Repaso
Algunas distribuciones:

Normal

Si $y \sim N(\mu, \sigma^2)$ entonces:

$$z = \frac{y-\mu}{\sigma} \sim N(0,1)$$

Muchas técnicas estadísticas suponen que la variable aleatoria en cuestión se distribuye normalmente. El Teorema Central del Límite es muchas veces una justificación para suponer normalidad.

Teorema Central del Límite.

Sean $x_1, x_2, \ldots, x_n$ v.a.i.i.d de una función de probabilidad $f_X(x)$ con media μ y varianza σ^2.

Sea $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n}$, para un tamaño de muestra **grande n**, la distribución de $\bar{x}$ es aproximadamente:

$$\bar{x} \dot{\sim} N(\mu, \sigma^2/n) \quad \text{ó} \quad \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \dot{\sim} N(0,1)$$

Ji-cuadrada

Una distribución muestral que se puede definir a través de v.a. normales es la distribución Ji-cuadrada χ^2

Si $z_1, z_2, \ldots, z_k$ son v.a.i.i.d. $N(0,1)$ entonces:

$$x = z_1^2 + z_2^2 + \ldots + z_k^2$$

tiene una distribución χ_k^2.

La densidad tiene la forma:

$$f(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{k/2-1} e^{-x/2} \quad x > 0$$

Ji-cuadrada

La distribución es asimétrica con $\mu = k$ y $\sigma^2 = 2k$.

Un ejemplo de v.a. con distribución Ji-cuadrada es el siguiente:

Suponga que $y_1, y_2, \ldots, y_n$ es una m.a. de $N(\mu, \sigma^2)$, entonces:

$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

t-Student

Si z tiene distribución normal estándar y x tiene distribución χ^2_k y z y x son independientes, entonces la v.a.

$$t = \frac{z}{\sqrt{x/k}} \sim t_k$$

La función de densidad tiene la forma:

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\,\Gamma(k/2)} \frac{1}{[(t^2/k)+1]^{(k+1)/2}} \qquad -\infty < t < \infty$$

La distribución es simétrica con $\mu = 0$ y $\sigma^2 = k/(k-2)$ para $k > 2$.

t-Student

Un ejemplo de v.a. con distribución t es:

Si $y_1, y_2, \ldots, y_n$ es una m.a. de $N(\mu, \sigma^2)$ entonces:

$$t = \frac{\bar{y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

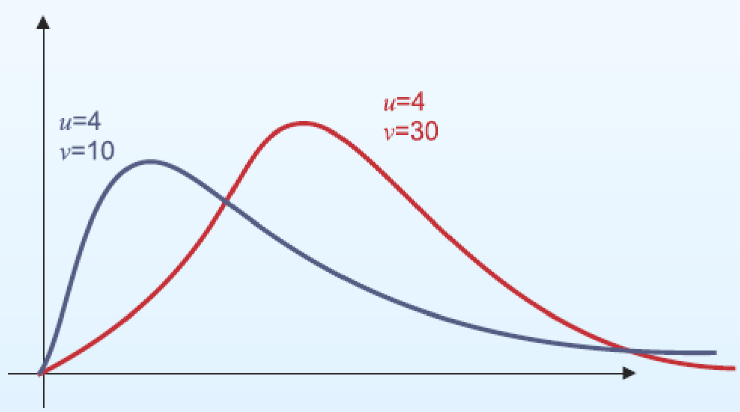
donde,

$$S = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$$

Distribución F

Si χ_u^2 y χ_v^2 son dos v.a. χ^2 independientes con u y v g.l. respectivamente, entonces

$$F = \frac{\chi_u^2/u}{\chi_v^2/v} \sim F_{u,v}$$



Distribución F

Como ejemplo de una estadística que se distribuye como F, suponga que tenemos dos poblaciones normales independientes con varianza común σ^2, es decir,

$y_{11}, y_{12}, \ldots, y_{1n_1}$ es una m.a. de la primera población

$y_{21}, y_{22}, \ldots, y_{2n_2}$ es una m.a. de la segunda población

entonces,

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$$

donde

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(y_{1i} - \bar{y}_1)^2}{n_1 - 1} \text{ y } S_2^2 = \frac{\sum_{i=1}^{n_2}(y_{2i} - \bar{y}_2)^2}{n_2 - 1}$$

Distribución F

Sabemos que:

$$S_1^2 = \frac{SS_1}{n_1 - 1} \text{ y } S_2^2 = \frac{SS_2}{n_2 - 1}$$

y

$$\frac{SS_1}{\sigma^2} = \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$$

$$\frac{SS_2}{\sigma^2} = \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$$

Por lo tanto:

$$\frac{\frac{(n_1-1)S_1^2}{(n_1-1)\sigma^2}}{\frac{(n_2-1)S_2^2}{(n_2-1)\sigma^2}} \sim F_{n_1-1,n_2-1}$$

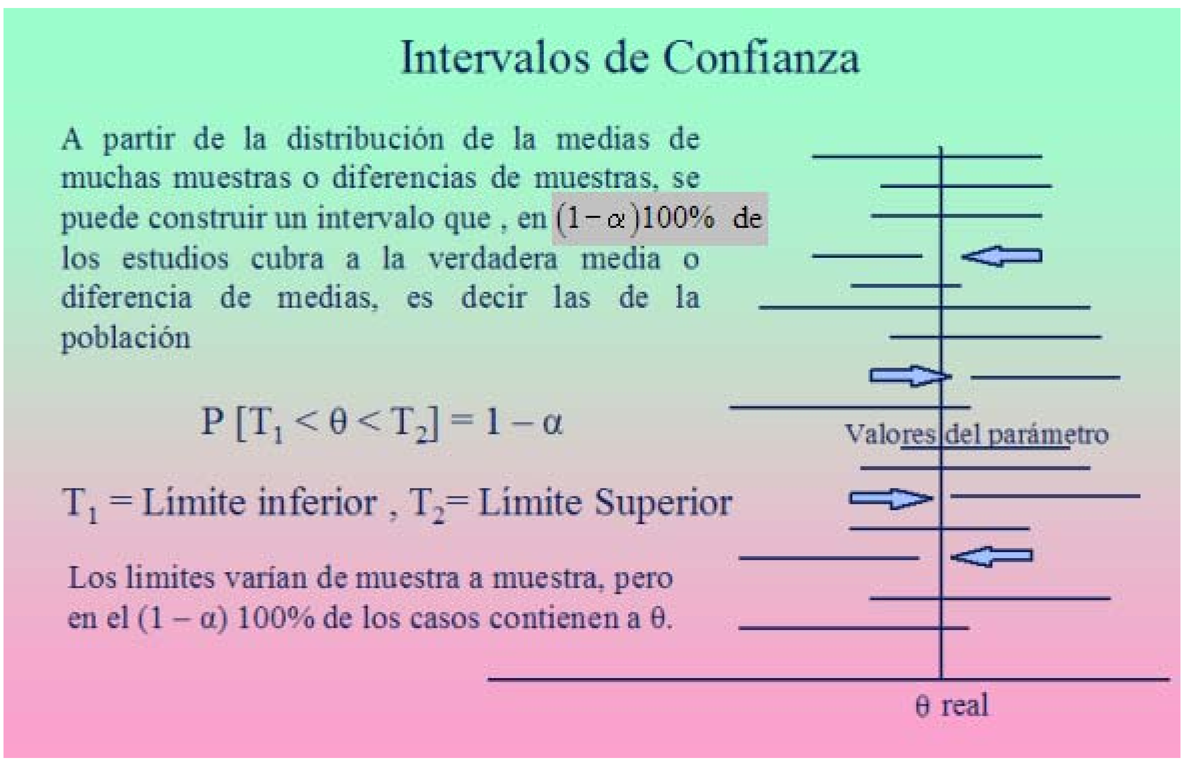
Estimación por intervalos (Intervalos de confianza)

- Conocemos $\hat{\theta}$ por algún método pero es deseable que vaya acompañado de alguna medida del posible error de estimación.

Sea $x_1, x_2, ..., x_n$ una m.a. de la densidad $f(\bullet, \theta)$ y sean $T_1 = t_1(x_1, x_2, ..., x_n)$ y $T_2 = t_2(x_1, x_2, ..., x_n)$ dos estadísticas tales que: $P[T_1 \leq \theta \leq T_2] = 1 - \alpha$, donde α no depende de θ, entonces el intervalo (T_1, T_2) es llamado intervalo de confianza al $(1-\alpha)100\%$ para el parámetro θ.

Interpretación: Si en un número grande de muestreos repetidos se construyen intervalos de confianza para θ, entonces el $(1-\alpha)100\%$ de éstos contendrán el verdadero valor de θ.

- Los intervalos de confianza tienen una interpretación frecuentista esto es, <u>no</u> sabemos si la aseveración es cierta para una muestra específica, pero sabemos que el método usado para calcular el intervalo de confianza produce aseveraciones correctas el $(1-\alpha)100\%$ de las veces.



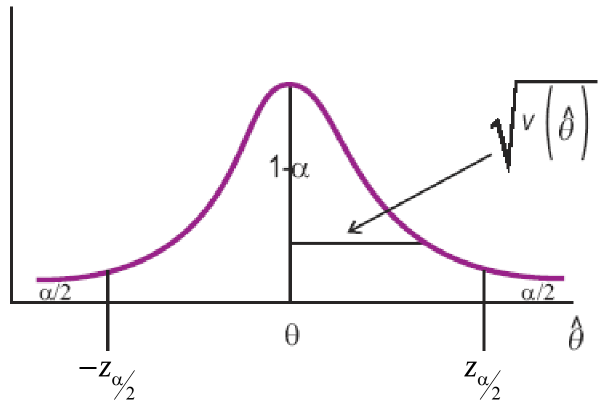
➢ Sea $\hat{\theta}$ un estimador del parámetro θ, si la distribución de $\hat{\theta}$ es desconocida, entonces **Asumimos n grande** y para construir un intervalo de confianza utilizamos teoría asintótica.
Hay dos procedimientos:

- Procedimiento 1.

Sea $\hat{\theta}$ un estimador del parámetro θ, entonces calculamos la varianza de $\hat{\theta}$, $Var\left(\hat{\theta}\right)$, a través del Teorema Central del Límite:

$$\hat{\theta} \overset{a}{\sim} N\left(\theta, Var\left(\hat{\theta}\right)\right)$$

luego estandarizamos: $\dfrac{\hat{\theta}-\theta}{\sqrt{Var\left(\hat{\theta}\right)}} \sim N(0,1) \Rightarrow$



PIVOTAL: función que depende exclusivamente de los datos y del parámetro de interés, además su distribución es conocida y no depende de parámetros desconocidos (por lo que esta totalmente determinada).

$$\Downarrow$$

$$P\left(-z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\sqrt{Var\left(\hat{\theta}\right)}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)} \le \hat{\theta}-\theta \le z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}\right) = 1-\alpha$$

$$\Rightarrow P\left(-\hat{\theta}-z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)} \le -\theta \le -\hat{\theta}+z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}\right) = 1-\alpha$$

$$P\left(\hat{\theta}-z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)} \le \theta \le \hat{\theta}+z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}\right) = 1-\alpha$$

Como $P\left[T_1 \le \theta \le T_2\right] = 1-\alpha$, entonces el intervalo de confianza al $(1-\alpha)100\%$ para el parámetro θ, para n grande es $(T_1, T_2) = \left(\hat{\theta}-z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}\ ,\ \hat{\theta}+z_{\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}\right)$.

De tablas de la $N(0,1)$

$1-\alpha = 0.99 \quad z_{.005} = 2.57$

$1-\alpha = 0.95 \quad z_{.025} = 1.96$

$1-\alpha = 0.90 \quad z_{.05} = 1.64$

- Procedimiento 2.

Sea $\hat{\theta}_{M.V.}$ un estimador máximo verosímil del parámetro θ, entonces **Bajo condiciones de regularidad.** Un estimador máximo verosímil, que depende de n, se distribuye asintóticamente como:

$$\hat{\theta}_{MV(n)} \overset{a}{\sim} N\left(\theta, \frac{1}{nE\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\}}\right) = N\left(\theta, I(\theta)^{-1}\right)$$

donde: $I(\theta)^{-1} = \frac{1}{nE\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\}}$

entonces estandarizando: $\frac{\hat{\theta}_{M.V} - \theta}{\sqrt{I(\theta)^{-1}}} \sim N(0,1)$

Por lo tanto el intervalo de confianza al $(1-\alpha)100\%$ para el parámetro θ, para n grande es:

$$(T_1, T_2) = \left(\hat{\theta}_{M.V} - z_{\alpha/2}\sqrt{I(\theta)^{-1}} \ , \ \hat{\theta}_{M.V} + z_{\alpha/2}\sqrt{I(\theta)^{-1}}\right).$$

<u>Pivotal exacta.</u>

Si $\hat{\theta}$, estimador del parámetro θ, tiene distribución conocida entonces podemos construir la pivotal sin usar teoría asintótica y determinar el intervalo de confianza.

- **Intervalo para μ**

Caso 1. Sea $x_1, x_2, ..., x_n$ una m.a. $N(\mu, \sigma^2)$ con σ^2 conocida.

El estimador para μ es: $\hat{\mu} = \overline{x}$.

$$\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{estandarizando: } \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = z \sim N(0,1) \rightarrow PIVOTAL$$

entonces el intervalo de confianza al $(1-\alpha)100\%$ para μ, es:

$$(T_1, T_2) = \left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \ , \ \overline{x} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right).$$

Caso 2. Sea $x_1, x_2, \ldots, x_n$ una m.a. $N(\mu, \sigma^2)$ con μ **y** σ^2 desconocidos.

El estimador para μ es: $\hat{\mu} = \bar{x}$. $\qquad \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

estandarizando: $\dfrac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = z \sim N(0,1) \rightarrow$ *No es PIVOTAL* ya que desconocemos a σ^2

Sabemos que: $\dfrac{\sum (x_i - \bar{x})^2}{\sigma^2} = X \sim \chi^2_{(n-1)}$

además si $z \sim N(0,1)$ y $X \sim \chi^2_{(k)}$ entonces $t = \dfrac{z}{\sqrt{\frac{X}{k}}} \sim t_{(k)}$

Por lo tanto:

$$t = \frac{z}{\sqrt{\frac{X}{k}}} = \frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\sum (x_i-\bar{x})^2}{\sigma^2 (n-1)}}} \sim t_{(n-1)} \Rightarrow \frac{\bar{x}-\mu}{\sqrt{\frac{\sum (x_i-\bar{x})^2}{n(n-1)}}} \sim t_{(n-1)} \Rightarrow \frac{\bar{x}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t_{(n-1)} \rightarrow \textit{PIVOTAL}$$

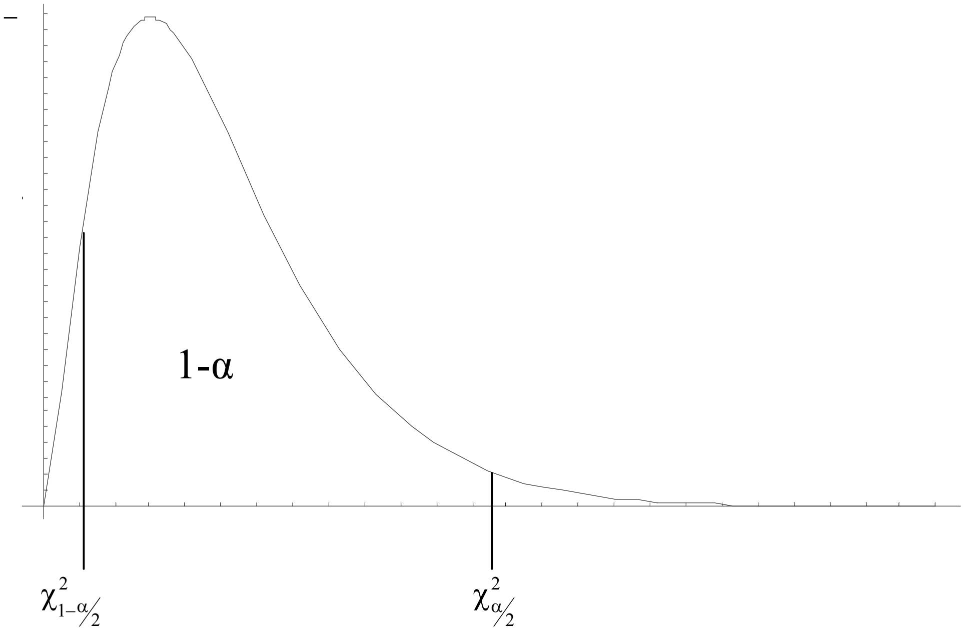
entonces el intervalo de confianza al $(1-\alpha)100\%$ para μ, es:

$$(T_1, T_2) = \left(\bar{x} - t^{\alpha/2}_{(n-1)} \sqrt{\frac{S^2}{n}} \ , \ \bar{x} + t^{\alpha/2}_{(n-1)} \sqrt{\frac{S^2}{n}} \right).$$

- **Intervalo para** σ^2

Sea $x_1, x_2, \ldots, x_n$ una m.a. $N(\mu, \sigma^2)$ con μ **y** σ^2 desconocidos.
Intervalo para σ^2

$$\frac{\sum (x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)} \Rightarrow \frac{(n-1)\sum (x_i-\bar{x})^2}{(n-1)\sigma^2} \sim \chi^2_{(n-1)} \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \rightarrow \textit{PIVOTAL}$$



$$\Rightarrow P\left(\chi^2_{\left(1-\alpha/2,\ n-1\right)} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\left(\alpha/2,\ n-1\right)}\right) = 1-\alpha \Rightarrow P\left(\chi^2_{\left(1-\alpha/2,\ n-1\right)} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\left(\alpha/2,\ n-1\right)}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{1}{\chi^2_{\left(\alpha/2,\ n-1\right)}} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{\chi^2_{\left(1-\alpha/2,\ n-1\right)}}\right) = 1-\alpha \Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{\left(\alpha/2,\ n-1\right)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\left(1-\alpha/2,\ n-1\right)}}\right) = 1-\alpha$$

Como $P\left[T_1 \le \theta \le T_2\right] = 1-\alpha$, entonces el intervalo de confianza al $(1-\alpha)100\%$ para σ^2, es

$$\left(T_1, T_2\right) = \left(\frac{(n-1)S^2}{\chi^2_{\left(\alpha/2,\ n-1\right)}}\ ,\ \frac{(n-1)S^2}{\chi^2_{\left(1-\alpha/2,\ n-1\right)}}\right)$$

- **Intervalo para p (proporción)**

Sea $x_1, x_2, ..., x_n$ una m.a. Bernoulli(p)
Intervalo de confianza para **p**

El estimador para p es: $\hat{p} = \frac{\sum x_i}{n}$ **es máximo verosímil**

Ahora $Var(\hat{p}) = \frac{p(1-p)}{n} = \frac{1}{nE\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\}} = I(\theta)^{-1}$

entonces Bajo condiciones de regularidad.

$$\hat{\theta}_{MV(n)} \overset{a}{\sim} N\left(\theta, I(\theta)^{-1}\right) \Rightarrow \hat{p} \overset{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \Rightarrow estimando: \hat{p} \overset{a}{\sim} N\left(p, \frac{\hat{p}(1-\hat{p})}{n}\right)$$

entonces estandarizando: $\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \overset{a}{\sim} N(0,1) \rightarrow PIVOTAL$

Por lo tanto el intervalo de confianza al $(1-\alpha)100\%$ para el parámetro p:

$$(T_1, T_2) = \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\; , \;\; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Intervalos de confianza para dos poblaciones.

- Intervalo para $\mu_1 - \mu_2$

Caso 1
Sea $x_i \perp y_i$

$$\left.\begin{array}{l} x_1, x_2, ..., x_n \sim N\left(\mu_1, \sigma_1^2\right), \; con \; \sigma_1^2 \; conocida \\ y_1, y_2, ..., y_m \sim N\left(\mu_2, \sigma_2^2\right), \; con \; \sigma_2^2 \; conocida \end{array}\right\} \Rightarrow \hat{\mu}_1 - \hat{\mu}_2 = \bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

estandarizando $\frac{(\bar{x}-\bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1) \rightarrow PIVOTAL$

Por lo tanto el intervalo de confianza al $(1-\alpha)100\%$ para para $\mu_1 - \mu_2$, es:

$$(T_1, T_2) = \left((\bar{x}-\bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \;\; , \;\; (\bar{x}-\bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

Caso 2.

Sea $x_i \perp y_i$

$$\left.\begin{array}{l} x_1, x_2, \ldots, x_n \sim N\left(\mu_1, \sigma_1^2\right),\ con\ \sigma_1^2 \text{ desconocida} \\ y_1, y_2, \ldots, y_m \sim N\left(\mu_2, \sigma_2^2\right),\ con\ \sigma_2^2 \text{ desconocida} \end{array}\right\} \Rightarrow \hat{\mu}_1 - \hat{\mu}_2 = \bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

estandarizando:

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1) \rightarrow \textit{No es PIVOTAL} \text{ ya que desconocemos } \sigma_1^2 \text{ y } \sigma_2^2$$

Sabemos que:

$$\left.\begin{array}{l} \dfrac{\sum (x_i - \bar{x})^2}{\sigma_1^2} = \dfrac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2_{(n-1)} \\ \dfrac{\sum (y_i - \bar{y})^2}{\sigma_2^2} = \dfrac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2_{(m-1)} \end{array}\right\} \Rightarrow X = \frac{(n-1)S_1^2}{\sigma_1^2} + \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2_{(n+m-2)}$$

además si $z = \dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$ y $X \sim \chi^2_{(k)}$ entonces $t = \dfrac{z}{\sqrt{\frac{X}{k}}} \sim t_{(k)}$

Por lo tanto:

$$t = \frac{z}{\sqrt{\frac{X}{k}}} = \frac{\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}}{\sqrt{\dfrac{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}{(n+m-2)}}} \sim t_{(n+m-2)} \rightarrow$$

No es PIVOTAL

ya que no se eliminan las varianzas poblacionales

Esto se conoce como Problema de Behrens-Fisher

Se asume que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, entonces: $t = \dfrac{\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}}}{\sqrt{\dfrac{\frac{(n-1)S_1^2+(m-1)S_2^2}{\sigma^2}}{n+m-2}}} \sim t_{(n+m-2)}$

$$\Rightarrow t=\frac{\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\dfrac{1}{n}+\dfrac{1}{m}\right)}}}{\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{\sigma^2(n+m-2)}}}=\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{(n+m-2)}\left(\dfrac{1}{n}+\dfrac{1}{m}\right)}}\sim t_{(n+m-2)}\rightarrow PIVOTAL$$

Por lo tanto el intervalo de confianza al $(1-\alpha)100\%$ para la diferencia de medias $\mu_1-\mu_2$, es:

$$\left((\bar{x}-\bar{y})-t_{(n+m-2)}^{\alpha/2}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{(n+m-2)}\left(\frac{1}{n}+\frac{1}{m}\right)}\ ,\ (\bar{x}-\bar{y})+t_{(n+m-2)}^{\alpha/2}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{(n+m-2)}\left(\frac{1}{n}+\frac{1}{m}\right)}\right)$$

- Intervalo de confianza para $\dfrac{\sigma_2^2}{\sigma_1^2}$

Sea $x_i \perp y_i$

$x_1,x_2,\ldots,x_n\sim N(\mu_1,\sigma_1^2)$, *con* μ_1,σ_1^2 desconocidas

$y_1,y_2,\ldots,y_m\sim N(\mu_2,\sigma_2^2)$, *con* μ_2,σ_2^2 desconocidas

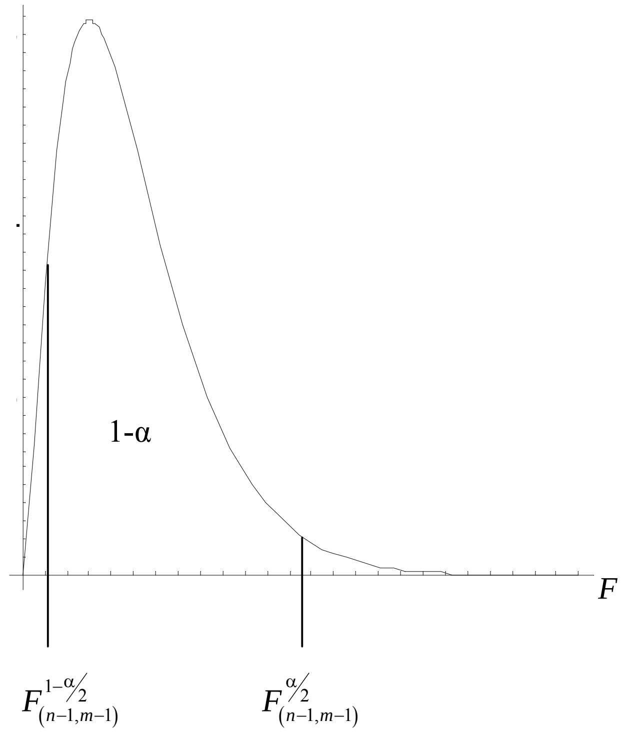
Sabemos que:

$$\frac{\sum(x_i-\bar{x})^2}{\sigma_1^2}=\frac{(n-1)S_1^2}{\sigma_1^2}\sim\chi^2_{(n-1)}$$

$$\frac{\sum(y_i-\bar{y})^2}{\sigma_2^2}=\frac{(m-1)S_2^2}{\sigma_2^2}\sim\chi^2_{(m-1)}$$

Además: $F=\dfrac{\chi_u^2/u}{\chi_v^2/v}\sim F_{(u,v)}$ por lo que:

$$F=\frac{\dfrac{(n-1)S_1^2}{(n-1)\sigma_1^2}}{\dfrac{(m-1)S_2^2}{(m-1)\sigma_2^2}}\sim F_{(n-1,m-1)}\Rightarrow F=\frac{\dfrac{S_1^2}{\sigma_1^2}}{\dfrac{S_2^2}{\sigma_2^2}}=\frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2}\sim F_{(n-1,m-1)}\rightarrow PIVOTAL$$



$$\Rightarrow P\left(F^{1-\alpha/2}_{(n-1,m-1)} \le \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \le F^{\alpha/2}_{(n-1,m-1)}\right) = 1-\alpha \Rightarrow P\left(\frac{S_2^2}{S_1^2} F^{1-\alpha/2}_{(n-1,m-1)} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2} F^{\alpha/2}_{(n-1,m-1)}\right) = 1-\alpha$$

Pero una de las propiedades de la distribución F es que: $F^{1-\alpha/2}_{(n-1,m-1)} = \dfrac{1}{F^{\alpha/2}_{(m-1,n-1)}}$, por lo tanto:

$$\Rightarrow P\left(\frac{S_2^2}{S_1^2} \frac{1}{F^{\alpha/2}_{(m-1,n-1)}} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{S_2^2}{S_1^2} F^{\alpha/2}_{(n-1,m-1)}\right) = 1-\alpha$$

Como $P\left[T_1 \le \theta \le T_2\right] = 1-\alpha$, entonces el intervalo de confianza al $\left(1-\alpha\right)100\%$ para $\dfrac{\sigma_2^2}{\sigma_1^2}$, es:

$$\left(T_1, T_2\right) = \left(\frac{S_2^2}{S_1^2} \frac{1}{F^{\alpha/2}_{(m-1,n-1)}}, \frac{S_2^2}{S_1^2} F^{\alpha/2}_{(n-1,m-1)}\right)$$

de forma similar el intervalo de confianza al $\left(1-\alpha\right)100\%$ para $\dfrac{\sigma_1^2}{\sigma_2^2}$, sería:

$$\left(T_1, T_2\right) = \left(\frac{S_1^2}{S_2^2} \frac{1}{F^{\alpha/2}_{(n-1,m-1)}}, \frac{S_1^2}{S_2^2} F^{\alpha/2}_{(m-1,n-1)}\right)$$

Ejemplo: Una empresa ha estado experimentando con dos disposiciones físicas distintas en su línea de ensamble. Se ha determinado que ambas disposiciones producen aproximadamente el mismo número de unidades terminadas al día. A fin de obtener una disposición que permita un mayor control del proceso, se busca adoptar de manera permanente la línea de ensamble que exhiba la varianza más pequeña en el número de unidades terminadas producidas al día. Dos muestras aleatorias independientes producen los siguientes resultados:

Línea de ensamble 1	Línea de ensamble 2
$n = 21 \; días$	$m = 25 \; días$
$S_1^2 = 1432$	$S_2^2 = 3761$

Establezca un intervalo de confianza de 95% para $\frac{\sigma_1^2}{\sigma_2^2}$, la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble. Con base en el resultado, ¿cuál de las dos líneas de ensamble permite un mayor control del proceso?

Queremos un intervalo de confianza de 95% $\Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$, necesitamos obtener los valores de F, usando tablas tenemos que:

$$F_{(m-1,n-1)}^{\alpha/2} = F_{(24,20)}^{0.025} = 2.41$$
$$F_{(n-1,m-1)}^{\alpha/2} = F_{(20,24)}^{0.025} = 2.33$$

Sustituyendo valores tenemos que el intervalo de confianza al 95% para $\frac{\sigma_1^2}{\sigma_2^2}$, sería:

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{(n-1,m-1)}^{\alpha/2}}, \frac{S_1^2}{S_2^2} F_{(m-1,n-1)}^{\alpha/2} \right) = \left(\frac{S_1^2}{S_2^2} \frac{1}{F_{(20,24)}^{0.025}}, \frac{S_1^2}{S_2^2} F_{(24,20)}^{0.025} \right) = \left(\frac{1432}{3761} \frac{1}{2.33}, \frac{1432}{3761} 2.41 \right) = (0.163, 0.918)$$

Por lo tanto $0.163 \le \frac{\sigma_1^2}{\sigma_2^2} \le 0.918$

Estimamos con 95% de confianza que la razón $\frac{\sigma_1^2}{\sigma_2^2}$ de las verdaderas varianzas de la población quedará dentro del intervalo $(0.163, 0.918)$ y puesto que los valores dentro de este intervalo son menores de 1, podemos confiar en que la varianza en el número de unidades terminadas en la línea $1\left(\sigma_1^2\right)$ es menor que la varianza correspondiente para la línea 2 $\left(\sigma_2^2\right)$. Por lo tanto la línea 1 de ensamble permite un mayor control del proceso.

- Intervalo de confianza para la diferencia de proporciones $p_1 - p_2$

Sea $x_i \perp y_i$

$x_1, x_2, ..., x_n \sim Bernoulli(p_1)$

$y_1, y_2, ..., y_m \sim Bernoulli(p_2)$

$$\left.\begin{aligned} \hat{p}_1 &\overset{a}{\sim} N\left(p_1, \frac{p_1(1-p_1)}{n}\right) \\ \hat{p}_2 &\overset{a}{\sim} N\left(p_2, \frac{p_2(1-p_2)}{m}\right) \end{aligned}\right\} \hat{p}_1 - \hat{p}_2 \overset{a}{\sim} N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}\right)$$

estandarizando:

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} \overset{a}{\sim} N(0,1) \rightarrow PIVOTAL$$

Por lo tanto el intervalo de confianza al $(1-\alpha)100\%$ para para $p_1 - p_2$, es:

$$(T_1, T_2) = \left((\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} \ , \ (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} \right)$$