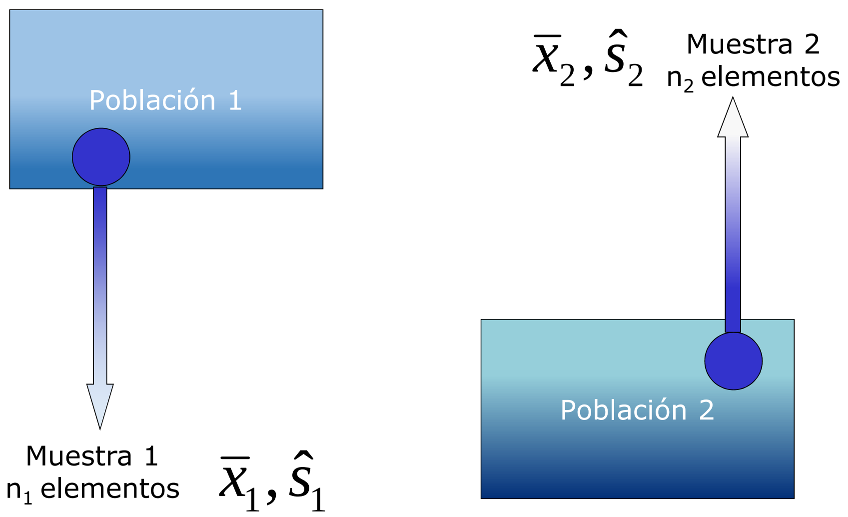
Tema 8
Comparación de poblaciones

Carlos Montes – uc3m

1. Introducción



2.1. Intervalo de confianza para la diferencia de medias con σ conocidas

Nos interesa la distribución de la variable:

$$\bar{x}_1 - \bar{x}_2$$

Con **muestras grandes** o **poblaciones normales:**

$$\frac{\bar{x}_1 - \mu_1}{\sigma / \sqrt{n_1}} \rightarrow Z \qquad \frac{\bar{x}_2 - \mu_2}{\sigma / \sqrt{n_2}} \rightarrow Z$$

$\bar{x}_1 - \bar{x}_2$ seguirá una ley Normal

2.1. Intervalo de confianza para la diferencia de medias con σ conocidas

Media: $\mu_1 - \mu_2$

Varianza: $\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}$

$$\bar{x}_1 - \bar{x}_2 \rightarrow N\left[\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}\right]$$

Entonces:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}}} \rightarrow N(0,1)$$

Análogamente a como hicimos en el caso del intervalo de confianza para la media poblacional, podemos escribir:

$$P\left(-Z_{\alpha/2} \leq \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}}} \leq Z_{\alpha/2}\right) = 1 - \alpha$$

Operando:

$$P\left((\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + \sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}\,Z_{\alpha/2}\right) = 1 - \alpha$$

Con lo que el intervalo queda:

DT estimador

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}\right\}}$$

Parámetro *Estimación* *Valor tablas*

2.1. Intervalo de confianza para la diferencia de medias con σ conocidas

Si las varianzas son iguales: $\left(\sigma_1^{\,2} = \sigma_2^{\,2}\right)$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow N(0,1)$$

Con lo que el intervalo queda:

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2}\,\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}}$$

2.2. Intervalo de confianza para la diferencia de medias con σ estimada

Con **muestras grandes**, la aproximación a la normal sigue siendo válida si sustituimos la varianza por su estimador:

Varianzas poblacionales distintas:

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{s}_1^{\,2}}{n_1} + \frac{\hat{s}_2^{\,2}}{n_2}} \right\}}$$

Varianzas poblacionales iguales:

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2}\,\hat{s}_T\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}}$$

$$\boxed{\hat{s}_T^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}}$$

Varianza experimental

Combinación más precisa de estimadores para σ^2

2.2. Intervalo de confianza para la diferencia de medias con σ estimada

¿Y si la muestra es pequeña?

Necesitamos que la población se distribuya normalmente.

Varianzas iguales

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,\alpha/2}\, \hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}}$$

$$\hat{s}_T^2 = \frac{(n_1-1)\hat{s}_1^2 + (n_2-1)\hat{s}_2^2}{n_1+n_2-2}$$

2.2. Intervalo de confianza para la diferencia de medias con σ estimada

Varianzas distintas:

$$\boxed{IC(1-\alpha): \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm t_{\nu,\alpha/2} \sqrt{\frac{\hat{s}_1^{\,2}}{n_1} + \frac{\hat{s}_2^{\,2}}{n_2}} \right\}}$$

$$\nu = \frac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2} \right)^2}{\frac{1}{n_1-1}\left(\frac{\hat{s}_1^2}{n_2} \right) + \frac{1}{n_2-1}\left(\frac{\hat{s}_2^2}{n_2} \right)}$$

Como, en general, no será entero, usamos el entero más próximo

3.1. Contraste para la diferencia de medias. Varianzas conocidas

Se desea contrastar la hipótesis de que las medias de dos poblaciones son iguales.

Muestras grandes, o **poblaciones normales**

Varianzas iguales

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow Z$$

3.1. Contraste para la diferencia de medias. Varianzas conocidas

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

Rechazamos si: $$\frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq Z_{\alpha/2} \quad \leq -Z_{\alpha/2}$$

$H_0: \mu_1 \leq \mu_2$
$H_1: \mu_1 > \mu_2$

Rechazamos si: $$\frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq Z_{\alpha}$$

$H_0: \mu_1 \geq \mu_2$
$H_1: \mu_1 < \mu_2$

Rechazamos si: $$\frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq -Z_{\alpha}$$

3.1. Contraste para la diferencia de medias. Varianzas conocidas

Varianzas distintas

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \to Z$$

Carlos Montes – uc3m

3.1. Contraste para la diferencia de medias. Varianzas conocidas

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \begin{matrix} \geq Z_{\alpha/2} \\ \leq -Z_{\alpha/2} \end{matrix}$

H_0: $\mu_1 \leq \mu_2$
H_1: $\mu_1 > \mu_2$

Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq Z_\alpha$

H_0: $\mu_1 \geq \mu_2$
H_1: $\mu_1 < \mu_2$

Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -Z_\alpha$

3.2. Contraste para la diferencia de medias. Varianzas estimadas

Con **muestras grandes,** la aproximación a la normal sigue siendo válida si sustituimos la varianza por su estimador:

Varianzas poblacionales iguales $\quad Z_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to Z$

Varianzas poblacionales distintas $\quad Z_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}} \to Z$

3.2. Contraste para la diferencia de medias. Varianzas estimadas

Y los contrastes se realizan de la misma manera:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \begin{matrix} \geq Z_{\alpha/2} \\ \leq -Z_{\alpha/2} \end{matrix}$

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}} \begin{matrix} \geq Z_{\alpha/2} \\ \leq -Z_{\alpha/2} \end{matrix}$

etc…

3.2. Contraste para la diferencia de medias. Varianzas estimadas

¿Y si la muestra es pequeña?

Necesitamos que la población se distribuya normalmente.

Varianzas poblacionales iguales

$$\hat{s}_T = \frac{(n_1-1)\hat{s}_1 + (n_2-1)\hat{s}_2}{n_1+n_2-2}$$

$$T_0 = \frac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightarrow t_{n_1+n_2-2}$$

Carlos Montes – uc3m

3.2. Contraste para la diferencia de medias. Varianzas estimadas

H_0: $\mu_1=\mu_2$
H_1: $\mu_1\neq\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \begin{matrix} > t_{n_1+n_2-2;\alpha/2} \\ < -t_{n_1+n_2-2;\alpha/2} \end{matrix}$

H_0: $\mu_1\leq\mu_2$
H_1: $\mu_1>\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha}$

H_0: $\mu_1\geq\mu_2$
H_1: $\mu_1<\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha}$

3.2. Contraste para la diferencia de medias. Varianzas estimadas

Varianzas distintas

$$T_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^{\,2}}{n_1}+\frac{\hat{s}_2^{\,2}}{n_2}}} \rightarrow t_\nu$$

$$\nu = \frac{\left(\frac{\hat{s}_1^2}{n_1}+\frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{\hat{s}_1^2}{n_2}\right)+\frac{1}{n_2-1}\left(\frac{\hat{s}_2^2}{n_2}\right)}$$

3.2. Contraste para la diferencia de medias. Varianzas estimadas

H_0: $\mu_1=\mu_2$
H_1: $\mu_1\neq\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^{\,2}}{n_1}+\frac{\hat{s}_2^{\,2}}{n_2}}} \begin{matrix} > t_{\nu;\alpha/2} \\ < -t_{\nu;\alpha/2} \end{matrix}$

H_0: $\mu_1\leq\mu_2$
H_1: $\mu_1>\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^{\,2}}{n_1}+\frac{\hat{s}_2^{\,2}}{n_2}}} > t_{\nu;\alpha}$

H_0: $\mu_1\geq\mu_2$
H_1: $\mu_1<\mu_2$ Rechazamos si: $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^{\,2}}{n_1}+\frac{\hat{s}_2^{\,2}}{n_2}}} < -t_{\nu;\alpha}$

4. Contraste para la comparación de medias con muestras emparejadas

Para detectar pequeñas diferencias entre medias es más conveniente tomar muestras por pares en condiciones semejantes.

Pares de datos de los mismos elementos.

- *Antes /después de cierto cambio.*
- *Antes / después de un tratamiento.*
- *Una misma medición con distintos aparatos.*

4. Contraste para la comparación de medias con muestras emparejadas

Con las n observaciones de X_1 y X_2 construimos una nueva variable:

$$Y = X_1 - X_2$$

$$\mu_Y = E(Y) = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$$

$$H_0 = \mu_1 - \mu_2 = \mu_Y = 0$$
$$H_1 = \mu_1 - \mu_2 = \mu_Y \neq 0$$

4. Contraste para la comparación de medias con muestras emparejadas

El estadístico del contraste será como el estudiado para la media:

$$T = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{n}} \to Z \qquad T_0 = \frac{\bar{y} - 0}{\hat{s}_y/\sqrt{n}} \to Z$$

$$\hat{s}_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

4. Contraste para la comparación de medias con muestras emparejadas

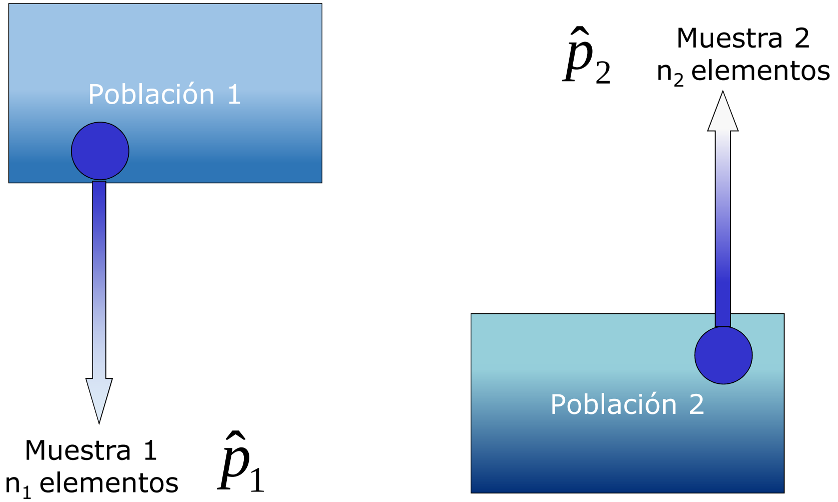
Rechazamos H_0 si:

$$\frac{\bar{y}}{\hat{s}_y/\sqrt{n}} \quad \begin{array}{lcl} \geq t_{(n-1);\alpha/2} & o & Z_{\alpha/2} \\ \leq -t_{(n-1);\alpha/2} & o & -Z_{\alpha/2} \end{array}$$

Muestra pequeña (población normal)

Muestra grande

5. Intervalo de confianza para la diferencia de proporciones



5. Intervalo de confianza para la diferencia de proporciones

$$\hat{p}_1 \to N\left(p_1, \sqrt{\frac{p_1 q_1}{n_1}}\right) \quad \hat{p}_2 \to N\left(p_2, \sqrt{\frac{p_2 q_2}{n_2}}\right)$$

La diferencia de proporciones seguirá una distribución:

$$(\hat{p}_1 - \hat{p}_2) \to N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

5. Intervalo de confianza para la diferencia de proporciones

Luego el intervalo de confianza queda:

$$\boxed{IC(1-\alpha):(p_1 - p_2) \in \left\{(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}\right\}}$$

6. Contraste para la comparación de proporciones

Se desea contrastar la hipótesis
de que la proporción de elementos con cierto atributo
es la misma en dos poblaciones.

$$H_0 : p_1 = p_2 = p_0$$
$$H_1 : p_1 \neq p_0$$

Se toman
muestras independientes
de tamaño n_1 y n_2
de ambas poblaciones

$$\hat{p}_1 \qquad \hat{p}_2$$

6. Contraste para la comparación de proporciones

Si H_0 es cierta, la estimación de mayor precisión para p_0 es:

$$\hat{p}_0 = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

$$\mathrm{var}(\hat{p}_1) = \frac{\hat{p}_0 \hat{q}_0}{n_1} \qquad \mathrm{var}(\hat{p}_2) = \frac{\hat{p}_0 \hat{q}_0}{n_2}$$

6. Contraste para la comparación de proporciones

Si la muestra es suficientemente grande:

$$Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \to Z$$

6. Contraste para la comparación de proporciones

H_0: $p_1 = p_2$
H_1: $p_1 \neq p_2$

Rechazamos si: $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \begin{array}{l} > Z_{\alpha/2} \\ < -Z_{\alpha/2} \end{array}$

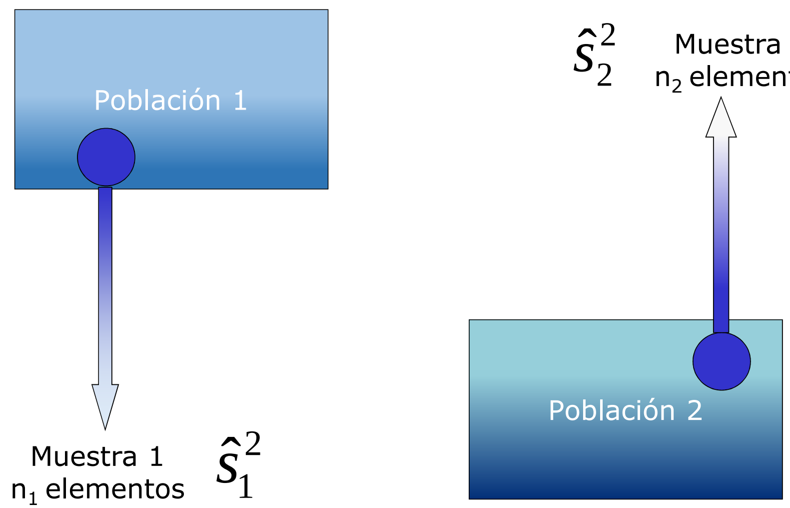
H_0: $p_1 \leq p_2$
H_1: $p_1 > p_2$

Rechazamos si: $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > Z_{\alpha}$

H_0: $p_1 \geq p_2$
H_1: $p_1 < p_2$

Rechazamos si: $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < -Z_{\alpha}$

7. Intervalo de confianza para la razón de varianzas en poblaciones normales



7. Intervalo de confianza para la razón de varianzas en poblaciones normales

Se demuestra que:

$$F = \frac{\hat{s}_1^2 / \sigma_1^2}{\hat{s}_2^2 / \sigma_2^2} \longrightarrow F_{(n_1-1, n_2-1)} \qquad \textit{(F de Fisher)}$$

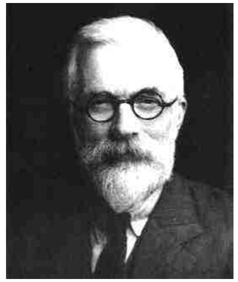
Grados de libertad del numerador

Grados de libertad del denominador

7. Intervalo de confianza para la razón de varianzas en poblaciones normales

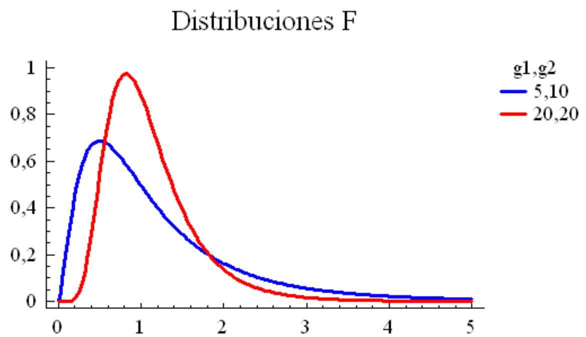
Ronald Aylmer Fisher
(1890-1962)

$$F_{n.m} = \frac{\chi_n^2 / n}{\chi_m^2 / m}$$



Compara la longitud de vectores aleatorios de variables normales independientes.

7. Intervalo de confianza para la razón de varianzas en poblaciones normales



Propiedad: $$\frac{1}{F_{(n_1, n_2, a)}} = F_{(n_2, n_1, 1-a)}$$

7. Intervalo de confianza para la razón de varianzas en poblaciones normales

$$F = \frac{\hat{s}_1^2 / \sigma_1^2}{\hat{s}_2^2 / \sigma_2^2} \longrightarrow F_{(n_1-1, n_2-1)}$$

$$\boxed{IC(1-\alpha): \frac{\sigma_1^2}{\sigma_2^2} \in \left\{ \frac{\hat{s}_1^2}{\hat{s}_2^2} F_{n_1-1, n_2-1; 1-\alpha/2} ; \frac{\hat{s}_1^2}{\hat{s}_2^2} F_{n_1-1, n_2-1; \alpha/2} \right\}}$$

8. Contraste para la igualdad de varianzas en poblaciones normales

$$F = \frac{\hat{s}_1^2 / \sigma_1^2}{\hat{s}_2^2 / \sigma_2^2} \rightarrow F_{(n_1-1, n_2-1)}$$

$H_0 : \sigma_1^2 = \sigma_2^2$ Si H_0 es cierta:

$$F_0 = \frac{\hat{s}_1^2}{\hat{s}_2^2} \rightarrow F_{(n_1-1, n_2-1)}$$

8. Contraste para la igualdad de varianzas en poblaciones normales

H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$ Rechazamos si: $\dfrac{\hat{s}_1^2}{\hat{s}_2^2}$ $> F_{n_1-1, n_2-1; \alpha/2}$ $< F_{n_1-1, n_2-1; 1-\alpha/2}$

H_0: $\sigma_1^2 \leq \sigma_2^2$
H_1: $\sigma_1^2 > \sigma_2^2$ Rechazamos si: $\dfrac{\hat{s}_1^2}{\hat{s}_2^2} > F_{n_1-1, n_2-1; \alpha}$

H_0: $\sigma_1^2 \geq \sigma_2^2$
H_1: $\sigma_1^2 < \sigma_2^2$ Rechazamos si: $\dfrac{\hat{s}_1^2}{\hat{s}_2^2} < F_{n_1-1, n_2-1; 1-\alpha}$