, los cuales tienen una cobertura conjunta de probabilidad de al menos $1 - \alpha$.

Tipos de control de error

Si $ER_m(P)$ es una tasa de error cualquiera, para una m dada en una distribución P para un conjunto de datos. El control de la tasa de error puede ser categorizado como:

- Control débil: Existe control débil si hay una muestra control finita de $ER_m(P_0)$, donde P_0 es una distribución que pudiera haber generado los datos si todas las hipótesis nulas fueran ciertas ("complete null")

- Control fuerte: Existe un control fuerte si hay una muestra control finita de $I \subseteq \{1, ..., m\}\, maxER_m(P_I)$ donde P_I es la distribución que generaría los datos si las pruebas de hipótesis nulas relacionadas a I fueran ciertos y el resto falso, esto quiere decir que la medida del error Tipo I está acotada por arriba sin importar la forma en la que estén las hipótesis nulas verdaderas o falsas.

- Control exacto: Hacemos notar que las únicas suposiciones para el control del FDR o FDX son usualmente relacionadas a la dependencia entre los estadísticos de prueba.

Table 1 Outcomes in testing m hypotheses

	H_0 not rejected	H_0 rejected	Total
H_0 True	$N_{0\|0}$	$N_{1\|0}$	M_0
H_0 False	$N_{0\|1}$	$N_{1\|1}$	M_1
Total	$m - R$	R	m

Figure 4: Resultados en m pruebas de hipótesis

Intervalos de confianza simultáneos.

De forma similar, si tenemos una familia de k parámetros $\{\theta_1, \theta_2, ..., \theta_k\}$, con un nivel de significancia α, los intervalos de confianza simultáneos son una familia de intervalos $(L_1, U_1), (L_2, U_2), ..., (L_k, U_k)$, en la cual cada intervalo tiene un nivel de significancia α, y para ajustar por multiplicidad necesitamos que cada intervalo sea lo suficientemente ancho y que los valores críticos sean mayores.

A continuación se mencionan algunos de los métodos para generar intervalos de confianza, en los cuales el ajuste a la creación de estos intervalos es interviniendo en la obtención de los **valores críticos.**

(Falta añadir historia de Bonferroni)

Método de Bonferroni (FWER)

Este método se trata básicamente de dividir αpara el valor crítico en k siendo k el número de formas en las que se parean los datos. Si hay m tratamientos, $C(m, 2)$ son las formas de parearlos. Dicho de otra manera, la nueva α a usar será de $\frac{\alpha}{C(m,2)}$.

$$t = \frac{\bar{y}_{i\bullet} - \bar{y}_{j\bullet}}{SE} > t_{N-g,\alpha/2/k}$$

$$|\bar{y}_{i\bullet} - \bar{y}_{j\bullet}| > SE \times t_{N-g,\alpha/2/k} \approx 2.9975 \times 3.38 \approx 10.13 = BSD.$$

La desigualdad anterior se conoce como Diferencia Significante de Bonferroni (BDS por sus siglas en inglés), y se dice que para cualquier par de tratamientos i, j son significativamente diferentes si su

diferencia es mayor a ese valor.

Cuando se aplica el método de Bonferroni, los p-valores marginales son esencialmente multiplicados por el número de hipótesis nulas a probar.

Limitaciones:

- No importa la dependencia de los $p - valores$
- El número de pruebas k debe ser finito.
- El método de Bonferroni funciona bien cuando el número de pruebas k es chico.
- Cuando el número de pruebas es grande (> 10), el método se vuelve muy conservador. El FWER puede ser mucho menor a α.

(Añadir por que funciona)

Proceso Tukey-Kramer para comparaciones en pares()

Para cualquier $i \neq j$, los intervalos de confianza simultaneos para $\mu_i - \mu_j$ estaán dados por:

$$\bar{y}_{i\bullet} - \bar{y}_{k\bullet} \pm \frac{q_\alpha(g, N-g)}{\sqrt{2}} SE(\bar{y}_{i\bullet} - \bar{y}_{k\bullet})$$

donde $SE(\bar{y}_{i\bullet} - \bar{y}_{k\bullet}) = \sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_k})}$.

Esto es porque

$$|t_0| = \frac{|\bar{y}_{i\bullet} - \bar{y}_{k\bullet}|}{\sqrt{MSE(\frac{1}{n} + \frac{1}{n})}} \leq \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{2MSE/n}} = \frac{q}{\sqrt{2}}$$

donde $q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{MSE/n}}$ y q tiene el rango de una $t - students$

$$\frac{q_\alpha(g, N-g)}{\sqrt{2}} \times \sqrt{MSE(\frac{1}{n} + \frac{1}{n})}$$

que es la Diferencia Significativa Honesta de Tukey (Tukey's HSD).

Método de Scheffé para comparar todos los contrastes

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

El método de Scheffe se aplica a un conjunto de estimaciones de todos los posibles contrastes entre los niveles de factores de las medias, es decir, prueba todos los posibles contrastes al mismo tiempo. Donde un contraste arbitrario se define por

$$C = \sum_{i=1}^{r} c_i \mu_i$$

donde

$$\sum_{i=1}^{r} c_i = 0$$

Otro problema de índole diferente, también ligado a las comparaciones múltiples, consiste en la comparación de distintos tratamientos con un control que suele emplearse en determinados campos de investigación, por ejemplo en estudios epidemiológicos.

Proceso Benjamini-Hotchberg

Sea V el número de hipótesis nulas rechazadas verdaderas y S el número de hipótesis nulas rechazadas falsas y R como su suma o el total de hipótesis nulas rechazadas, la proporción de errores (FDR) puede verse como la siguiente variable aleatoria.

$$Q = \frac{V}{V+S}$$

Claro que $V + S = 0$, entonces $Q = 0$, ya que no hay falsos rechazos.

Si consideramos las hipótesis nulas $H_1, H_2, ..., H_m$ de las cuales obtenemos los $p-$valores $P_1, P_2, ..., P_m$. Ordenamos estos valores de forma ascendente de manera que $P_{(1)} \leq P_{(2)} \leq ... \leq P_{(m)}$ con correspondientes hipótesis nulas $H_{(i)}$.

El procedimiento Benjamini-Hotchberg establece:

Si k es el mayor valor de i *t.q.* $P_k \leq \frac{k}{m}\alpha$ rechazamos todas las hipótesis nulas $H_{(i)}$ para $i = 1, ..., k$

((Para cualquier $i \neq j$, los intervalos de confianza simultáneos para $\mu_i - \mu_j$.))

(Falta conexion)

Tenemos que:

Para un α dado, se debe hallar el mayor k tal que $P_k \leq \frac{k}{m}\alpha$

Rechazar la hipótesis nula (i.e., declarar descubrimientos) para todos las $H_{(i)}$ para $i = 1, ..., k$

Este procedimiento es válido cuando las m pruebas son independientes

$$E(Q) \leq \frac{m_0}{m}\alpha \leq \alpha \tag{1}$$

PRUEBA

Por inducción matemática.

En el caso en el que $m = 1$, se cumple. Asumimos que el lema es verdadero para cualquier $m_I \leq m$ y mostramos que sirve para $m + 1$. Si $m_0 = 0$, todas hipótesis nulas son falsas, Q es 0 y

$$E(Q \mid P_1 = p_1, ..., P_m = p_m) = 0 \leq \frac{m_0}{m+1}q*$$

Si $m_0 > 0$, denotamos $P_i^{'}$, $i = 1, 2, ..., m_0$, los $p-$valores correspondientes a las hipótesis nulas verdaderas, y al más grande de estos como $P_{(m_0)^I}$. Estos son $U(0, 1)$ variables aleatorias independientes. Para simplificar la notación asumimos que m_1 $p-$valores que son hipótesis nulas falsas toman el orden $p_1 \leq p_2 \leq ... \leq p_{m_1}$. Definimos j_0 como el más grande j_{m_1} que satisface

$$p_j \leq \frac{m_0 + j}{m+1}q*$$

y denotamos el lado derecho de la ecuación anterior en j_0 por p''.

Condicionando $P_{(m_0)}^{'} = p$,

$E(Q \mid P_{m_0+1} = p_1, ..., P_m = p_{m_1}) = \int_0^{p''} E(Q \mid P_{(m_0)}^{'} = p, P_{m_0+1} = p_1, ..., P_m = p_{m_1}) f_{P_{(m_0)}}^{'}(p)dp+$

con $f_{P_{(m_0)}}(p) = m_0 p^{(m_0-1)}$. En la primera pp''. Entonces todas las hipótesis nulas $m_0 + j_0$ son rechazadas, y $Q \equiv m_0/(m_0 + j_0)$. Evaluando la integral primero y luego usando la desigualdad de p_j obtenemos

$$\frac{m_0}{m_0 + j_0}(p'')^{m_0}\frac{m_0}{m_0 + j_0}\frac{m_0 + j_0}{m+1}q * (p'')^{m_0-1} = \frac{m_0}{m+1}q * (p'')^{m_0-1}$$

En la segunda parte de la ecuación de integrales, consideramos de manera separada cada $p_{j_0} < p_j \leq P_{(m_0)}^{'} = p < p_{j+1}$, junto con $p_{j0} \leq p'' < P_{(m0)}^{'} = p < p_{j_0+1}$. Es importante notar que por la manera en la cual j_0 y p'' están definidos, no se rechazan hipótesis como resultado de los valores $p, p_{j+1}, p_{j+2}, ..., p_{m_1}$. Por lo tanto, cuando todas las hipótesis son ciertas y falsas se consideran juntas, y sus $p-$valores se ordenan, una hipótesi $H_{(i)}$ puede ser rechazada si solo si existe $k, ikm_0 + j - 1$, para el cual $p_{(k)} \{k/(m+1)\} q*$

$$\frac{p_{(k)}}{p} \leq \frac{k}{m_0 + j - 1}\frac{m_0 + j - 1}{(m+1)p}q* \tag{2}$$

Cuando condicionamos $P^{'}_{(m_0)} = p$, $P^{'}_i/p$ para $i = 1, 2, ..., m_0 - 1$ se distribuyen como $m_0 - 1$ variables aleatorias independientes $U(0, 1)$ y p_i/p para $i = 1, 2, ..., j$ son números que corresponden a hipótesis nula falsa entre 0 y 1. Usando la desigualdad (2) para probar $m_0 + j - 1 = m' \leq m$ hipótesis es equivalente a usar el procedimiento:

Si probamos $H_1, H_2, ..., H_m$ que se basa en los correspondientes $p-$valores $P_1, P_2, ..., P_m$. Sean $P_{(1)} \leq P_{(2)} \leq ... \leq P_{(m)}$ $p-$valores ordenados y denotamos por $H_{(i)}$ la hipótesis nula correspondiente a $P(i)$. Definimos el siguiente procedimiento de Bonferroni múltiple: sea k el mayor $i$para el cual $P_{(i)} \leq \frac{i}{m} q*$: entonces se rechazan todas las $H_{(i)}$ $i = 1, 2, ..., k$.

De modo que con la constante $\{(m_0 + j - 1)/(m + 1)p\}q*$ tomando el rol de $q*$. Aplicando la hipótesis de inducción, tenemos

$$E(Q \mid P^{'}_{(m_0)} = p, P_{m_0+1} = p_1, ..., P_m = p_{m_1} \leq \frac{m_0 - 1}{m_0 + j - 1} \frac{m_0 + j - 1}{(m + 1)p} q* = \frac{m_0 - 1}{(m + 1)p} q* \tag{3}$$

La frontera en la desigualdad (2) depende de p, pero no en el segmento $p_j < p < p_{j+1}$ para el cual se evaluó, entonces:

$$\int_{p''}^{1} E(Q \mid P'_{(m_0)} = p, P_{m_0+1} = p_1, ..., P_m = p_{m_1}) f'_{P_{(m_0)}}(p)dp \leq \int_{p''}^{1} \frac{m_0 - 1}{(m + 1)p} q * m_0 p^{(m_0-1)} dp \tag{4}$$

$$= \frac{m_0}{m + 1} q * \int_{p''}^{1} (m_0 - 1) p^{(m_0-2)} dp = \frac{m_0}{m + 1} q * \{1 - p^{''(m_0-1)}\} \tag{5}$$

y considerando la desigualdad

$$\frac{m_0}{m_0 + j_0} (p^n)^{m_0} \leq \frac{m_0}{m_0 + j_0} \frac{m_0 + j_0}{m + 1} q * (p^n)^{m_0-1} = \frac{m_0}{m + 1} q * (p^n)^{m_0-1} \tag{6}$$

por (5) y (6) queda demostrado (1).

Conclusiones

dfdffdd

Bibliografía

- Benjamini, Yoav; D. Yekutieli. The control of the false discovery rate in multiple testing under dependency, Y., The Annals of Statistics, The Annals of Statistics, https://doi.org/10.1214/aos/1013699998, 2001/8/1

- Benjamini, Yoav; Hochberg, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B 57 (1995), no. 1, 289–300.

- Farcomeni A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. Stat Methods Med Res. 2008 Aug;17(4):347-88. doi: 10.1177/0962280206079046. Epub 2007 Aug 14. PMID: 17698936.

- Bretz, F., Hothorn, T. & Westfall, P. (2016). Multiple Comparisons Using R (English Edition) (1.a ed.). Chapman and Hall/CRC. Huang, Y. (s. f.). STAT22200 Linear Models And Experimental Designs. https://www.stat.uchicago.edu/%7Eyibiteaching/stat222/2021/

- Wikipedia contributors. (2022). False discovery rate. https://en.wikipedia.org/wiki/False_discovery_rate