
Inferencia paramétrica

INFERENCIA SOBRE LAS VARIANZAS DE DOS LEYES NORMALES INDEPENDIENTES

Sólo tiene sentido si las muestras de las dos leyes normales son *independientes*, es decir, si cada una se elige de forma independiente de la otra, por lo que pueden tener tamaños iguales o distintos.

INTERVALO DE CONFIANZA DE NIVEL $100(1 - \alpha)$ DE σ_1^2 / σ_2^2 PARA DOS LEYES NORMALES CON MUESTRAS INDEPENDIENTES de tamaños n_1 y n_2 .

$$\left(\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{n_1-1, n_2-1, (1-(\alpha/2))}} , \frac{s_1^2}{s_2^2} \cdot F_{n_2-1, n_1-1, (1-(\alpha/2))} \right)$$

Los extremos de un intervalo para el cociente de dos desviaciones típicas son las raíces cuadradas positivas de los anteriores.

CONTRASTES SOBRE LAS VARIANZAS DE DOS LEYES NORMALES INDEPENDIENTES

Hipótesis nulas: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_0 : \sigma_1^2 \leq \sigma_2^2$ \circ $H_0 : \sigma_1^2 \geq \sigma_2^2$.

Sólo puede hacerse con dos muestras independientes de tamaños n_1 y n_2 .

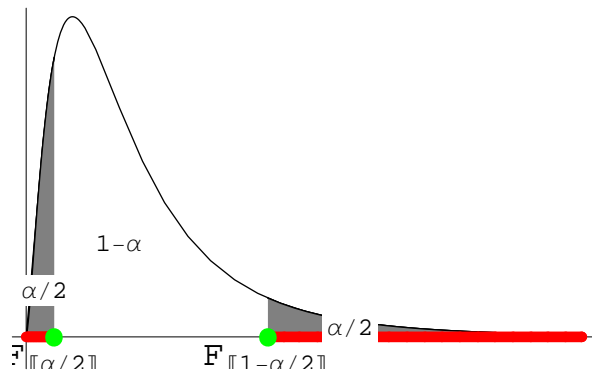
El estadístico de contraste es $F = s_1^2/s_2^2$, que sigue una ley F_{n_1-1, n_2-1} , si H_0 es cierta.

Hipótesis alternativas

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

$$H_1 : \sigma_1^2 < \sigma_2^2$$



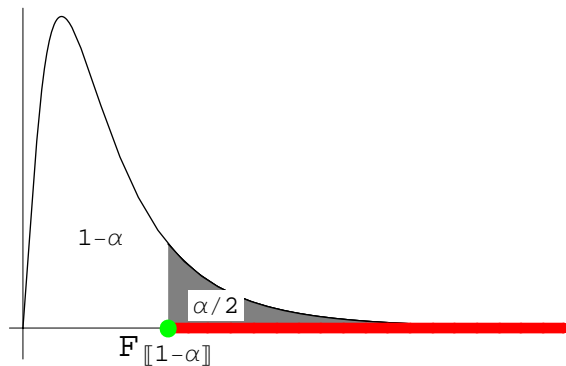
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Criterios de decisión al nivel de significación α

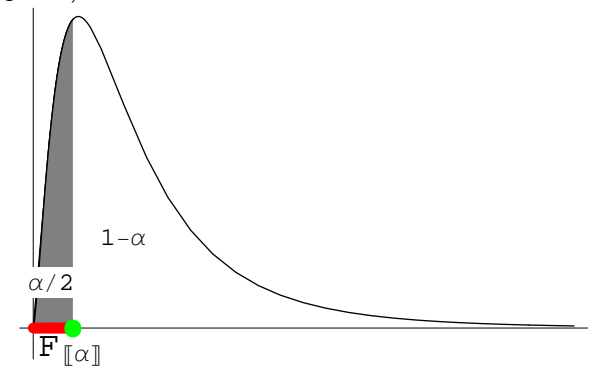
$$\begin{cases} \text{Rechazar } H_0 \text{ si } f_0 \geq F_{n_1-1, n_2-1, 1-(\alpha/2)} \\ \text{Rechazar } H_0 \text{ si } f_0 \leq \frac{1}{F_{n_2-1, n_1-1, 1-(\alpha/2)}} \end{cases}$$

$$\text{Rechazar } H_0 \text{ si } f_0 \geq F_{n_1-1, n_2-1, 1-\alpha}$$

$$\text{Rechazar } H_0 \text{ si } f_0 \leq \frac{1}{F_{n_2-1, n_1-1, 1-\alpha}}$$



$$H_1 : \sigma_1^2 > \sigma_2^2$$



$$H_1 : \sigma_1^2 < \sigma_2^2$$

INFERENCIA SOBRE LAS MEDIAS DE DOS LEYES NORMALES

Muestras relacionadas o apareadas Son muestras que se forman eligiendo **n** pares de unidades, una de cada población, de tal modo que las unidades de cada par sean muy homogéneas entre sí, a la vez que son muy diferentes de los demás pares.

Las dos muestras X_1, X_2, \dots, X_n **e** Y_1, Y_2, \dots, Y_n ***tienen que ser del mismo tamaño:*** $n = n_1 = n_2$.

INFERENCIA SOBRE LAS MEDIAS DE DOS LEYES NORMALES CON DOS MUESTRAS RELACIONADAS.

Para cada par (X_i, Y_i) se define la variable $D_i = X_i - Y_i$, que representa la diferencia entre los datos de ese par. Las n variables D_i forman una muestra aleatoria de una ley normal de media $\mu_D = \mu_1 - \mu_2$ y varianza σ_D^2 , que siempre se supone desconocida.

En estas muestras se definen la media muestral $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$

y la cuasivarianza de las diferencias $s_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$

INTERVALOS DE CONFIANZA DEL $100(1 - \alpha)$ DE $\mu_D = \mu_1 - \mu_2$ EN DOS LEYES NORMALES CON MUESTRAS RELACIONADAS de tamaño n

EL INTERVALO ES $\left(\bar{D} - \sqrt{s_D^2/n} \cdot t_{n-1, 1-\frac{\alpha}{2}} , \bar{D} + \sqrt{s_D^2/n} \cdot t_{n-1, 1-\frac{\alpha}{2}} \right)$

Si el tamaño de ambas muestras es $n \geq 30$, pueden aproximarse los percentiles de la t por los de la $N(0, 1)$

CONTRASTES SOBRE μ_D CON MUESTRAS RELACIONADAS

Hipótesis nulas: $H_0 : \mu_D = 0$, $H_0 : \mu_D \leq 0$ • $H_0 : \mu_D \geq 0$,

El estadístico de contraste es $T = \frac{\bar{D}}{\sqrt{s_D^2/n}}$, que sigue la ley t_{n-1} cuando H_0 es cierta.

Hipótesis alternativas

$$H_1 : \mu_D \neq 0$$

$$H_1 : \mu_D > 0$$

$$H_1 : \mu_D < 0$$

Criterios de decisión para el nivel α

$$\text{Rechazar } H_0 \text{ si } |t_0| \geq t_{n-1, 1-\frac{\alpha}{2}}$$

$$\text{Rechazar } H_0 \text{ si } t_0 \geq t_{n-1, 1-\alpha}$$

$$\text{Rechazar } H_0 \text{ si } t_0 \leq -t_{n-1, 1-\alpha}$$

INTERVALOS DE CONFIANZA DEL $100(1 - \alpha)$ DE $\mu_1 - \mu_2$ EN DOS LEYES NORMALES CON MUESTRAS INDEPENDIENTES de tamaños n_1 y n_2 .

a) Si se conocen las varianzas de las leyes normales (σ_1^2 y σ_2^2),

EL INTERVALO ES
$$\left((\bar{x} - \bar{y}) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x} - \bar{y}) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

b) Si no se conocen las varianzas, pero se suponen iguales, se calcula la

cuasivarianza ponderada,
$$s_p^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2} = \frac{SC_X + SC_Y}{n_1 + n_2 - 2}$$

EL INTERVALO ES

$$\left((\bar{x} - \bar{y}) - t_{n_1+n_2-2, 1-\frac{\alpha}{2}} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, (\bar{x} - \bar{y}) + t_{n_1+n_2-2, 1-\frac{\alpha}{2}} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right)$$

Si en b) son $n_1 \geq 30$ y $n_2 \geq 30$, $t_{n_1+n_2-2, 1-\frac{\alpha}{2}}$ se aproxima por $Z_{1-\frac{\alpha}{2}}$ de la ley $N(0, 1)$.

- c) ***Si no se conocen las varianzas, se suponen distintas*** y los tamaños muestrales son iguales o mayores que 30,

EL INTERVALO ES
$$\left((\bar{x} - \bar{y}) - \mathbf{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ; (\bar{x} - \bar{y}) + \mathbf{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

- d) ***Si no se conocen las varianzas, se suponen distintas*** y algún tamaño muestral es menor que 30,

EL INTERVALO ES:

$$\left((\bar{x} - \bar{y}) - \left(t_{[f], 1-(\alpha/2)} \right) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ; (\bar{x} - \bar{y}) + \left(t_{[f], 1-(\alpha/2)} \right) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

donde $[f]$ es la parte entera del número

$$f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1} \right)} \quad \text{(Criterio de Welch)}$$

**CONTRASTES SOBRE LAS MEDIAS DE DOS LEYES NORMALES CON
MUESTRAS INDEPENDIENTES de tamaños n_1 y n_2 .**

Hipótesis nulas: $H_0 : \mu_1 = \mu_2$, $H_0 : \mu_1 \leq \mu_2$ o $H_0 : \mu_1 \geq \mu_2$,

a) Si se conocen las varianzas de las leyes normales (σ_1^2 y σ_2^2), el estadístico de contraste es

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \text{ que sigue la ley } N(0, 1) \text{ si } H_0 \text{ es cierta.}$$

Hipótesis alternativas

Criterios de decisión al nivel α

P-valores

$$H_1 : \mu_1 \neq \mu_2$$

Rechazar H_0 si $|z_0| \geq \mathbf{Z}_{1-\frac{\alpha}{2}}$

$$\mathbf{P} = 2P(Z \geq |z_0|)$$

$$H_1 : \mu_1 > \mu_2$$

Rechazar H_0 si $z_0 \geq \mathbf{Z}_{1-\alpha}$

$$\mathbf{P} = P(Z \geq z_0)$$

$$H_1 : \mu_1 < \mu_2$$

Rechazar H_0 si $z_0 \leq -\mathbf{Z}_{1-\alpha}$

$$\mathbf{P} = P(Z \leq z_0)$$

b) Si no se conocen las varianzas, no se suponen iguales, y los tamaños de las muestras son superiores a 30, el estadístico de contraste es

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \text{ que sigue la ley } N(0, 1) \text{ si } H_0 \text{ es cierta.}$$

Las reglas de decisión son análogas a las del caso *a*).

c) **Si no se conocen las varianzas, pero se suponen iguales**, se calcula la **cuasivarianza ponderada**, S_p^2 y el estadístico de contraste es

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ que sigue la ley } t_{n_1+n_2-2} \text{ si } H_0 \text{ es cierta.}$$

Donde

$$S_p^2 = \frac{SC_1 + SC_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Hipótesis alternativas

$$H_1 : \mu_1 \neq \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

Criterios de decisión al nivel α

$$\text{Rechazar } H_0 \text{ si } |t_0| \geq t_{n_1+n_2-2, 1-\frac{\alpha}{2}}$$

$$\text{Rechazar } H_0 \text{ si } t_0 \geq t_{n_1+n_2-2, 1-\alpha}$$

$$\text{Rechazar } H_0 \text{ si } t_0 \leq -t_{n_1+n_2-2, 1-\alpha}$$

Si los tamaños de las dos muestras son mayores o iguales que 30, el estadístico T sigue aproximadamente una ley $N(0, 1)$ y las reglas de decisión son análogas a las del caso α).

d) Si no se conocen las varianzas, no se suponen iguales, y alguno de los tamaños muestrales es inferior a 30, sólo existen criterios aproximados de decisión, entre ellos el **criterio de Welch**

El criterio de Welch consiste en tomar el estadístico de contraste

$$T = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \text{ cuya distribución, cuando } H_0 \text{ es cierta, es una ley } t \text{ de Student cuyo}$$

número de grados de libertad es la parte entera del número f , donde

$$f = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

NOTA: Si la hipótesis nula de alguno de estos contrastes se refiere a una diferencia no nula entre las medias, se designa a esta diferencia con la letra griega δ_0 (delta) y las hipótesis nulas serán $H_0 : \mu_1 - \mu_2 = \delta_0$, $H_0 : \mu_1 - \mu_2 \leq \delta_0$ o $H_0 : \mu_1 - \mu_2 \geq \delta_0$, contra las respectivas alternativas.

Los contrastes se efectúan como en los casos anteriores, sustituyendo el valor 0 de las hipótesis alternativas por el valor δ_0 y restando esta cantidad a los numeradores de los distintos estadísticos de contraste.

RESUMEN:

En un contraste sobre las medias de dos variables normales, con muestras independientes de tamaños n_1 y n_2 :

Las hipótesis nulas son : $H_0 : \mu_1 = \mu_2$, $H_0 : \mu_1 \leq \mu_2$ o $H_0 : \mu_1 \geq \mu_2$,

Los estadísticos de contraste y sus distribuciones pueden ser:

$$\sigma_1^2 \text{ y } \sigma_2^2 \text{ conocidas: } \Rightarrow Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

$$\sigma_1^2 \text{ y } \sigma_2^2 \text{ desconocidas: } \left\{ \begin{array}{l} \text{iguales } T = \frac{\bar{X} - \bar{Y}}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} : \left\{ \begin{array}{ll} T \sim t_{n_1+n_2-2} & \text{si } n_1 < 30 \text{ y/o } n_2 < 30 \\ T \sim N(0, 1) & \text{si } n_1 \geq 30 \text{ y } n_2 \geq 30 \end{array} \right. \\ \text{distintas } T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} : \left\{ \begin{array}{ll} T \sim t_{[f]} & \text{si } n_1 < 30 \text{ y/o } n_2 < 30 \\ T \sim N(0, 1) & \text{si } n_1 \geq 30 \text{ y } n_2 \geq 30 \end{array} \right. \end{array} \right.$$

OBSERVACIÓN IMPORTANTE:

Cuando hacemos un contraste con un programa informático, la salida que nos da valor del p-valor P . Rechazaremos la hipótesis nula comparando esta P con un valor α , que elegiremos nosotros y que como máximo valdrá 0.1.

Si $P \leq \alpha$ ***se rechaza*** H_0

Contraste estadísticamente significativo a un nivel α

esto no significa que H_0 sea falsa, sino que es improbable que sea cierta, o dicho de otro modo, con los datos de los que se dispone no hay suficiente evidencia para aceptarla.

Si $P > \alpha$ ***NO se rechaza*** H_0

Contraste estadísticamente NO significativo a un nivel α

es decir, con los datos de los que se dispone no hay suficiente evidencia para rechazarla.

En resumen, un test no demuestra nada, solo sugiere que la hipótesis nula es muy probable que sea cierta o es muy probable que sea falsa.