Diferencia de Medias Muestrales

Se presenta el caso en que es necesario estimar la diferencia de las medias poblacionales. Si se cuenta con dos muestras, se pueden presentar varios casos:

a) Caso σ_1 y σ_2 son conocidos:

Se puede construir la variable normalizada *Z*:

$$\boxed{Z = \frac{(\overline{x_1} - \overline{x_2}) - \mu_{\overline{x_1}-\overline{x_2}}}{\sigma_{\overline{x_1}-\overline{x_2}}} = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1}\sigma_1^2 + \frac{1}{n_2}\sigma_2^2}}} \longrightarrow \quad \sim N(0,1)$$

Visto anteriormente

Se puede plantear que las medias poblacionales sean iguales $(\mu_1 - \mu_2) = 0$ o que las mismas difieran en una cantidad *Y* : $(\mu_1 - \mu_2) = Y$

Para el método de estimación por IC: $\quad P\left(-Z\alpha_{/2} < Z < Z\alpha_{/2}\right) = 1 - \alpha$

$$P\left((\overline{x}_1 - \overline{x}_2) - Z\alpha_{/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + Z\alpha_{/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 1 - \alpha$$

b) Caso σ_1 y σ_2 son desconocidos y se sabe que $\sigma_1 \neq \sigma_2$:

$$\boxed{T = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} S_1^2 + \frac{1}{n_2} S_2^2}}} \longrightarrow \quad \sim t(0, \nu)$$

El problema radica en elegir cuál de los ν utilizar ya que tenemos dos muestras:

ν_1 y ν_2

Se utiliza un valor de grados de libertad ponderado entre ambas muestras:

$$\nu_p = \frac{\left({S_1^2}/{n_1} + {S_2^2}/{n_2} \right)^2}{\frac{\left({S_1^2}/{n_1} \right)^2}{\nu_1} + \frac{\left({S_2^2}/{n_2} \right)^2}{\nu_2}}$$

tomar el valor entero más próximo

Para el método de estimación por IC:

$$P\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) = 1 - \alpha$$

c) Caso σ_1 y σ_2 son desconocidos pero $\sigma_1 = \sigma_2$:

También utilizamos la $\sim t(0, \nu)$: Construimos una variable $T = \dfrac{\overline{x} - \mu}{S/\sqrt{n}} = \dfrac{Z}{\sqrt{V/\nu}}$

con $V = \dfrac{\nu . S^2}{\sigma^2}$ Que sigue una $\sim \chi^2(\nu, \alpha)$. En este caso de diferencia de medias

$V = \dfrac{\nu_1 S_1^2}{\sigma_1^2} + \dfrac{\nu_2 S_2^2}{\sigma_2^2}$ y $\nu = \nu_1 + \nu_2$ De esta forma:

$$T = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\cancel{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} . \frac{1}{\underbrace{\sqrt{\left(\dfrac{\nu_1 S_1^2}{\cancel{\sigma^2}} + \dfrac{\nu_2 S_2^2}{\cancel{\sigma^2}}\right) . \dfrac{1}{\nu_1 + \nu_2}}}_{S_p}}$$

$$T = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

Con: $S_p^2 = \dfrac{\nu_1 S_1^2 + \nu_2 S_2^2}{\nu_1 + \nu_2}$

Para el método de estimación por IC:

$$P\left((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

Ejemplo 2: Dos muestras de la concentración de Fósforo en el río se toman en dos lugares diferentes, uno corriente arriba y otro corriente debajo de una fábrica química.

a) Encontrar un intervalo de confianza del 95% para $(\mu_1 - \mu_2)$ suponiendo $\sim N(\mu, \sigma)$ y $\sigma_1 \neq \sigma_2$

b) Comparar la hipótesis $\mu_1 = \mu_2$ con un 95% de confianza.

		n	$\overline{x}\ [mg/l]$	$S\ [mg/l]$
M 1	Corriente abajo	15	3,84	3,07
M 2	Corriente arriba	12	1,49	0,80

¿ A qué caso de los vistos corresponde ?

Corresponde al caso (b) σ_1 y σ_2 son desconocidos y se sabe que $\sigma_1 \neq \sigma_2$

$$P\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}} \right) = 1 - \alpha$$

Calculo el grado de libertad ponderado:

$$\nu_p = \frac{\left({S_1}^2/n_1 + {S_2}^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{\nu_1} + \frac{\left(S_2^2/n_2\right)^2}{\nu_2}} = 16{,}3 \cong 16$$

$$\bar{x}_1 - \bar{x}_2 = 3{,}48 - 1{,}49 = 2{,}35\left[{mg}/{l}\right] \quad ; \quad \sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}} = 0{,}826\left[{mg}/{l}\right]$$

$t_{\alpha/2} = t_{0{,}025} = 2{,}12$ con $\nu_p = 16$ Tabla A.4

$\boxed{IC = [0{,}6 < \mu_1 - \mu_2 < 4{,}10]}$ Con un 95% de confianza

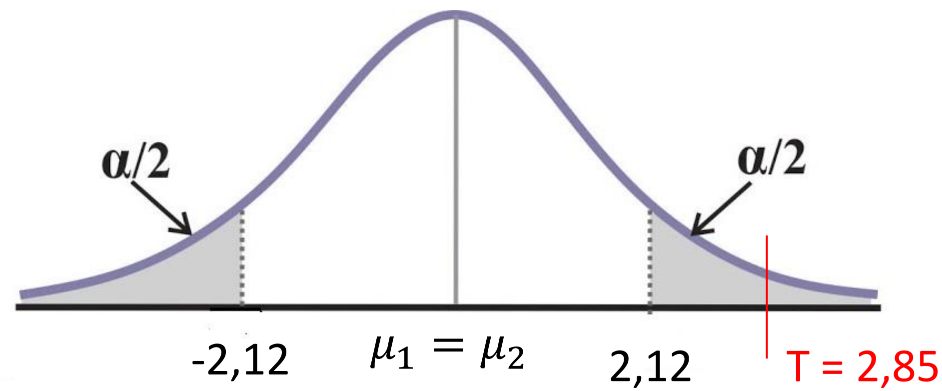
b) Comparar la hipótesis $\mu_1 = \mu_2$ con un 95% de confianza.

Estimación puntual

$$T = \frac{(\overline{x_1} - \overline{x_2}) - \cancel{(\mu_1 - \mu_2)}}{\sqrt{\frac{1}{n_1}S_1^2 + \frac{1}{n_2}S_2^2}} \qquad \boxed{=0}$$

$$T = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{1}{n_1}S_1^2 + \frac{1}{n_2}S_2^2}} = \frac{2{,}35}{0{,}826} = 2{,}845$$

$$P\left(-t\alpha_{/2} < T < t\alpha_{/2}\right) = 0{,}95$$



$$P(-2{,}12 < 2{,}845 < 2{,}12) = 0{,}95$$ → 2,845 está fuera del intervalo

$\boxed{\text{No puede afirmarse: } \mu_1 = \mu_2}$

CONTAMINA

Ahora vemos para qué sirve la estimación por IC y la Estimación Puntual

Estimación de proporciones

Muy utilizado en encuestas

Sea un experimento binomial → variable dicotómica $\begin{cases} \text{Éxito} \\ \text{Fracaso} \end{cases}$

Un estimador puntual de V.A. $x \sim b(x; n, p) = \binom{n}{x} p^x q^{n-x}$ es: $\boxed{\hat{P} = \frac{x}{n}}$

Con $x =$ número de éxitos (o fracasos) y $n =$ número de pruebas.

La proporción $\hat{P} = {}^{x}/_{n}$ de la muestra de V.A. x se utilizará como estimador puntual de p que representa la proporción de la población.

Si no se espera que p (desconocida) esté demasiado cerca de 0 (fracaso) o 1 (éxito) se puede establecer un IC considerando la distribución muestral $\hat{P}$.

¿ Por qué ponemos esta restricción ?

Si $n \to \infty$ vimos que $\sim b \ \to \sim N(\mu, \sigma)$: $\longrightarrow$ $Z = \dfrac{\hat{P} - \mu_{\hat{P}}}{\sigma_{\hat{P}}}$

$$\mu_{\hat{P}} = E(\hat{P}) = E\left(\frac{x}{n}\right) = \frac{1}{n} E(x) = \frac{1}{n}.(np) = p$$

$$\sigma_{\hat{P}}^2 = \sigma_{x/n}^2 = \frac{1}{n^2}\sigma_x^2 = \frac{npq}{n^2} = \frac{pq}{n}$$

$$V(ax+b) = a^2 V(x)$$

Usando la aproximación normal de la binomial:

$$\boxed{P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) = 1 - \alpha}$$

Con

$$\boxed{Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}}}$$

Para construir un estimador por IC:

$$P\left(\hat{P} - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{P} + Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

Problema !!! Los extremos del intervalo dependen de p (desconocida)

Si $n \to \infty$ el error de sustituir $\hat{P}$ por p es pequeño.

$$\boxed{P\left(\hat{P} - Z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}} < p < \hat{P} + Z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}}\right) = 1 - \alpha}$$

Con las condiciones: n grande

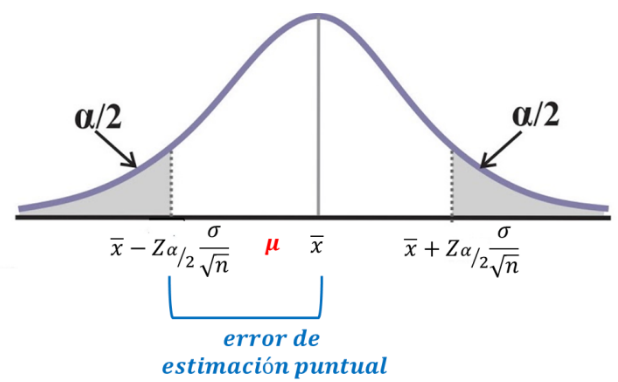
$p \gg 0$ y $p \ll 1$

Al igual que en los casos anteriores de estimación puntual de parámetros poblacionales: $\hat{P} \neq p$ y por lo tanto se comete un error de estimación puntual.

$$Error = \left|p - \hat{P}\right|$$

Con el método de estimación por IC se puede calcular el Error de Estimación Puntual con un nivel de confianza $(1-\alpha)$.

$$e = Z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}}$$



Determinemos ahora el tamaño de la muestra para asegurar, con un $(1-\alpha)\%$, un determinado valor de e:

$$n = \left(\frac{Z_{\alpha/2}}{e}\right)^2 \hat{P}\hat{Q}$$

Aquí el problema es que $\hat{P}$ depende de n.

$$\boxed{\hat{P} = \frac{x}{n}}$$

Se toma primeramente una muestra de $n \geq 30$ para calcular $\hat{P}$, luego con $\hat{P}$ se estima n para obtener un error especificado.

Dos muestras: Estimación de la diferencia de proporciones

Estimar la diferencia de dos proporciones p_1 y p_2

Se calcula $\hat{P}_1 = \frac{x_1}{n_1}$ y $\hat{P}_2 = \frac{x_2}{n_2}$ a partir de dos muestras de tamaño n_1 y n_2

La diferencia de proporciones muestrales $\hat{P}_1 - \hat{P}_2$ se utilizará para estimar la diferencia de proporciones poblacionales $p_1 - p_2$

Al ser $\hat{P}_1$ y $\hat{P}_2$ independientes, es posible utilizar la $\sim N(\mu, \sigma)$ para la diferencia.

$$\boxed{\mu_{\hat{P}_1 - \hat{P}_2} = p_1 - p_2}$$

$$\boxed{\sigma^2_{\hat{P}_1 - \hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Por lo tanto:

$$P\left(-Z\alpha_{/2} < Z < Z\alpha_{/2}\right) = 1 - \alpha \quad \text{con} \quad Z = \frac{\left(\hat{P}_1 - \hat{P}_2\right) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Para construir IC se sustituye Z

$$\boxed{P\left(\left(\hat{P}_1 - \hat{P}_2\right) - Z\alpha_{/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < p_1 - p_2 < \left(\hat{P}_1 - \hat{P}_2\right) + Z\alpha_{/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) = 1 - \alpha}$$

En las que p_1, q_1, p_2 y q_2 pueden ser sustituidas por $\hat{P}_1$, $\hat{Q}_1$, $\hat{P}_2$ y $\hat{Q}_2$ siempre que:

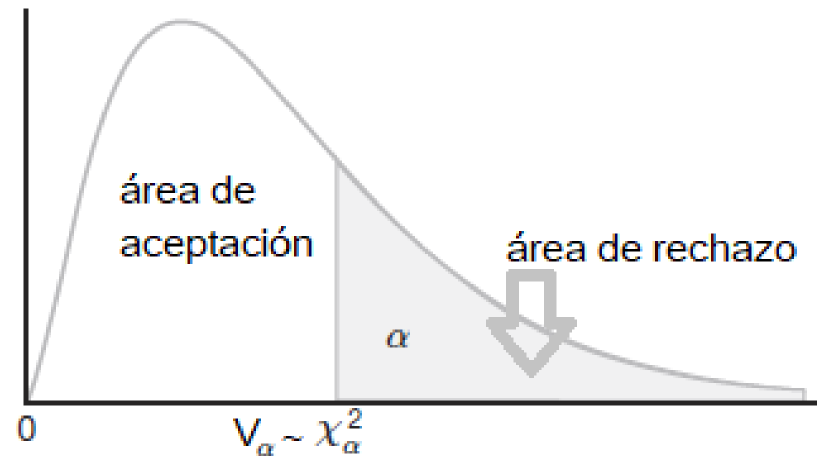
$$\boxed{\begin{array}{c} n \ \text{ grande} \\ \text{y} \qquad p \gg 0 \ \text{ y } \ p \ll 1 \end{array}}$$

Estimación de la Varianza por IC

En el método de estimación puntual habíamos visto que para estimar σ^2 a partir de S^2 se utiliza la $\sim\chi^2$ con $\nu = (n-1)$ grados de libertad.

La V.A. $V = \dfrac{\nu . S^2}{\sigma^2}$ tiene $\sim\chi$

Si $x \sim N(\mu, \sigma)$



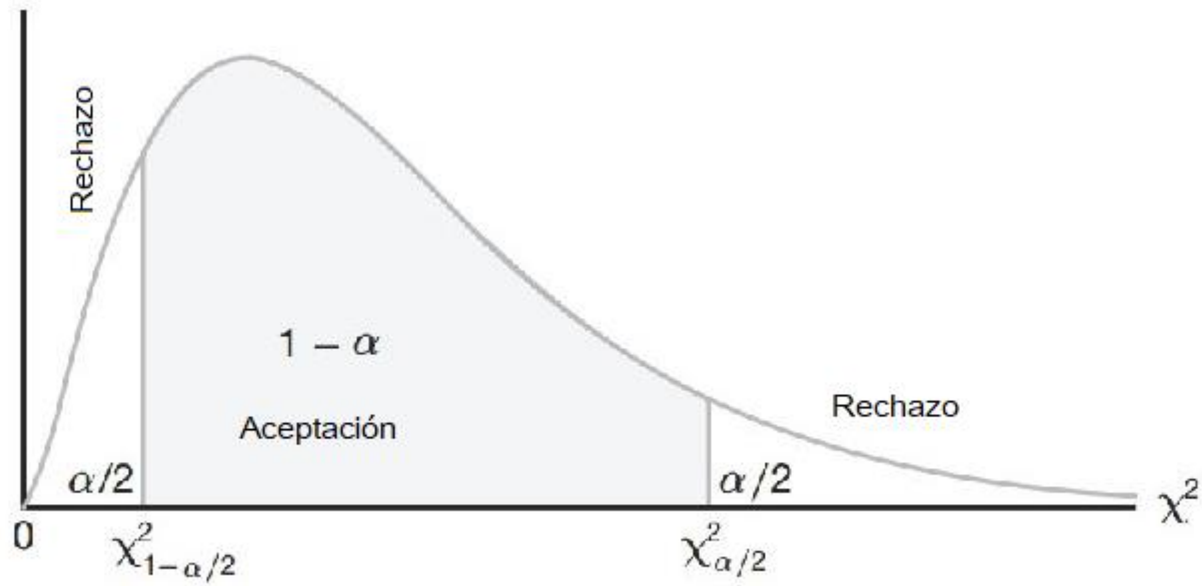
Para aceptar o rechazar la hipótesis de que la muestra de varianza S^2 provenga de una población de varianza σ^2:

1. Elegimos α.
2. Calculamos V.
3. Encontramos V_α.
4. Comparamos V con V_α.
5. Decidimos aceptar o rechazar la hipótesis.

Ahora nos proponemos estimar σ^2 a partir de S^2 con $(1-\alpha)$ nivel de confianza por IC:

Es decir, planteamos la probabilidad por IC

$$P\left(V_{1-\alpha/2} < V < V\alpha_{/2}\right) = 1 - \alpha$$ Bilateral



$$P\left(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

V

Tiene que quedar S^2 al medio

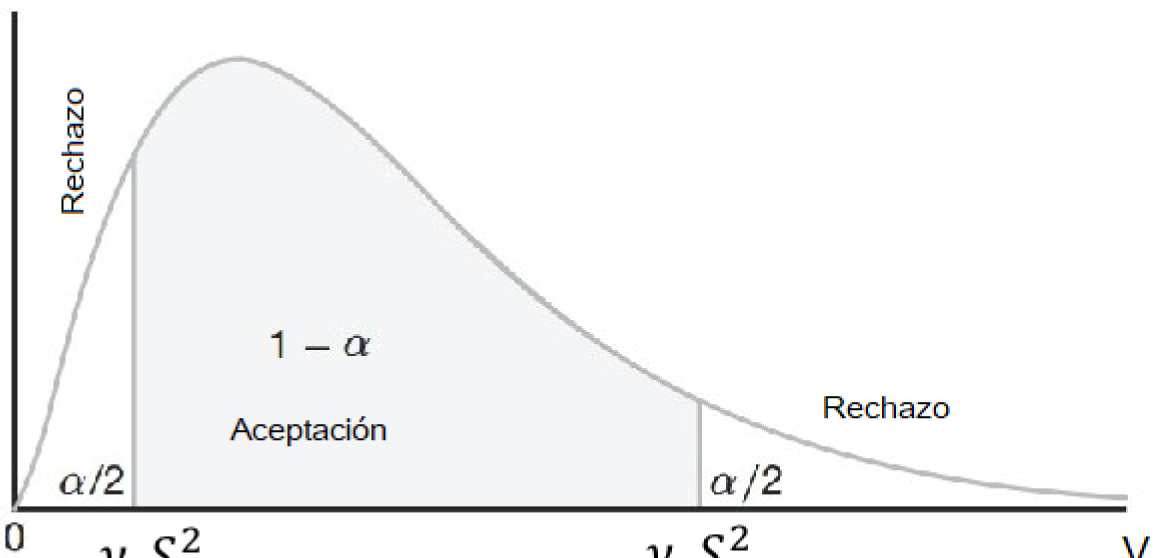
$$P\left(V_{1-\alpha/2} < \frac{\nu . S^2}{\sigma^2} < V_{\alpha/2}\right) = 1 - \alpha$$

Dividiendo m.a.m. por $\nu . S^2$

$$P\left(\frac{V_{1-\alpha/2}}{\nu . S^2} < \frac{1}{\sigma^2} < \frac{V_{\alpha/2}}{\nu . S^2}\right) = 1 - \alpha$$

Invertimos las desigualdades

$$\boxed{P\left(\frac{\nu . S^2}{V_{\alpha/2}} < \sigma^2 < \frac{\nu . S^2}{V_{1-\alpha/2}}\right) = 1 - \alpha}$$



$$\frac{\nu . S^2}{V_{\alpha/2}} \qquad \frac{\nu . S^2}{V_{1-\alpha/2}}$$

Ejemplo 3: Encuentre un IC del 95% para la varianza de todos los paquetes de semillas que se distribuyen en una empresa agrícola suponiendo distribución normal de los pesos de esos paquetes.

Pesos en Kg de una muestra de 10 paquetes: 464, 461, 458, 470, 461, 459, 458, 469, 452, 460

$$S^2 = \frac{n.\sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2}{n.(n-1)} = \frac{10.2127312 - 4612^2}{10.9} = \frac{2576}{90} = 28{,}6\hat{2}$$

Elegimos $\alpha = 0{,}05$ y de la Tabla A.5 usando $\nu = 9$ encontramos:

$V\alpha/_2 = V_{0{,}025} = 19{,}023$ y $V_{1-\alpha/_2} = V_{0{,}975} = 2{,}700$

Por lo tanto el IC es: $IC = \left[\frac{\nu.S^2}{V\alpha/_2} ; \frac{\nu.S^2}{V_{1-\alpha/_2}}\right]$ $IC = \left[\frac{9.28{,}6\hat{2}}{19{,}023} ; \frac{9.28{,}6\hat{2}}{2{,}700}\right]$

$$\boxed{IC = \left[13{,}542 \ ; \ 95{,}\widehat{407}\right]}$$

$$\boxed{P\left(13{,}542 < \sigma^2 < 95{,}\widehat{407}\right) = 0.95}$$

Dos muestras: Estimación de la razón de Varianzas por IC

Como antes, hacemos uso de la $\sim f$ para estimar la razón de varianzas

$$F = \frac{U/_{\nu_1}}{V/_{\nu_2}}$$

Con U y V dos V.A. independientes donde ambas tienen $\sim\chi^2$ con ν_1 grados de libertad para U y ν_2 grados de libertad para V.

Esta distribución nos ayuda a determinar con cierta probabilidad si S_1^2 y S_2^2 provenían de muestras de la misma población.

Ahora estamos interesados en estimar la razón de dos varianzas poblacionales σ_1^2 y σ_2^2 a partir de muestras de varianzas S_1^2 y S_2^2 con $(1-\alpha)$ nivel de confianza utilizando IC

Se plantea entonces la Probabilidad de forma bilateral:

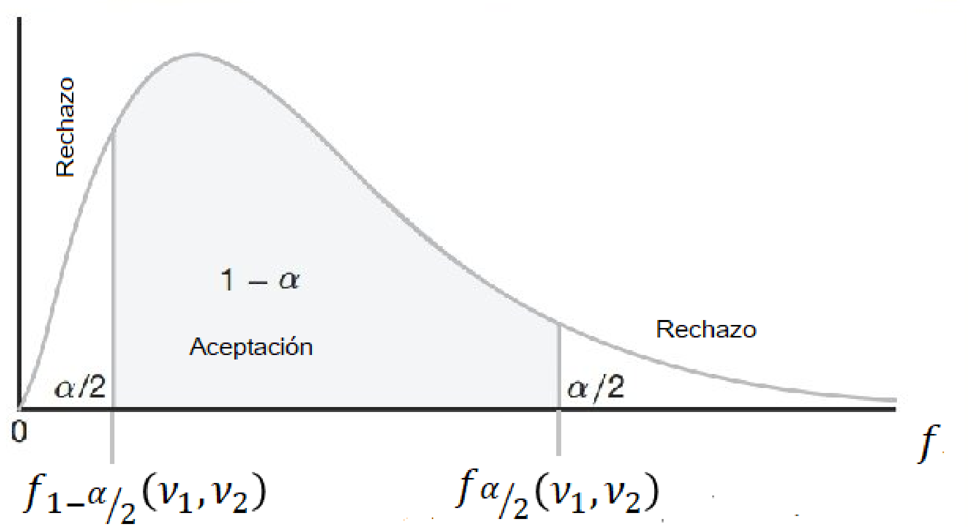
$$P\left[f_{1-\alpha/_2}(\nu_1,\nu_2) < F < f_{\alpha/_2}(\nu_1,\nu_2)\right] = 1-\alpha$$

$$P\left[f_{1-\alpha/2}(\nu_1,\nu_2) < F < f_{\alpha/2}(\nu_1,\nu_2)\right] = 1-\alpha$$ Con $$F = \frac{U/\nu_1}{V/\nu_2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 . S_1^2}{\sigma_1^2 . S_2^2}$$

$$P\left[f_{1-\alpha/2}(\nu_1,\nu_2) < \frac{\sigma_2^2 . S_1^2}{\sigma_1^2 . S_2^2} < f_{\alpha/2}(\nu_1,\nu_2)\right] = 1-\alpha$$ Multiplicando m.a.m. S_2^2/S_1^2

$$P\left[\frac{S_2^2}{S_1^2} . f_{1-\alpha/2}(\nu_1,\nu_2) < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} . f_{\alpha/2}(\nu_1,\nu_2)\right] = 1-\alpha$$ Invirtiendo las desigualdades

$$\boxed{P\left[\frac{S_1^2}{S_2^2} . \frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} . \frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)}\right] = 1-\alpha}$$



$$P\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\alpha/2}(\nu_1,\nu_2)}\right] = 1-\alpha$$

Pero también puede ser expresada de otra forma sabiendo que:

$$f_{1-\alpha}(\nu_1,\nu_2) = \frac{1}{f_\alpha(\nu_2,\nu_1)}$$

$$\boxed{P\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot f_{\alpha/2}(\nu_2,\nu_1)\right] = 1-\alpha}$$

Ejemplo 4: Del Ejemplo 2 donde dos muestras de la concentración de Fósforo en el río se toman en dos lugares diferentes, uno corriente arriba y otro corriente abajo de una fábrica química. Suponiendo que las variables de concentración de fósforo en el río siguen distribuciones normales, la razón de las varianzas muestrales S_1^2/S_2^2 es calculada y utilizada para estimar la razón de las varianzas poblacionales de concentración de fósforo por IC. Con este criterio mostrar que las varianzas poblacionales son distintas con un nivel de confianza del 98%

		n	$\overline{x}\ [mg/l]$	$s\ [mg/l]$
Muestra 1	Corriente abajo	15	3,84	3,07
Muestra 2	Corriente arriba	12	1,49	0,80

Solución: Con $(1-\alpha) = 0{,}98 \rightarrow \alpha = 0{,}02$

De Tabla A.6: $\rightarrow \begin{cases} f_{\alpha/2}(\nu_1,\nu_2) = f_{0,01}(14{,}11) = 4{,}30 \\ f_{\alpha/2}(\nu_2,\nu_1) = f_{0,01}(11{,}14) = 3{,}87 \end{cases}$

$$P\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\cdot f_{\alpha/2}(\nu_2,\nu_1)\right] = 1-\alpha$$

$$IC = \left[\frac{3{,}07^2}{0{,}80^2}\cdot\frac{1}{4{,}30} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{3{,}07^2}{0{,}80^2}\cdot 3{,}87\right] = [1{,}851; 7{,}549] \longrightarrow$$

Este IC no contiene el cociente = 1