

Antecedentes para comprender el artículo

Considerando el mismo contexto de las pruebas de hipótesis simples, se puede generalizar esta idea al considerar una colección de $m > 1$ hipótesis $H_{01}, H_{02}, \dots, H_{0m}$ donde el objetivo es realizar una prueba estadística simultánea para determinar cuántas y cuáles de ellas son rechazadas considerando un cierto nivel de confianza α . Dicho procedimiento recibe el nombre de **prueba de hipótesis múltiple** o **problema de comparaciones múltiples**.

Para extender la metodología para pruebas de hipótesis simples al caso múltiple se tienen que considerar algunos aspectos importantes. Supuesto de independencia: es posible que puedan existir al menos un par de índices i y j tales que el rechazo de H_{0i} podría influir (positiva o negativamente) en las posibilidades del rechazo de H_{0j} . Muchos de los procedimientos clásicos dentro de la metodología de comparaciones múltiples requieren el supuesto de independencia entre las hipótesis; sin embargo, existen diversas modificaciones que prueban ser bastante robustas para ciertas instancias de dependencia. Otro aspecto importante a considerar, y quizás el más importante, es que, al extender el escenario de las pruebas de hipótesis simples al caso múltiple se incorpora el efecto de la multiplicidad al análisis. Ejemplo:

Ejemplo: 1 *Supóngase que un experimentador desea probar estadísticamente si una moneda determinada es justa. Para ello realiza 10 lanzamientos, de los cuales 9 resultan en cara. Si asumimos como cierta la hipótesis de que la moneda es justa entonces la probabilidad de que se observe un resultado al menos tan extremo como ese sería de $(10 + 1)(\frac{1}{2})^{10} = 0,0107$, con lo que podemos concluir que no es razonable asumir que la moneda es justa con base en la información obtenida.*

¿Qué pasa si en lugar de una moneda fueran 100? ¿Cuál es la probabilidad de que en 100 experimentos con monedas justas, al menos una muestre 9 o más caras en 10 lanzamientos?

Otro ejemplo, si una prueba simple se hace a un 5 % de confianza, afirmamos que existe un 95 % de probabilidad de que la hipótesis nula sea rechazada incorrectamente. Sin embargo, si se realizan $m = 100$ pruebas de hipótesis simultáneamente, donde todas son ciertas, el número esperado de rechazos incorrectos es 5, mientras que, si las pruebas son independientes, la probabilidad de rechazar al menos una hipótesis incorrectamente es de $1 - (0,05)^{100} = 0,994$. En este contexto, el error de rechazar una hipótesis nula que es cierta se conoce comúnmente como falso positivo o error de tipo I como en el caso de las pruebas de hipótesis simples.

Lo anterior exhibe un aspecto importante de la teoría de las pruebas de hipótesis múltiples y la necesidad de nuevas formas de cuantificar el error, para considerando la multiplicidad.

Un algoritmo estándar consiste en dos pasos fundamentales,

1. Elegir y calcular un estadístico de prueba T_j para cada hipótesis individual j .
2. 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.

El Problema 1 está dentro del contexto de pruebas de hipótesis simples. Como se discutió a detalle en clases, la elección del estadístico de prueba adecuado T_j dependerá enteramente del contexto experimental

y del tipo de hipótesis de trabajo que se tenga. Y el problema 2, es parte de la teoría descrita en el artículo de Benjamini.

En el artículo se describe una tasa de error en particular, pero las más comunes son 3 (considerando la nomenclatura del artículo):

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

$$PCER = \mathbb{E}(V)/m.$$

La idea detrás del PCER yace en la intención de heredar, al menos en cierto sentido, el nivel de significancia α de las pruebas individuales. Para ver esto, supongamos, por ejemplo, que todas las hipótesis son ciertas y que se prueban individualmente a un nivel de significancia común α . Luego, V es una variable aleatoria cuya distribución es binomial con probabilidad de éxito dada por la probabilidad de rechazar una hipótesis cierta, que es precisamente α . Por tanto,

$$PCER = E(V)/m = m\alpha/m = \alpha.$$

En general, si m hipótesis son probadas a un nivel α de significancia, entonces el PCER sería siempre α . En otras palabras, PCER tiende a ignorar el efecto de la multiplicidad y por ello, por lo general no es muy recomendable.

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

$$FWER = \mathbb{P}(V \geq 1).$$

Lo describen bien en el link del Readme de la tesis.

- **Tasa de Falsos Descubrimientos (FDR)** Propuesto por primera vez en el artículo de Benjamini and Hochberg.

1. Artículo: Introducción

En esa época la metodología tradicional de las pruebas múltiples era solo considerar el control de FWER (y el PCER que equivale a ignorar o completo de la multiplicidad) para el efecto de multiplicidad. La metodología de FWER presentan algunos problemas: potencia, muy conservador. Por lo cuál, el objetivo del artículo se centra en proponer un nuevo punto de vista para el problema de multiplicidad. No hay mucho que decir, solo que existía una problemática con las MCP en ese entonces.

2. False Discovery Rate

La tasa de descubrimientos falsos (FDR) (es el valor esperado del número de errores de tipo I (V) que se cometen entre las hipótesis rechazadas (R). Es decir, $FDR = \mathbb{E}(Q)$ donde $Q = V/R$ si $R > 0$ y $Q = 0$ si $R = 0$. Menciona dos propiedades importantes del FDR y su relación con FWER, todas las m hipótesis son ciertas, implica que $Q = V/R$ sólo puede tomar dos posibles valores, $Q = 0$ si $V = 0$ o $Q = 1$ si $V > 0$ ya que todos los rechazos serán equivocados. Es decir, Q representará una variable aleatoria Bernoulli cuya probabilidad de éxito, y valor esperado, será igual a la probabilidad de que ocurra un rechazo equivocado cualquiera $P(V > 0)$, que es precisamente la definición de la FWER, es decir, $FDR = FWER$. Y además se cumple que

$$PCER \leq FWER \leq FDR.$$

(En el artículo solo lo prueban para FWER y FDR), lo anterior se puede interpretar: si mediante un procedimiento determinado controlamos a PCER a un nivel de significancia α , es decir, dicho procedimiento garantiza que $PCER \leq \alpha$, entonces habremos garantizado también el control de FWER y FDR al mismo nivel de significancia.

3. Procedimiento de control FDR

Considere la pruebas H_1, H_2, \dots, H_m , basado en los p-values correspondientes $P_{(1)}, P_{(2)}, \dots, P_{(m)}$. Sean $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$ los p-values están ordenados y denotar por la hipótesis nula $H_{(i)}$ correspondiente a $P_{(i)}$. Defina el siguiente procedimiento de prueba múltiple de Bonferroni-type:

$$\text{sea } k \text{ la } i \text{ más grande para la cual } P_{(i)} \leq \frac{i}{m}q^*; \quad (1)$$

$$\text{luego rechaza todo } H_{(i)} = 1, 2, \dots, k. \quad (2)$$

Teorema: 1 *Para estadísticas de prueba independientes y para cualquier configuración de hipótesis nulas falsas, el procedimiento anterior controla el FDR en q^* .*

Algo importante que resaltar es que esta metodología funciona para prueba independientes!!

Explica la relación de la metodología de Hochberg para contralar el *FWER*.

Posteriormente muestran un ejemplo en donde si se considera la metodología de Bonferroni para controlar el FWER con un $\alpha = 0,05$ se rechazan 3 hipótesis y si se considera nueva metodología se rechazan 4. Pues ser un buen ejemplo a mostrar.

4. Comparaciones de Potencia

Básicamente este capítulo es para observar como el procedimiento nuevo resulta ser una mejora considerable respecto del metodo de Bonferonni cuando se controla FWER en términos de potencia. Es decir, esta es otra gran ventaja de utilizar la nueva metodología.