ESTADÍSTICA

Inferencia Estadística

Índice

1. Introducción

La Inferencia Estadística se centra en analizar e interpretar las observaciones para obtener conclusiones sobre la población en estudio.

Supongamos que queremos estudiar la edad de los profesores de universidad en España. Primero tendremos que definir la población de interés como el conjunto de todos los profesores de España. Deseamos determinar parámetros que caractericen la posible distribución que siguen estos datos.

Si se dispusiera del tiempo y de los recursos necesarios se podría hacer un censo a la población y dichos parámetros quedarían perfectamente delimitados. Normalmente no es posible y es necesario elegir una muestra representativa de la población.

Los problemas de inferencia estadística se clasifican :

- Si se desea proporcionar un rango de variación para un determinado parámetro desconocido de la distribución tenemos Estimación por intervalos de confianza.

- Si se desea corroborar o invalidar una determinada afirmación acerca de los parámetros de una distribución o de la propia distribución tenemos Contraste de Hipótesis.

Estimación por intervalos

Este método consiste en encontrar un intervalo dentro del cuál se encuentra el parámetro desconocido con una probabilidad muy alta que llamaremos nivel de confianza. El intervalo de confianza es un rango de valores entre los que posiblemente se encuentre el verdadero valor del parámetro.

Sea X una v.a. con función de densidad f (x, θ) de parámetro desconocido θ, una m.a.s (x_1, ..., x_n), y el nivel de confianza (1 − α) donde α es el nivel de significación.

Un intervalo de confianza (1 − α) es:

$$\theta_1\left(x_1,\ldots,x_n\right) \leq \theta \leq \theta_2\left(x_1,\ldots,x_n\right)$$

Para nivel de confianza 95% (1-α = 0.95) significa que: si disponemos de 100 muestras en 95 de ellas el IC que se obtiene contendría el verdadero valor del parámetro.

Estimación por contrastes

Los contrastes de hipótesis dan respuesta estadística a los interrogantes que nos planteamos en investigaciones empíricas. Estamos interesados en aceptar o rechazar ciertas hipótesis que plantearemos acerca de los parámetros de una distribución

Hipótesis nula H_0 es la hipótesis que se contrasta y se mantiene a menos que se demuestre que es falsa, sin embargo, puede ser rechazada por los datos. Cuando rechazamos H_0 es porque la evidencia empírica de los datos hace mucho más plausible la hipótesis alternativa H_1.

Tipos de contrastes:

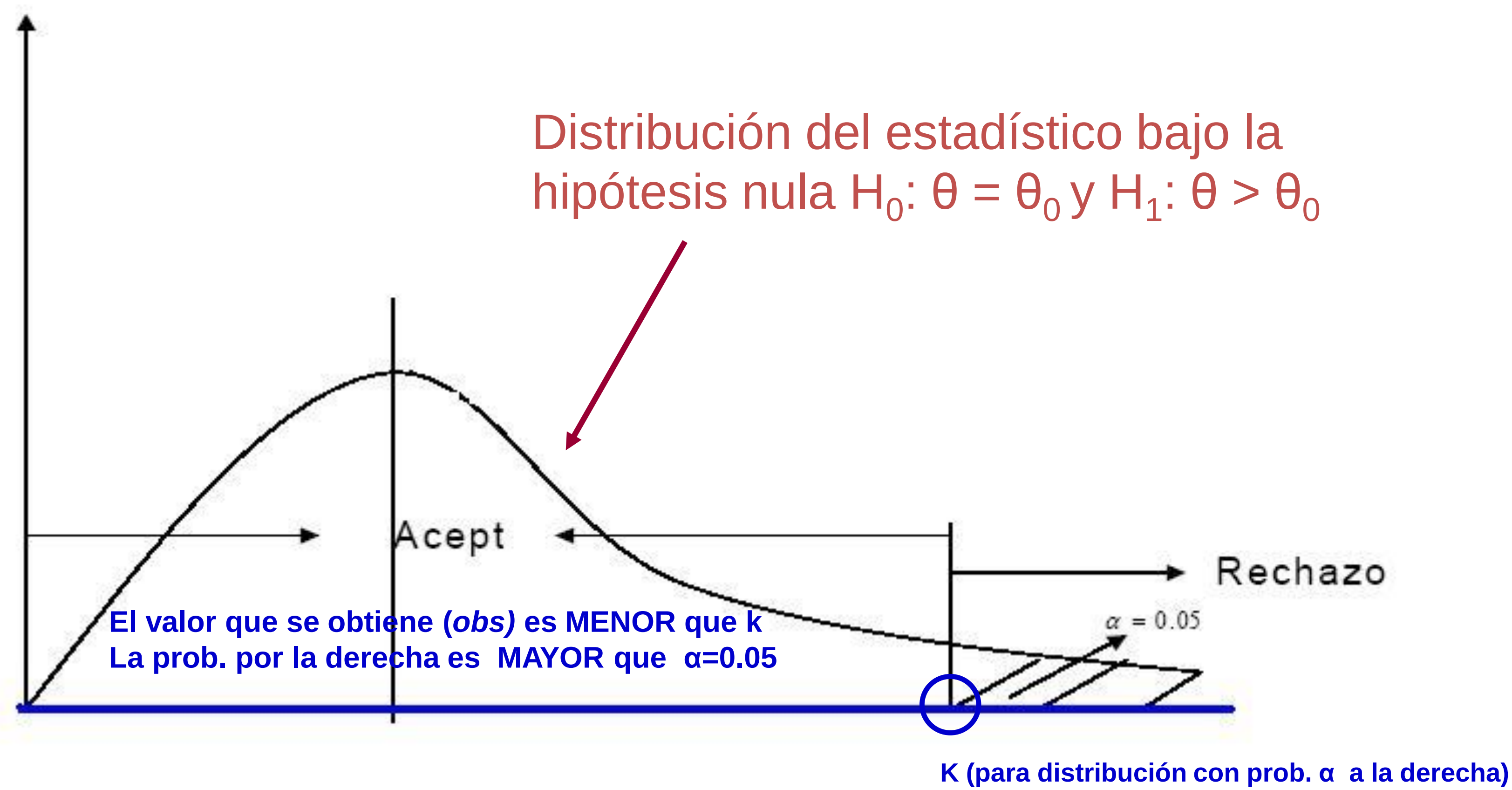
1.Contrastes bilaterales, ya que al rechazar la hipótesis nula no sabemos en qué sentido puede ser falsa

$$H_1 : \theta \neq \theta_0 \quad \text{donde } \theta_0 \text{ es un valor conocido}$$

2.Contrastes unilaterales, de tal manera que si rechazamos la hipótesis nula o bien es porque verdadero valor del parámetro es menor que la cantidad especificada en la nula o mayor.

$$H_1 : \theta < \theta_0 \quad \text{donde } \theta_0 \text{ es un valor conocido} \quad \text{ó}$$
$$H_1 : \theta > \theta_0$$

Estimación gráficamente



Es posible comparar con el valor obtenemos en el eje x.

Otra opción es comparar con la probabilidad, p_valor.

El p_valor calcula siempre la probabilidad de la hipótesis alternativa.

2. Una muestra. Media, Varianza y Proporción
a) y b) c) d)

Sea una variable aleatoria X y una m.a.s. $(x_1,...,x_n)$.

a) Intervalo de confianza para la MEDIA con **varianza conocida** siendo una población normal ó para un tamaño muestral grande.

El estadístico y su pivote correspondiente que consideraremos es

$$\bar{X} = \frac{\sum^{n} X_i}{n} \to N(\mu, \sigma/\sqrt{n}) \qquad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$$

Buscamos en la Normal un valor de $z_{\alpha/2}$ tal que: $P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$

Despejamos el parámetro μ.
Por tanto el IC para la media a nivel (1-α) es:

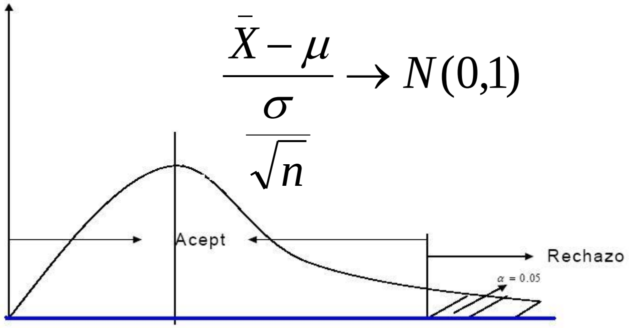
$$IC_\alpha(\mu) = \left[\bar{x} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{error maximo}}\right] = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Contraste hipótesis para la MEDIA con **varianza conocida** siendo una población normal ó para un tamaño muestral grande.

Queremos realizar el siguiente contraste unilateral.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$



Por tanto, si $z_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$ se rechaza H_0

Equivalentemente, si $p_{valor} = P\left(Z > z_{OBS}\right) < \alpha$ se rechaza H_0

Contraste hipótesis para la MEDIA con **varianza conocida** siendo una población normal ó para un tamaño muestral grande.

Queremos realizar el siguiente contraste bilateral.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Por tanto, si $z_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{\alpha/2}$ ó si $z_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{\alpha/2}$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = P\left(|Z| > z_{OBS}\right) = 2P\left(Z > |z_{OBS}|\right) < \alpha$$ se rechaza H_0

Cuando n es grande $S_X^2 = \sigma_X^2$, para cualquier distribución de la v.a. X (normal o no).

Ejemplo:

De una población normal N(μ, σ = 2.5) se extrae una m.a.s. de tamaño 30, siendo $\sum_{i=1}^{n} x_i = 77$. Calcula un intervalo de confianza de la media de la población al nivel de confianza 0.95.

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$$

La media de la población es: $\frac{\sum_{i=1}^{n} x_i}{30} = \frac{77}{30} = 2.57$

teniendo en cuenta que la población de la que procede la muestra es normal con varianza conocida, el intervalo que nos piden es para 1 − α = 0.95, es decir, α = 0.05

$$P\left(\bar{x} - z_{0.05/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{0.05/2}\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$IC_{0.95}(\mu) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] =$$

$$= \left[2.57 - 1.96\frac{2.5}{\sqrt{30}} ; 2.57 + 1.96\frac{2.5}{\sqrt{30}}\right] = [1.68;3.46]$$

Sea una variable aleatoria $X \in N(\mu, \sigma^2)$ y una m.a.s. $(x_1,...,x_n)$.

b) Intervalo de confianza para la MEDIA de una población normal con **varianza desconocida**. El estadístico y su pivote correspondiente que consideraremos es

$$\bar{X} = \frac{\sum^{n} X_i}{n} \rightarrow N(\mu, \sigma/\sqrt{n}) \qquad \frac{\bar{X}-\mu}{S_X/\sqrt{n}} \rightarrow t_{n-1}$$

t-student (simétrica)

Buscamos en t-student un valor de $t_{n-1,\, \alpha/2}$ tal que: $P\left(-t_{\alpha/2,n-1} \le \frac{\bar{x}-\mu}{s_x/\sqrt{n}} \le t_{\alpha/2,n-1}\right) = 1-\alpha$

Despejamos el parámetro μ.
Por tanto el IC para la media a nivel (1-α) es:

$$IC_\alpha(\mu) = \left[\bar{x} - t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}};\ \bar{x} + t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}}\right]$$

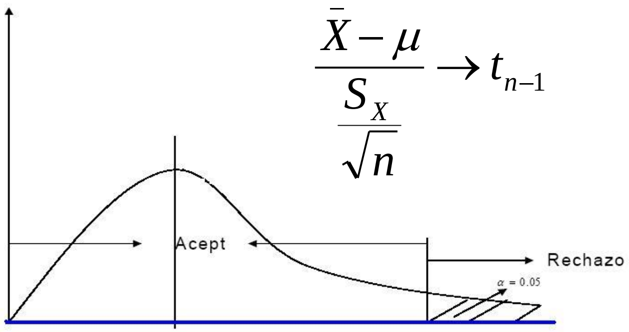
$$s_X^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$

Intervalo de confianza para la MEDIA de una población normal con **varianza desconocida**.

Queremos realizar el siguiente contraste unilateral.

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0$$

$$\frac{\bar{X} - \mu}{\frac{S_X}{\sqrt{n}}} \rightarrow t_{n-1}$$



Por tanto, si $t_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} > t_{n-1,\alpha}$ se rechaza H_0

Equivalentemente, si $p_{valor} = P\left(t_{n-1} > t_{OBS}\right) < \alpha$ se rechaza H_0

Intervalo de confianza para la MEDIA de una población normal con **varianza desconocida**.

Queremos realizar el siguiente contraste bilateral.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Por tanto, si $t_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} > t_{n-1,\alpha/2}$ ó si $t_{OBS} = \dfrac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} < -t_{n-1,\alpha/2}$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = P\left(|\, t_{n-1} | > t_{OBS}\right) = 2P\left(t_{n-1} > |\, t_{OBS} \,|\right) < \alpha$$ se rechaza H_0

Cuando n es grande $S_X^2 = \sigma_X^2$, para cualquier distribución de X (normal o no).
Y es que la t-student se comporta como una normal para n grande.

Ejemplo:

La Oficina de Control de Calidad del canal de Isabel II con vistas a establecer los límites máximos y mínimos válidos para el consumo de agua de la localidad de Cobeña tomó una muestra de tamaño 10, de la que se obtuvo una media $\bar{x}$ = 998 cm^3 y desviación típica s_x =5 cm^3. Se supone que la cantidad de agua contenida en el pozo sigue una ley normal. Calcular los límites de confianza al nivel del 95%.

$$\frac{\bar{X}-\mu}{S_X/\sqrt{n}} \rightarrow t_{n-1}$$

Nivel de confianza: 1- α = 0.95
$t_{0.025,9}$ = 2.262

$$IC_\alpha(\mu) = \left[\bar{x} - t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}};\bar{x} + t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}}\right] =$$

$$= \left[998 - 2.262\frac{5}{\sqrt{10}};998 + 2.262\frac{5}{\sqrt{10}}\right] = [994.41;1001.58]$$

Ejemplo:

Supongamos una población normal de la que se extrae una muestra de tamaño 10 con los siguientes resultados: 10, 11, 12, 13, 14, 15, 11, 12, 13, 14. Contraste la hipótesis nula media igual a 12 frente a la alternativa media distinta de 12, al nivel de significación del 5%.

$$\frac{\bar{X}-\mu}{S_X/\sqrt{n}} \rightarrow t_{n-1}$$

$$\begin{cases} H_0 : \mu = 12 \\ H_1 : \mu \neq 12 \end{cases} \qquad t_{obs} = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{12.5-12}{\frac{1.581139}{\sqrt{10}}} = 1$$

$$|t_{obs}| = 1 < t_{n-1,\alpha/2} = t_{9,0,025} = 2.262157$$ no se rechaza Ho

p-valor = $2 \cdot P(t_9 > 1) = 2 \cdot 0.1717182 = 0.3434 > 0.05 = \alpha$ no se rechaza Ho

Al tratarse de um contraste bilateral, también sería adecuado utilizar el IC para el contraste.

$$12.5 - 2.26\frac{1.58}{\sqrt{10}} < \mu < 12.5 + 2.26\frac{1.58}{\sqrt{10}} \Longrightarrow IC_{95\%}(\mu) = (11.36892, 13.63108)$$

12 є (11.36, 13.63) no se rechaza Ho

Sea una variable aleatoria X $\in$ N(μ, σ^2) y una m.a.s. $(x_1,...,x_n)$.

c) Intervalo de confianza para la VARIANZA de una población normal. El estadístico que consideraremos es

$$\frac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_{n-1}$$ Chi-cuadrado positiva y asimétrica

Buscamos en Chi-cuadrado un valor de $\chi^2_{n-1,\ \alpha/2}$ tal que: $P\left(\chi^2_{n-1,\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha$

Despejamos el parámetro σ^2.

Por tanto el IC para la varianza a nivel (1-α) es:

$$IC_\alpha(\sigma^2) = \left[\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}\right]$$

Intervalo de confianza para la VARIANZA de una población normal.

Queremos realizar el siguiente contraste unilateral.

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 < \sigma_0^2$$

Por tanto, si $\chi^2_{OBS} = \dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,\alpha}$ se rechaza H_0

Equivalentemente, si $p_{valor} = P\left(\chi^2_{n-1} < \chi^2_{OBS}\right) < \alpha$ se rechaza H_0

Tomamos una muestra de tamaño n grande y calculamos la proporción de cierta característica como los elementos r que la presentan $p = \frac{r}{n}$

d) Intervalo de confianza para la PROPORCIÓN con n grande. Sabemos que la distribución muestral de las proporciones es

$$\frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \rightarrow N(0,1)$$

Y por tanto, el intervalo de confianza a nivel 1-α es:

$$IC_{\alpha}(\pi) = \left[p - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}};\ p + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} \right]$$

para π = p

ó si **error máximo** para π = 0.5

En R al intervalo le realiza un ajuste, luego NO se obtiene el mismo valor, pero si parecido.

Intervalo de confianza para la PROPORCIÓN con n grande.

Queremos realizar el siguiente contraste bilateral.

$$H_0 : \pi = \pi_0$$
$$H_1 : \pi \neq \pi_0$$

Por tanto, si $z_{OBS} = \dfrac{p-\pi_0}{\sqrt{\pi_0\left(1-\pi_0\right)/n}} > z_{\alpha/2}$ ó si $z_{OBS} = \dfrac{p-\pi_0}{\sqrt{\pi_0\left(1-\pi_0\right)/n}} < -z_{\alpha/2}$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = P\left(\left|Z\right| > z_{OBS}\right) = 2P\left(Z > \left| z_{OBS} \right|\right) < \alpha$$ se rechaza H_0

En R aparece $\chi^2_{OBS} = z^2_{OBS}$	Por lo que habría que hacer la raíz para obtener el valor de z_{OBS}

Ejemplo:

Una muestra aleatoria de 1000 personas obtiene que 250 son usuarias de un servicio.
1, 0, 1, 0, 0, 0, …, 1 En nuestro caso: 250 usuarios (los 1) y 750 no usuarios (los 0)

a) Obtenga el intervalo de confianza del 95% para la verdadera proporción de usuarios.

b) Contraste la hipótesis nula de que la proporción es 0.27 frente a la alternativa menor que 0.27, al nivel de significación del 3%.

$$\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \to N(0{,}1)$$

a)

$$\frac{250}{1000} - 1{,}96\sqrt{\frac{0{,}25\times 0{,}75}{1000}} < \pi < \frac{250}{100} + 1{,}96\sqrt{\frac{0{,}25\times 0{,}75}{1000}}$$

b)

$$\begin{cases} H_0 : \pi = 0.27 \\ H_1 : \pi < 0.27 \end{cases}$$

$$\boxed{z_{obs} = \frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}25-0{,}27}{\sqrt{\frac{0{,}27\times(1-0{,}27)}{1000}}} = -1.425}$$

$z_{obs} = -1.425 > -1.8807 = -z_{0.03}$ no rechazar Ho

p-valor = $P(Z < -1.425) = P(Z > 1.425) = 0.0771 > 0.03 = \alpha$ no rechazar Ho

3. Dos muestras. Media, Varianza y Proporción
A),B),C)yD) E) F)

Sean dos v.a. y dos m.a.s. $(x_1,...,x_n)$ e $(y_1,...,y_n)$ dependientes.

A) Comparación de MEDIAS **con varianzas desconocidas** para muestras normales **dependientes ó pareadas**. Definimos las diferencias $(x_1-y_1,...,x_n-y_n)$, y obtenemos la v.a: $D = X - Y$. Por tanto el estadístico y su pivote son:

$$\bar{D} = \bar{X} - \bar{Y} \qquad \frac{\bar{D} - (\mu_X - \mu_Y)}{\frac{S_D}{\sqrt{n}}} \sim t_{n-1}$$

Por tanto el IC para la diferencia de medias a nivel (1-α) es:

$$\bar{d} - t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}} < \mu_X - \mu_Y < \bar{d} + t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

Comparación de MEDIAS **con varianzas desconocidas** para muestras normales **dependientes ó pareadas**.

Queremos realizar el siguiente contraste bilateral.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

Por tanto, si $|t_{OBS}| = \left| \dfrac{\bar{d} - \bar{d}_0}{s_d / \sqrt{n}} \right| > t_{n-1,\alpha/2}$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = P\left(|t_{n-1}| > t_{OBS}\right) = 2P\left(t_{n-1} > |t_{OBS}|\right) < \alpha$$ se rechaza H_0

Las muestras <u>pareadas</u> se distribuyen **como Normal** si:
* Muestras grandes, puesto que $S_D = \sigma_D$
* Varianzas conocidas, siendo $S_D = \sigma_D = \sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$

Ejemplo:

Se elige una muestra de 15 consumidores que evalúan dos productos. Se pide a cada participante que evalúe el producto A y el producto B (en una escala de 0 a 10). Se obtienen los siguientes resultados:

Consumidor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Producto A	2	5	6	6	3	4	1	6	7	4	7	6	6	3	**5**
Producto B	4	8	6	3	7	1	6	5	3	8	9	6	3	7	6

Suponga que las valoraciones de ambos productos se distribuyen normalmente.

a) Determine un intervalo de confianza al 95% para la diferencia (producto A – producto B) entre las valoraciones medias de ambos productos.

b) Al nivel de significación del 5%, ¿se puede afirmar que las valoraciones medias de los productos son significativamente distintas?

$$\frac{\bar{D}-(\mu_X-\mu_Y)}{\frac{S_D}{\sqrt{n}}} \sim t_{n-1}$$

a) $$-0.733 - 2.144787\frac{3.011091}{\sqrt{15}} < \mu_D < -0.733 + 2.144787\frac{3.011091}{\sqrt{15}}$$

$$\mathbf{I.C._{95\%}(\mu_A-\mu_B) = (-2.40,\ \ 0.93)}$$

b) $$\boxed{\begin{cases} H_0: \mu_A - \mu_B = 0 \\ H_1: \mu_A - \mu_B \neq 0 \end{cases}}$$

$$t_{obs} = \frac{-0.733}{\frac{3.011091}{\sqrt{15}}} = -0.943$$

$|\,t_{obs}| = 0.9432 < t_{15-1,0.025} = 2.144787$ **No rechazar H_0**

p-valor = $2\cdot P(t_{14} > |-0.943|) = 0.36 > 0.05 = \alpha$ No rechazar H_0

Sean dos v.a. y dos m.a.s. $(x_1,...,x_m)$ e $(y_1,...,y_n)$ independientes.

B) Comparación de MEDIAS con **varianzas conocidas** para poblaciones normales e independientes o con tamaños muestrales grandes**.**

El estadístico a considerar es: $$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\left(\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)}} \rightarrow N(0,1)$$

Queremos realizar el siguiente contraste bilateral.

$$\boxed{\begin{array}{l} H_0: \mu_X = \mu_Y \\ H_1: \mu_X \neq \mu_Y \end{array}}$$

Si $$|z_{OBS}| = \frac{\left|(\bar{x}-\bar{y})-(\mu_{X0}\mu_{Y0})\right|}{\sqrt{\left(\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)}} > z_{\alpha/2}$$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = P\left(|Z| > z_{OBS}\right) = 2P\left(Z > |z_{OBS}|\right) < \alpha$$ se rechaza H_0

Cuando n es grande $S_X^2 = \sigma_X^2$, para cualquier distribución de X (normal o no) e independientemente de si son varianzas iguales o distintas.

Sean dos v.a. N(μ_i, σ_i^2) y dos m.a.s. ($x_1,\ldots,x_m$) e ($y_1,\ldots,y_n$) independientes.

C) Comparación de MEDIAS con **varianzas desconocidas e iguales** para poblaciones normales e independientes**.** El estadístico a considerar es:

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_C\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}} \to t_{m+n-2} \qquad s_c^2=\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}$$

Queremos realizar el siguiente contraste bilateral.

$$\boxed{\begin{array}{l} H_0: \mu_{\mathrm{X}}=\mu_{\mathrm{Y}} \\ H_1: \mu_{\mathrm{X}}\neq\mu_{\mathrm{Y}} \end{array}}$$

Si $|t_{OBS}|=\dfrac{\left|(\bar{x}-\bar{y})-(\mu_{X0}\mu_{Y0})\right|}{s_c\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}>t_{m+n-2,\alpha/2}$ se rechaza H_0

Equivalentemente, si

$p_{valor}=P\left(\left|t_{m+n-2}\right|>t_{OBS}\right)=2P\left(t_{m+n-2}>|t_{OBS}|\right)<\alpha$ se rechaza H_0

Ejemplo:

Se eligen dos muestras, cada una de 15 consumidores, para probar dos productos A y B. Se pide a cada participante que evalúe un único producto que evalúa en una escala de 0 a 10. Se obtienen los siguientes resultados:

Producto A	2	5	6	6	3	4	1	6	7	4	7	6	6	3	5
Producto B	4	8	6	3	7	1	6	5	3	8	9	6	3	7	6

Suponga que las valoraciones de ambos productos se distribuyen normalmente con la misma varianza.
a) Determine un intervalo de confianza al 95% para la diferencia (producto A – producto B) entre las valoraciones medias de ambos productos.
b) Al nivel de significación del 5%, ¿se puede afirmar que la evaluación medial del producto A es peor que la del B?

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_c\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}} \to t_{m+n-2}$$

a)

$$s_c^2 = \frac{(15-1)\times 1.830951^2 + (15-1)\times 2.263583^2}{15+15-2} = 4.238$$

$$(4.733 - 5.466) \pm 2.048407\sqrt{4.238}\sqrt{\frac{1}{15}+\frac{1}{15}} \Rightarrow$$

I.C.$_{95\%}$(μ_X-μ_Y) = (-2.273, 0.806)

b)

$$\begin{cases} H_0: \mu_A - \mu_B = 0 \\ H_1: \mu_A - \mu_B < 0 \end{cases}$$

$$t_{OBS} = \frac{(4.733-5.466)-0}{\sqrt{4.238\left(\frac{1}{15}+\frac{1}{15}\right)}} = -0.9755$$

t_{OBS} = - 0.9755 < -$t_{28,0.05}$ = -1.701131 No se rechaza H_0

p-valor = P(t_{28} <-0,9755) = 0.1688 > 0.05 = α No se rechaza H_0

Sean dos v.a. y dos m.a.s. $(x_1,\ldots,x_m)$ e $(y_1,\ldots,y_n)$ independientes.

D) Comparación de MEDIAS con **varianzas desconocidas y distintas** para poblaciones normales e independientes ó con tamaños muestrales grandes**.** El estadístico a considerar es:

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\dfrac{S_X^2}{m}+\dfrac{S_Y^2}{n}}} \to t_k$$

k aprox. de Welch

$$k=\frac{\left(s_x^2/m+s_y^2/n\right)^2}{\dfrac{\left(s_x^2/m\right)^2}{m-1}+\dfrac{\left(s_y^2/n\right)^2}{n-1}}$$

Queremos realizar el siguiente contraste bilateral.

$$\begin{aligned} H_0 &: \mu_X=\mu_Y \\ H_1 &: \mu_X\neq\mu_Y \end{aligned}$$

Si $\left|\dfrac{(\bar{x}-\bar{y})-(\mu_{X0}\,\mu_{Y0})}{\sqrt{\dfrac{s_x^2}{m}+\dfrac{s_y^2}{n}}}\right| > t_{k,\alpha/2}$ se rechaza H_0

Equivalentemente, si

$p_{valor}=P\left(|t_k|>t_{OBS}\right)=2P\left(t_k>|t_{OBS}|\right)<\alpha$ se rechaza H_0

Cuando n es grande $S_X^2=\sigma_X^2$, para cualquier distribución de X (normal o no).
Y es que la t-student se comporta como una normal para n grande.

Sean dos v.a. N(μ_i, σ_i^2) y dos m.a.s. (x_1,...,x_m) e (y_1,...,y_n) independientes.

E) Comparación de VARIANZAS para poblaciones normales e independientes**.** El estadístico a considerar es:

$$S_X^2 > S_Y^2 \rightarrow \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} = \frac{S_X^2}{S_Y^2}\frac{\sigma_Y^2}{\sigma_X^2} \rightarrow F_{m-1,n-1}$$

F-Snedecor positiva y asimétrica

Queremos realizar el siguiente contraste bilateral.

$$H_0 : \sigma_X^2 = \sigma_Y^2 \Rightarrow \sigma_X^2 / \sigma_Y^2 = 1$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2 \Rightarrow \sigma_X^2 / \sigma_Y^2 \neq 1$$

Por tanto, si $F_{OBS} = \frac{s_x^2 / \sigma_{X0}^2}{s_y^2 / \sigma_{Y0}^2} < F_{m-1,n-1,\alpha/2}$ ó si $F_{OBS} = \frac{s_x^2 / \sigma_{X0}^2}{s_y^2 / \sigma_{Y0}^2} > F_{m-1,n-1,1-\alpha/2}$ se rechaza H_0

Equivalentemente, si

$$p_{valor} = 2 * \min\left\{P\left(F_{m-1,n-1} < F_{OBS}\right), P\left(F_{m-1,n-1} > F_{OBS}\right)\right\} < \alpha$$ se rechaza H_0

Comparación de VARIANZAS para poblaciones normales e independientes**.**

Queremos realizar el siguiente contraste unilateral.

$$H_0: \sigma_X^2 \le \sigma_Y^2 \Rightarrow \sigma_X^2 / \sigma_Y^2 \le 1$$

$$H_1: \sigma_X^2 > \sigma_Y^2 \Rightarrow \sigma_X^2 / \sigma_Y^2 > 1$$

Por tanto, si $F_{OBS} = \dfrac{s_x^2 / \sigma_{X0}^2}{s_y^2 / \sigma_{Y0}^2} > F_{m-1,n-1,\alpha}$ se rechaza H_0

Equivalentemente, si $p_{valor} = P\left(F_{m-1,n-1} > F_{OBS}\right) < \alpha$ se rechaza H_0

En R hace (factor1/factor2) luego si se pide el contraste inverso tendremos que ***reodenar*** *los niveles del factor*.

Ejemplo:

Se eligen dos muestras, cada una de 15 consumidores, para probar dos productos A y B. Se pide a cada participante que evalúe un único producto que evalúa en una escala de 0 a 10. Se obtienen los siguientes resultados:

Producto A	2	5	6	6	3	4	1	6	7	4	7	6	6	3	5
Producto B	4	8	6	3	7	1	6	5	3	8	9	6	3	7	6

Suponga que las valoraciones de ambos productos se distribuyen normalmente. Al nivel de significación del 5%, ¿se puede afirmar que las varianzas de las valoraciones de los productos son significativamente distintas?

$$\frac{S_X^2}{S_Y^2}\frac{\sigma_Y^2}{\sigma_X^2} \to F_{m-1,n-1}$$

Lo hacemos haría R por defecto:

$$\begin{cases} H_0 : \sigma_A^2 / \sigma_B^2 = 1 \\ H_1 : \sigma_A^2 / \sigma_B^2 \neq 1 \end{cases}$$

$$\boxed{F_{obs} = \frac{s_A^2}{s_B^2}\frac{\sigma_B^2}{\sigma_A^2} = \frac{1.830951^2}{2.263583^2} = 0.6543}$$

$$F_{14,14,0.025} = 0,33573$$
$$F_{14,14,0.975} = 2,97859$$

$0.33573 < F_{obs} = 0.6543 < 2.97859$ No rechazar H_0

p-valor = $2 \cdot \min\{P(F_{14,14} > 0.6543), P(F_{14,14} < 0.6543)\} =$
$= 2 \cdot 0.2186471 = 0.437 > 0.05$ No rechazar H_0

Sean dos v.a. y dos m.a.s. $(x_1,...,x_m)$ e $(y_1,...,y_n)$ independientes.

F) Comparación de PROPORCIONES para tamaños muestrales grandes**.**

El estadístico a considerar es:

$$\frac{(p_X - p_Y) - (\pi_X - \pi_Y)}{\sqrt{\frac{\pi_X(1-\pi_X)}{n_X} + \frac{\pi_Y(1-\pi_Y)}{n_Y}}} \to N(0,1)$$

Para $\pi_X = p_X$; $\pi_Y = p_Y$

ó si **error máximo**

para $\pi_X = 0.5$; $\pi_Y = 0.5$

Queremos realizar el siguiente contraste bilateral. $\begin{cases} H_0 : \pi_X = \pi_Y \\ H_1 : \pi_X \neq \pi_Y \end{cases}$

Por tanto, si $z_{OBS} = \left| \frac{(p_X - p_Y) - (\pi_{X0} - \pi_{Y0})}{\sqrt{\frac{p_0(1-p_0)}{n_X} + \frac{p_0(1-p_0)}{n_Y}}} \right| > z_{\alpha/2}$ se rechaza H_0

$$p_0 = \frac{p_x n_x + p_Y n_Y}{n_X + n_Y}$$

Equivalentemente, si

$$p_{valor} = P(|Z| > z_{OBS}) = 2P(Z > | z_{OBS} |) < \alpha$$ se rechaza H_0

Ejemplo:

Para analizar la efectividad de una campaña electoral se tomó una muestra aleatoria de 1000 votantes antes del inicio obteniendo que 300 declaraban su intención de votar al partido A. Una muestra aleatoria de 800 votantes realizada al final de la campaña obtuvo que 280 declaraban su intención de votar al partido A.

a) Obtenga un intervalo de confianza del 90% para la diferencia de proporción de votantes al partido A antes y después de la campaña.

b) Contraste la hipótesis de la proporción de votantes al partido A no ha variado con la campaña electoral, al nivel de significación del 5%.

$$\frac{(p_X - p_Y) - (\pi_X - \pi_Y)}{\sqrt{\frac{\pi_X(1-\pi_X)}{n_X} + \frac{\pi_Y(1-\pi_Y)}{n_Y}}} \to N(0,1)$$

a) $$(0.3 - 0.35) \pm 1{,}645\sqrt{\frac{0.3(1-0.3)}{1000} + \frac{0.35(1-0.35)}{800}}$$

$$\mathbf{I.C._{90\%}(\pi_X - \pi_Y) = (-0.08657,\ \ -0.013)}$$

b) $$\begin{cases} H_0 : \pi_X = \pi_Y \\ H_1 : \pi_X \neq \pi_Y \end{cases}$$

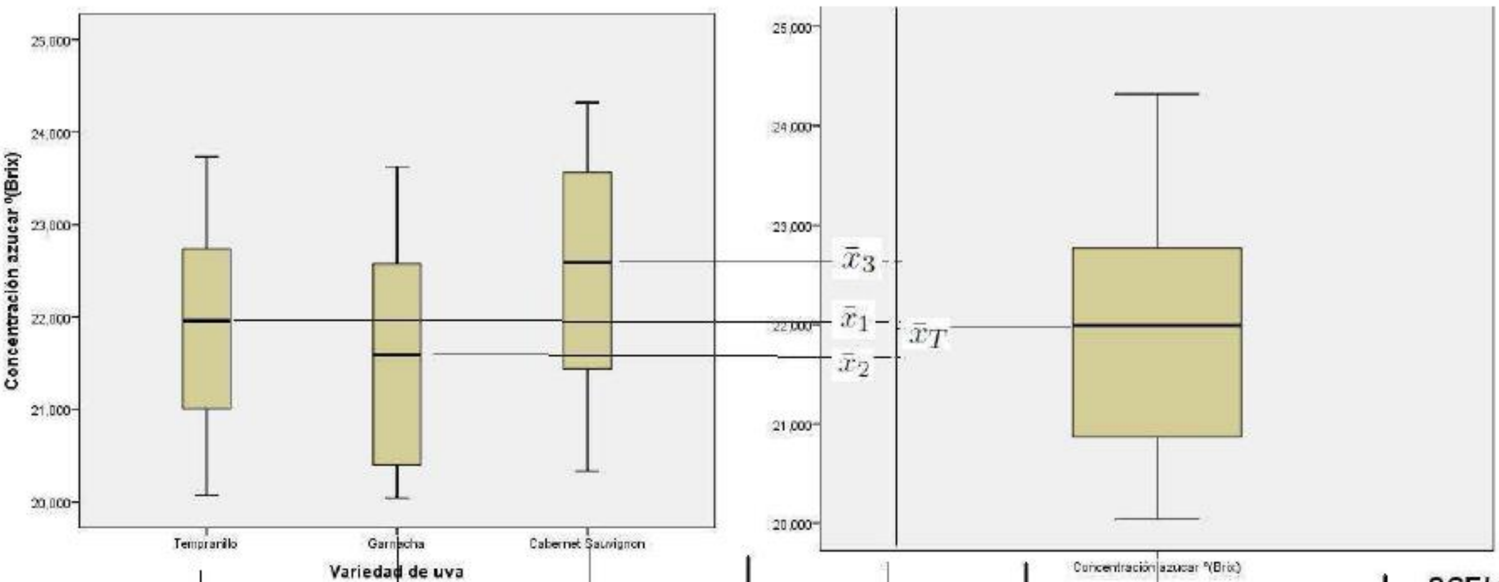
$$z_{obs} = \frac{(0.3 - 0.35)}{\sqrt{\frac{0.32222(1-0.32222)}{1000} + \frac{0.32222(1-0.32222)}{800}}} = -2.25$$

$$p_0 = \frac{300+280}{1000+800} = 0.32222$$

$$\mathbf{|\,z_{OBS}| = 2.25 > z_{\alpha/2} = 1.96}$$ **Rechazar H_0**

p-valor = 2 · P(Z > |-2.25|) = 2 · 0.01222 = 0.024 < 0.05 Rechazar H_0

4. Tabla ANOVA



Variabilidad dentro de la muestra 1 sirve para estimar σ^2

Variabilidad dentro de la muestra 2 sirve para estimar σ^2

Variabilidad dentro de la muestra 3 sirve para estimar σ^2

Con las tres se estima σ^2 variabilidad inevitable, la mínima.

SCD=116.003

Las diferencias entre las medias de cada grupo y la media total $\bar{x}_1-\bar{x}_T, \bar{x}_2-\bar{x}_T, \bar{x}_3-\bar{x}_T$ permiten estimar la variabilidad entre los grupos

SCE=12.992

Variabilidad de toda la muestra

Variabilidad total

SCT=128.995

$$\frac{SCE/\ r\text{-}1}{SCD/\ n\text{-}r}$$

Si H0 es cierta este valor debe de ser pequeño.

Cuando toma un valor demasiado grande se rechaza H0

Si H0 es cierta la distribución es:

$$\frac{SCE/\ r\text{-}1}{SCD/\ n\text{-}r} = F_{r-1,\, n-r}$$

SCD + SCE = SCT

ESTADISTICOS

Una población

- Para la media μ

Situación	Estadístico y distribución	Requisitos de uso
Con σ^2 conocida	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z$	X normal o n grande
Con σ^2 desconocida	$\frac{\bar{X}-\mu}{S/\sqrt{n}} = t_{n-1}$	X normal
Con σ^2 desconocida	$\frac{\bar{X}-\mu}{S/\sqrt{n}} = Z$	n grande

- Para la varianza σ^2

Estadístico y distribución	Requisitos de uso
$\frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1}$	X normal

- Para la proporción π

Estadístico y distribución	Requisitos de uso
$\frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} = Z$	n grande

Para comparar dos poblaciones

- Para la diferencia de medias $\mu_1 - \mu_2$

Situación	Estadístico y distribución	Requisitos de uso
σ_1^2 y σ_2^2 conocidas	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = Z$	X normal o n_1, n_2 grandes
σ_1^2 y σ_2^2 desconocidas pero iguales	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = t_{n_1+n_2-2}$	X normal
σ_1^2 y σ_2^2 desconocidas	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = Z$	n_1, n_2 grandes

- Para la comparación de varianzas σ_1^2/σ_2^2

Estadístico y distribución	Requisitos de uso
$\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} = F_{n_1-1,n_2-1}$	X_1, X_2 normales

- Para la diferencia de proporciones $\pi_1 - \pi_2$

Estadístico y distribución	Requisitos de uso
$\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}} = Z$	n_1, n_2 grandes