

Inferencia Respecto a un Vector de Medias Simple. Con Matriz Varianza Covarianza Conocida

En el caso univariado, para hacer inferencias respecto a la media μ de una distribución se considera que X_1, \dots, X_n es una muestra aleatoria de una distribución $\text{Normal}(\mu, \sigma^2)$ entonces cada $X_i \sim \text{Normal}(\mu, \sigma^2)$. Se utiliza el estadístico de prueba $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim \text{Normal}(0,1)$ si la σ^2 es conocida, y si σ^2 no se conoce se utiliza el

estadístico de prueba t de Student $t = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \sim t(n-1)$. Se puede reescribir el

estadístico de prueba $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ de la siguiente manera:

$Z^2 = n(\bar{X} - \mu_0)\Sigma^{-1}(\bar{X} - \mu_0) \sim \chi^2(p)$. En el caso multivariado, cuando se conoce la matriz varianza covarianza y se supone que $\underline{X}_i \stackrel{\text{iid}}{\sim} \text{Normal}(\underline{\mu}, \Sigma)$, donde:

1. $\underline{X}_i = (X_{i1}, \dots, X_{ip})'$ es el vector $p \times 1$ de la observación i.
2. $\underline{\mu} = (\mu_1, \dots, \mu_p)'$ es el vector de medias $p \times 1$.
3. $\Sigma = \{\sigma_{jh}\}$ es la matriz de covariancias $p \times p$.

Y se asume que Σ se conoce y que deseamos probar las hipótesis.

$$H_0 : \underline{\mu} = \underline{\mu}_0 \text{ vs } H_1 : \underline{\mu} \neq \underline{\mu}_0$$

$$H_0 = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \dots \\ \mu_{p0} \end{bmatrix}, \quad H_1 = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \dots \\ \mu_{p0} \end{bmatrix}$$

donde $\underline{\mu}_0$ es algún vector conocido especificado por la hipótesis nula, entonces se debe utilizar la estadística de prueba $Z^2 = n(\bar{X} - \mu_0)\Sigma^{-1}(\bar{X} - \mu_0) \sim \chi^2(p)$, teniendo en cuenta que la prueba es unilateral a la derecha.

Ejemplo 1: Se seleccionan 50 porciones de cebiche de pescado en Surquillo. Se tiene interés en 4 nutrientes:

X_1 : Contenido de energía en kcal/100 g

X_2 : Contenido de agua en g/100 g

X_3 : Contenido de carbohidratos totales en g/100 g

X_4 : Contenido de proteínas en g/100 g

$$\text{Asumiendo que } \Sigma = \begin{bmatrix} 9 & -4.5 & 3.5 & 3.5 \\ -4.5 & 4 & -2.5 & -2.5 \\ 3.5 & -2.5 & 2 & 1.5 \\ 3.5 & -2.5 & 1.5 & 2 \end{bmatrix} \text{ pruebe: } H_0 = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} 90 \\ 75.3 \\ 13 \\ 7.3 \end{bmatrix}$$

Inferencia Respecto a un Vector de Medias Simple. Con Matriz Varianza Covarianza Desconocida

En el caso univariado, para hacer inferencias respecto a la media μ de una distribución se considera que X_1, \dots, X_n es una muestra aleatoria de una distribución Normal(μ, σ^2) entonces cada $X_i \sim \text{Normal}(\mu, \sigma^2)$. Se utiliza el estadístico de prueba

$Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim \text{Normal}(0,1)$ si la σ^2 es conocida, y si σ^2 no se conoce se utiliza el

estadístico de prueba t de Student $t = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \sim t(n-1)$. Se puede reescribir el

estadístico de prueba de la siguiente manera: $t^2 = n(\bar{X} - \mu_0)(S^2)^{-1}(\bar{X} - \mu_0) \sim F(1, n-1)$

Para el caso multivariado se supone que $\underline{X}_i \stackrel{\text{iid}}{\sim} \text{Normal}(\underline{\mu}, \Sigma)$, donde:

1. $\underline{X}_i = (X_{i1}, \dots, X_{ip})'$ es el vector $p \times 1$ de la observación i.
2. $\underline{\mu} = (\mu_1, \dots, \mu_p)'$ es el vector de medias $p \times 1$.
3. $\Sigma = \{\sigma_{jh}\}$ es la matriz de covariancias $p \times p$.

Asumiendo que Σ no se conoce y que deseamos probar las hipótesis.

$$H_0 : \underline{\mu} = \underline{\mu}_0 \text{ vs } H_1 : \underline{\mu} \neq \underline{\mu}_0$$

Donde $\underline{\mu}_0$ es algún vector conocido especificado por la hipótesis nula.

Prueba Estadística T^2 de Hotelling

Es la extensión multivariada del estadístico $t^2 = n(\bar{X} - \mu_0)(S^2)^{-1}(\bar{X} - \mu_0) \sim F(1, n-1)$ y se representa de la siguiente manera:

$$T^2 = n(\bar{\underline{X}} - \underline{\mu}_0)'(S)^{-1}(\bar{\underline{X}} - \underline{\mu}_0) \text{ donde:}$$

1. $\bar{\underline{X}} = \frac{1}{n} \sum_{i=1}^n \underline{X}_i$ es el vector de medias muestrales.

2. $S = \frac{1}{n} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})'$ es la matriz de covarianza muestral.

3. $\frac{1}{n}S$ es la matriz de covarianza de $\bar{\underline{X}}$.

A $X = \{x_{ij}\}$ se le llama matriz de datos $n \times p$. Podemos reescribir:

1. $\bar{\underline{X}} = \frac{1}{n} \sum_{i=1}^n \underline{X}_i = n^{-1} X' \mathbf{1}_n$ es el vector de medias muestrales.

2. $S = \frac{1}{n} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})' = \frac{1}{n-1} X'_c X_c$ es la matriz de covarianza muestral.

3. $X_c = CX$ con $C = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n'$ que denota una matriz centrada.

Inferencia usando T^2 de Hotelling

Considerando H_0 , T^2 de Hotelling sigue una distribución en escala de F.

$$T^2 \sim \frac{(n-1)p}{n-p} F(p, n-p)$$

Se rechaza la H_0 si $T^2 > \frac{(n-1)p}{n-p} F(p, n-p)(\alpha)$

Donde $F(p, n-p)(\alpha)$ es el percentil 100α de la distribución $F(p, n-p)$ y α es el nivel de significación de la prueba.

Ejemplo 2: En 200 porciones de 100ml de vino se registraron: el pH, el % de alcohol, la densidad en mg y el contenido de cloruro en mg. Realizar inferencias respecto al vector de medias con matriz varianza covarianza desconocida.

Inferencias sobre dos vectores de medias Con dos muestras dependientes

Al igual que en el caso univariado, la prueba T^2 de Hotelling de dos muestras dependientes multivariadas realiza con la prueba de una muestra en una puntuación de diferencia.

Se quiere probar:

$$H_0 : \underline{\mu}_D = \underline{\mu}_1 - \underline{\mu}_2 = \underline{\mu}_0 \text{ vs } H_1 : \underline{\mu}_D = \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\mu}_0$$

Ejemplo 3: Una compañía de calzado está considerando eliminar un modelo de calzado existente (Modelo 2) y quedarse con el Modelo 1. La compañía hizo que los mismos sujetos evaluaran tanto el Modelo 1 como el Modelo 2, y observaron si había una diferencia significativa entre los dos modelos, lo que les ayudaría a decidir si reemplazar el Modelo 2 por el Modelo 1.

Inferencias sobre dos vectores de medias

Con dos muestras independientes

Se quiere probar:

$$H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\mu}_0 \text{ vs } H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\mu}_0$$

$$H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \dots \\ \mu_{p1} \end{bmatrix} - \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \dots \\ \mu_{p2} \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \dots \\ \mu_{p0} \end{bmatrix}, \quad H_1 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \dots \\ \mu_{p1} \end{bmatrix} - \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \dots \\ \mu_{p2} \end{bmatrix} \neq \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \dots \\ \mu_{p0} \end{bmatrix}$$

Ejemplo 4: Considere las mediciones de dos especies de escarabajo (*Haltica oleracea* y *Haltica carduorum*). Estas mediciones son la longitud del tórax (x1 en micras) y la longitud del ala (x2 en 0.01 mm).

Inferencias sobre más de dos vectores de medias Con tres o más muestras independientes

Esto se estudiará cuando lleguemos a MANOVA.

Las hipótesis serán las siguientes:

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_G$$

$$H_0 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \dots \\ \mu_{p1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \dots \\ \mu_{p2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1G} \\ \mu_{2G} \\ \dots \\ \mu_{pG} \end{bmatrix}$$

$$H_1 : \text{Al menos un } \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \dots \\ \mu_{pi} \end{bmatrix} \text{ es diferente. Para } i = 1, \dots, G$$