Contrastes múltiples

17 de marzo de 2023

Definición 1. Llamaremos contraste múltiple a un contraste simultáneamente sobre varios parámetros.

Ejemplo 1. El análisis de varianza contrasta la hipótesis nula $H_0\colon \mu_1 = \cdots = \mu_I$ frente a $H_1\colon \exists\, i, j,\ \mu_i \neq \mu_j$, por lo que es un contraste múltiple.

Ejemplo 2. El contraste de homoscedasticidad de Bartlett contrasta $H_0\colon \sigma_1 = \cdots = \sigma_I$ frente a $H_1\colon \exists\, i, j,\ \sigma_i \neq \sigma_j$, luego también es un contraste múltiple.

La hipótesis nula de tales contrastes puede formularse como intersección de hipótesis nulas parciales.

Ejemplo 3. La hipótesis nula del contraste de Bartlett es $H_0 = \bigcap_{i,j} H_0^{ij}$ donde $H_0^{ij}\colon \sigma_i = \sigma_j$. Sería la intersección de $k = \binom{I}{2}$ hipótesis nulas.

La hipótesis nula global H_0 se rechazaría al rechazar alguna de las hipótesis parciales H_0^i, luego la región crítica global sería

$$\mathrm{RC} = \bigcup_{i=1}^{k} \mathrm{RC}_i \tag{1}$$

donde RC_i es la región crítica asociada al i-ésimo contraste parcial.

Para garantizar que

$$\Pr[\mathrm{RC} \mid H_0] \leqslant \alpha$$

se pueden considerar varios criterios.

1. Bonferroni

Aplicando la desigualdad de Boole,

$$\Pr\left(\bigcup_{i=1}^{k} A_i\right) \leqslant \sum_{i=1}^{k} \Pr A_i$$

a las regiones críticas de (1) y resolviendo cada contraste parcial a nivel de significación $\frac{\alpha}{k}$ se tiene

$$\Pr(\mathrm{RC} \mid H_0) = \Pr\left(\bigcup_{i=1}^{k} \mathrm{RC}_i \;\middle|\; H_0\right) \leqslant \sum_{i=1}^{k} \Pr(\mathrm{RC}_i \mid H_0^i) \leqslant \sum_{i=1}^{k} \frac{\alpha}{k} = \alpha$$

Ejemplo 4. En un contraste para comparar la igualdad de tres medias,

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \mu_3 \\ H_1 &: \quad \mu_1 \neq \mu_2 \vee \mu_1 \neq \mu_3 \vee \mu 2 \neq \mu_3 \end{aligned}$$

para conseguir un nivel de significación global $\alpha = 0{,}05$ se realizaría cada contrastre $H_0^{ij} : \mu_i = \mu_j$, $H_1^{ij} : \mu_i \neq \mu_j$ a nivel de significación $\frac{\alpha}{3} = \frac{0{,}05}{3} = 0{,}01666...$

El método de Bonferroni es muy popular porque

- Es muy fácil de aplicar.
- Se puede usar siempre.

Su inconveniente fundamental es que el tamaño del contraste global puede estar muy alejado de α, con la consiguiente pérdida de potencia.

2. Šidák

Cuando las regiones críticas de los contrastes parciales se basan en estadísticos independientes, entonces

$$\alpha = \Pr(\mathrm{RC} \mid H_0) = 1 - \Pr(\mathrm{RA} \mid H_0) = 1 - \Pr\left(\bigcap_{i=1}^{k} \mathrm{RA}_i \;\middle|\; H_0 \right) =$$

$$= 1 - \prod_{i=1}^{k} \Pr(\mathrm{RA}_i \mid H_0^i) = 1 - \prod_{i=1}^{k} (1 - \alpha_i)$$

luego bastaría tomar como nivel de significación de cada contraste parcial

$$\alpha_i = 1 - \sqrt[k]{1 - \alpha}$$

Aunque esta corrección es menos conservadora que la de Bonferroni y, por tanto, más potente, suele serlo por poco.

Ejemplo 5.

$$\left. \begin{array}{l} \alpha = 0{,}05 \\ k = 10 \end{array} \right\} \Longrightarrow \left\{ \begin{array}{l} \text{Bonferroni } \alpha_i = 0{,}005 \\ \text{Šidák } \alpha_i = 0{,}005\,116 \end{array} \right.$$

3. Holm

Este criterio mejora la corrección de Bonferroni, pues

- También puede usarse siempre (no requiere independencia).
- En caso de rechazarse alguna hipótesis nula parcial, puede rechazar más hipótesis que el de Bonferroni.

Considérese la siguiente notación:

p_i p-valor del contraste H_0^i/H_1^i, con $i = 1, \ldots, k$.

$p_{(i)}$ p-valor ordenado, tal que $p_{(1)} \leqslant \ldots \leqslant p_{(k)}$.

$H_0^{(i)}$ hipótesis nula del contraste parcial con p-valor $p_{(i)}$.

El criterio de Holm consiste en rechazar $H_0^{(1)}, \ldots, H_0^{(m)}$ y no rechazar $H_0^{(m+1)}, \ldots, H_0^{(k)}$ si

$$\forall\, i \leqslant m \quad p_{(i)} \leqslant \frac{\alpha}{k-i+1} \qquad \wedge \qquad p_{(m+1)} > \frac{\alpha}{k-m}$$

Vamos a comprobar que el procedimiento de Holm mantiene un nivel de significación global α:

- Sea $I_0 = \{i \mid H_0^{(i)} \text{ es cierta}\}$
- Sea $k_0 = \#I_0$ el número de hipótesis nulas ciertas.
- Supongamos que $H_0^{(j)}$ es la primera hipótesis nula cierta pero rechazada.
- $H_0^{(1)}, \ldots, H_0^{(j-1)}$ son falsas.
- $j - 1 \leqslant k - k_0$
- $k_0 \leqslant k - j + 1$
- $j \leqslant m \implies p_{(j)} \leqslant \frac{\alpha}{k-j+1} \leqslant \frac{\alpha}{k_0}$
- Por tanto, si rechazamos una hipótesis nula verdadera, tiene que haber una hipótesis nula verdadera con p-valor $\leqslant \frac{\alpha}{k_0}$.
- Por tanto,

$$\begin{aligned}
\text{Pr[error global de tipo I]} &= \Pr\left(\text{rechazar alguna } H_0^{(i)} \text{ verdadera}\right) \\
&\leqslant \Pr\left(\exists\, i \in I_0, p_i \leqslant \frac{\alpha}{k_0}\right) \\
&\leqslant \sum_{i \in I_0} \Pr\left(p_i \leq \frac{\alpha}{k_0}\right) \\
&= \sum_{i \in I_0} \Pr\left(\text{RC}_i \text{ a nivel } \frac{\alpha}{k_0}\right) \\
&\leqslant \sum_{i \in I_0} \frac{\alpha}{k_0} = \alpha
\end{aligned}$$

porque p-valor $\leqslant a$ si y sólo si la muestra pertenece a la región crítica construida a nivel de significación a. De hecho, si la población es continua, $p_i \overset{H_0}{\hookrightarrow} \mathcal{U}(0,1)$.

4. R

A veces, en vez de dividir el α global, se expresan los p-valores de los contrastes sencillos multiplicados por el correspondiente coeficiente. Por ejemplo, al usar en R las funciones `pairwise.prop.test`, `pairwise.t.test` y `pairwise.wilcox.test`:

```
> summary (cw <- chickwts)

     weight             feed
 Min.   :108.0   casein   :12
 1st Qu.:204.5   horsebean:10
 Median :258.0   linseed  :12
 Mean   :261.3   meatmeal :11
 3rd Qu.:323.5   soybean  :14
 Max.   :423.0   sunflower:12


> pairwise.t.test (cw$weight, cw$feed, "none") # p-valores crudos

Pairwise comparisons using t tests with pooled SD

data:  cw$weight and cw$feed

          casein  horsebean linseed meatmeal soybean
horsebean 2.1e-09 -         -       -        -
linseed   1.5e-05 0.01522   -       -        -
meatmeal  0.04557 7.5e-06   0.01348 -        -
soybean   0.00067 0.00032   0.20414 0.17255  -
sunflower 0.81249 8.2e-10   6.2e-06 0.02644  0.00030

P value adjustment method: none


> pairwise.t.test (cw$weight, cw$feed, "bonferroni")

Pairwise comparisons using t tests with pooled SD

data:  cw$weight and cw$feed

          casein  horsebean linseed meatmeal soybean
horsebean 3.1e-08 -         -       -        -
linseed   0.00022 0.22833   -       -        -
meatmeal  0.68350 0.00011   0.20218 -        -
soybean   0.00998 0.00487   1.00000 1.00000  -
sunflower 1.00000 1.2e-08   9.3e-05 0.39653  0.00447
```

```
P value adjustment method: bonferroni


> pairwise.t.test (cw$weight, cw$feed) # "holm" por omisión

Pairwise comparisons using t tests with pooled SD

data:  cw$weight and cw$feed

          casein  horsebean linseed meatmeal soybean
horsebean 2.9e-08 -         -       -        -
linseed   0.00016 0.09435   -       -        -
meatmeal  0.18227 9.0e-05   0.09435 -        -
soybean   0.00532 0.00298   0.51766 0.51766  -
sunflower 0.81249 1.2e-08   8.1e-05 0.13218  0.00298

P value adjustment method: holm
```

Se puede solicitar la corrección de p-valores directamente:

```
> p.adjust (c(0.02,0.03), "none")
[1] 0.02 0.03
> p.adjust (c(0.02,0.03), "bonferroni")
[1] 0.04 0.06
> p.adjust (c(0.02,0.03), "holm")
[1] 0.04 0.04

> p.adjust (c (0.001, 0.005, 0.01, 0.015, 0.02, 0.04), "bonf")
[1] 0.006 0.030 0.060 0.090 0.120 0.240
> p.adjust (c (0.001, 0.005, 0.01, 0.015, 0.02, 0.04), "holm")
[1] 0.006 0.025 0.040 0.045 0.045 0.045
```