TEMA 5

ESTIMACIÓN POR INTERVALOS

Estimación por Intervalos

- **INTRODUCCIÓN**
- **CONCEPTOS BÁSICOS**
    - **Nivel de confianza**
    - **Amplitud**
    - **Precisión de un intervalo**
- **MÉTODOS DE CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA**
    - **El método pivotal**
- **INTERVALOS DE CONFIANZA NOTABLES**

INTRODUCCIÓN

- Si se quiere llegar a asignar determinadas garantías o "*confianza*" a los resultados de un proceso inferencial de estimación, cabe la posibilidad de ampliar la óptica de la ***Estimación Puntual*** analizada en el tema anterior pasando a la estimación mediante ***Intervalos de Confianza***.

- En términos estadísticos las "*garantías*" asignables consisten en *afirmaciones de tipo probabilístico.*

- La estimación de una magnitud desconocida mediante un ***Intervalo de Confianza*** consiste en *obtener unos límites aleatorios que contendrán al parámetro desconocido con una probabilidad fijada de antemano.*

CONCEPTOS BÁSICOS

Amplitud, confianza y precisión

- Los extremos de un Intervalo de Confianza son aleatorios, por lo que podrán o no contener al verdadero parámetro y será posible evaluar la probabilidad de que así ocurra. A la probabilidad de que un Intervalo de Confianza contenga al parámetro poblacional objeto de análisis se le denomina ***nivel de confianza*** y la denotaremos por $1-\alpha$.
- Cuanto mayor sea la ***amplitud*** del intervalo, su ***precisión*** será menor. La precisión está relacionada con la *capacidad informativa*.
- Si se requiere una *mayor confianza* se originan *mayores amplitudes* de los intervalos, y por tanto la información que proporcionan acerca del parámetro será *menos precisa*.

MÉTODO PIVOTAL

Sea $X_1, ..., X_n$ m.a.s. de $X \sim f(x|\theta)$ $x \in X$, $\theta \in \Theta$

Definición (*Estadístico Pivote*):

Un estadístico pivote $T(X_1,...,X_n,\theta)$ es un estadístico muestral caracterizado por las siguientes propiedades:

- $T(X_1,\ldots,X_n,\theta)$ es una función monótona respecto de θ.
- La distribución de $T(X_1,\ldots,X_n,\theta)$ es conocida y no depende de θ.
- El estadístico $T(X_1,\ldots,X_n,\theta)$ es calculable en la muestra.

MÉTODO PIVOTAL

Ejemplo: *Hemos visto que en una m.a.s extraída de una población $N(\mu,\sigma)$, el estadístico*

$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

Si σ es conocida, entonces el estadístico tipificado es pivote para μ:

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

MÉTODO PIVOTAL

OBSERVACIÓN: Los requisitos exigidos a un estadístico pivote permiten derivar un intervalo de confianza en los siguientes términos:

- Fijado 1-α, como la distribución del pivote es libre y conocida, existen dos valores $k_{inf}(\alpha)$ y $k_{sup}(\alpha)$ tales que:

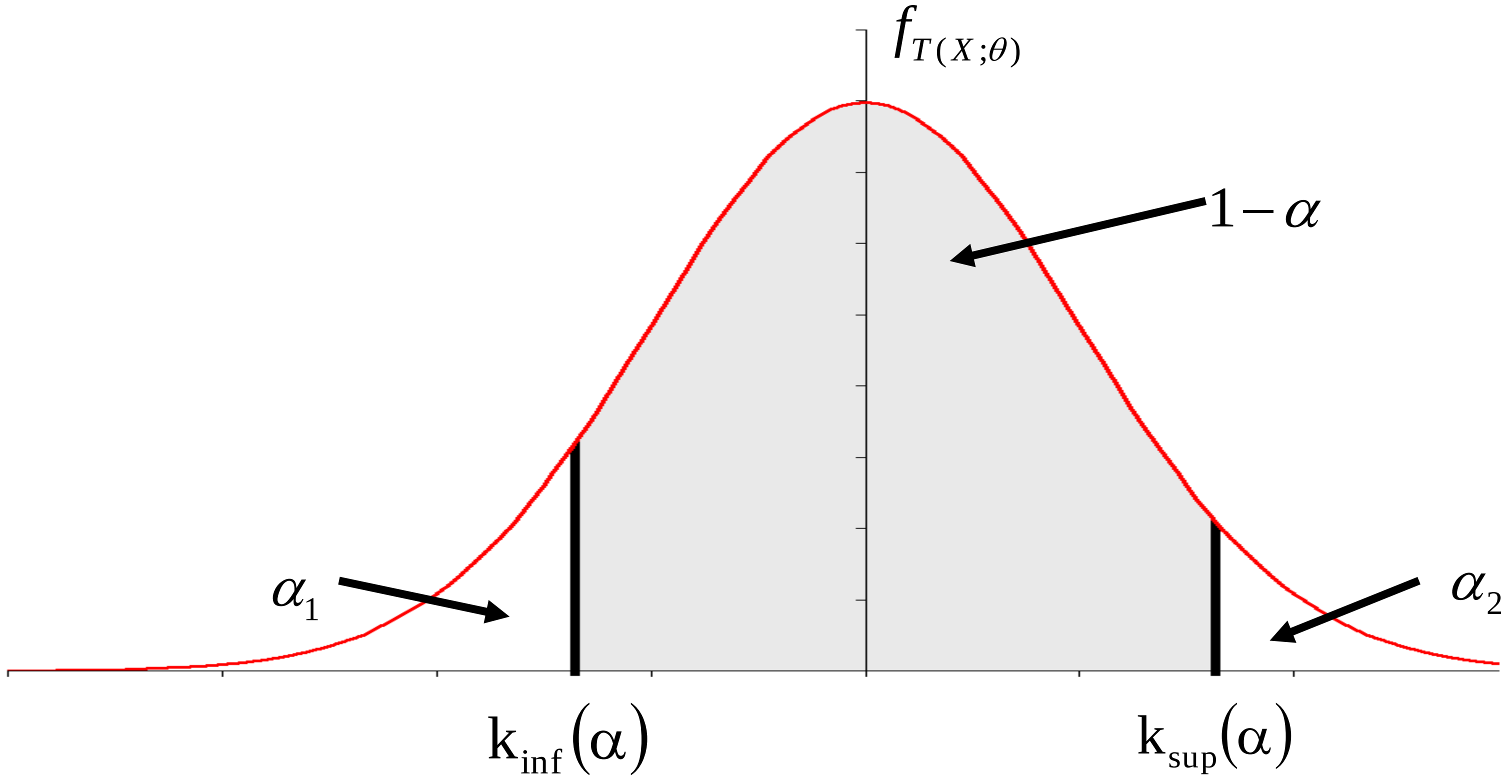
$$\boxed{P\left(k_{inf}(\alpha) \leq T(X_1,\ldots,X_n;\theta) \leq k_{sup}(\alpha)\right) = 1-\alpha}$$

- Debido a que el pivote es una función monótona en θ, mediante manipulaciones, podemos despejar el parámetro de interés:

$$\boxed{P\left(\hat{\theta}_{inf}(X_1,\ldots,X_n;\alpha) \leq \theta \leq \hat{\theta}_{sup}(X_1,\ldots,X_n;\alpha)\right) = 1-\alpha}$$

MÉTODO PIVOTAL

$$P\left(k_{\inf}(\alpha) \le T(X_1,\ldots,X_n;\theta) \le k_{\sup}(\alpha)\right) = 1-\alpha$$



Se cumple que $\alpha_1 + \alpha_2 = \alpha$

MÉTODO PIVOTAL

I.C. de mínima amplitud

Los valores $k_{inf}(\alpha)$ y $k_{sup}(\alpha)$ no son únicos, y el argumento utilizado en su determinación está relacionado con la amplitud del intervalo al que conducen.

Sea $A(k_{inf}(\alpha), k_{sup}(\alpha))$ la amplitud del intervalo de confianza:

$$A\left(k_{inf}(\alpha), k_{sup}(\alpha)\right) = \hat{\theta}_{sup}(X_1, \ldots, X_n; \alpha) - \hat{\theta}_{inf}(X_1, \ldots, X_n; \alpha)$$

Elegiremos $k_{inf}(\alpha)$ y $k_{sup}(\alpha)$ como solución de:

$$\begin{aligned} &\text{Min} \quad A\left(k_{inf}(\alpha), k_{sup}(\alpha)\right) \\ &\alpha_1, \alpha_2 \in (0,1) \\ &\text{s.a.} \quad \alpha_1 + \alpha_2 = \alpha \end{aligned}$$

INTERVALOS NOTABLES

I.C. para la media de una población

A) Población Normal

A.1) Varianza poblacional conocida

A.2) Varianza poblacional desconocida

A.2.i) Tamaño muestral elevado

A.2.ii) Tamaño muestral pequeño

B) Población no Normal

I.C. para la varianza de una normal

C) Población Normal

INTERVALOS NOTABLES

A.1) I.C. para la media μ de una N(μ,σ) con σ conocida

Sea $X_1, \ldots, X_n$ una m.a.s. a partir de $X \sim N(\mu, \sigma)$

Se tiene $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1)$

Fijado 1-α ∃ $z_{inf}(\alpha)$ y $z_{sup}(\alpha)$ tales que

$$P\left(z_{inf}(\alpha) \leq \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \leq z_{sup}(\alpha) \right) = 1 - \alpha$$

$$P\left(z_{inf}(\alpha)\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq z_{sup}(\alpha)\frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$P\left(\overline{X} - z_{sup}(\alpha)\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - z_{inf}(\alpha)\frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

INTERVALOS NOTABLES

A.1) I.C. para la media μ de una N(μ,σ) con σ conocida

$$\text{Minimizar Amplitud}$$

$$\text{Amplitud} = \left(\overline{X} - z_{\inf}(\alpha)\frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X} - z_{\sup}(\alpha)\frac{\sigma}{\sqrt{n}}\right) =$$

$$= z_{\sup}(\alpha)\frac{\sigma}{\sqrt{n}} - z_{\inf}(\alpha)\frac{\sigma}{\sqrt{n}} = \left(z_{\sup}(\alpha) - z_{\inf}(\alpha)\right)\frac{\sigma}{\sqrt{n}}$$

$$\text{Sujeto a}$$

$$P\left(z_{\inf}(\alpha) \leq Z \leq z_{\sup}(\alpha)\right) = F\left(z_{\sup}(\alpha)\right) - F\left(z_{\inf}(\alpha)\right) = 1 - \alpha$$

INTERVALOS NOTABLES

Para ello planteamos la siguiente función:

$$H\left(z_{sup}(\alpha), z_{inf}(\alpha), \lambda\right) = \left(z_{sup}(\alpha) - z_{inf}(\alpha)\right)\frac{\sigma}{\sqrt{n}} - \lambda\left(1 - \alpha - F\left(z_{sup}(\alpha)\right) + F\left(z_{inf}(\alpha)\right)\right)$$

y resolvemos el sistema:

$$\frac{\partial H\left(z_{sup}(\alpha), z_{inf}(\alpha), \lambda\right)}{\partial z_{sup}(\alpha)} = \frac{\sigma}{\sqrt{n}} - \lambda f\left(z_{sup}(\alpha)\right) = 0$$

$$\frac{\partial H\left(z_{sup}(\alpha), z_{inf}(\alpha), \lambda\right)}{\partial z_{inf}(\alpha)} = -\frac{\sigma}{\sqrt{n}} + \lambda f\left(z_{inf}(\alpha)\right) = 0$$

$$\frac{\partial H\left(z_{sup}(\alpha), z_{inf}(\alpha), \lambda\right)}{\partial \lambda} = -1 + \alpha + F\left(z_{sup}(\alpha)\right) - F\left(z_{inf}(\alpha)\right) = 0$$

INTERVALOS NOTABLES

A.1) I.C. para la media μ de una N(μ,σ) con σ conocida

Por tanto,

$$f\left(z_{sup}(\alpha)\right)=f\left(z_{inf}(\alpha)\right) \quad y \quad F\left(z_{sup}(\alpha)\right)-F\left(z_{inf}(\alpha)\right)=1-\alpha$$

$$\Rightarrow \quad z_{sup}(\alpha)=-z_{inf}(\alpha)=z_{\alpha/2}$$

$$\boxed{I.C_{1-\alpha}(\mu)=\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)}$$

NOTA.- Si X no es normal pero el tamaño muestral n es elevado, el resultado es válido gracias al T.C.L que nos aproxima la distribución de la media muestral.

A.2.i) I.C. para la media μ de una $N(\mu,\sigma)$ con σ desconocida y n grande

El hecho de que el tamaño muestral sea elevado, hace que la estimación de la varianza poblacional mediante la cuasi-varianza muestral sea fiable (consistente).

$$S_1^2 \to \sigma^2 \Rightarrow \overline{X} \xrightarrow[\text{consistencia}]{} N\left(\mu, \frac{S_1}{\sqrt{n}}\right) \Rightarrow \frac{\overline{X}-\mu}{S_1}\sqrt{n} \xrightarrow[\text{consistencia}]{} N(0,\ 1)$$

$$\boxed{I.C_{1\text{-}\alpha}(\mu) = \left(\overline{X} - z_{\alpha/2}\frac{S_1}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S_1}{\sqrt{n}}\right)}$$

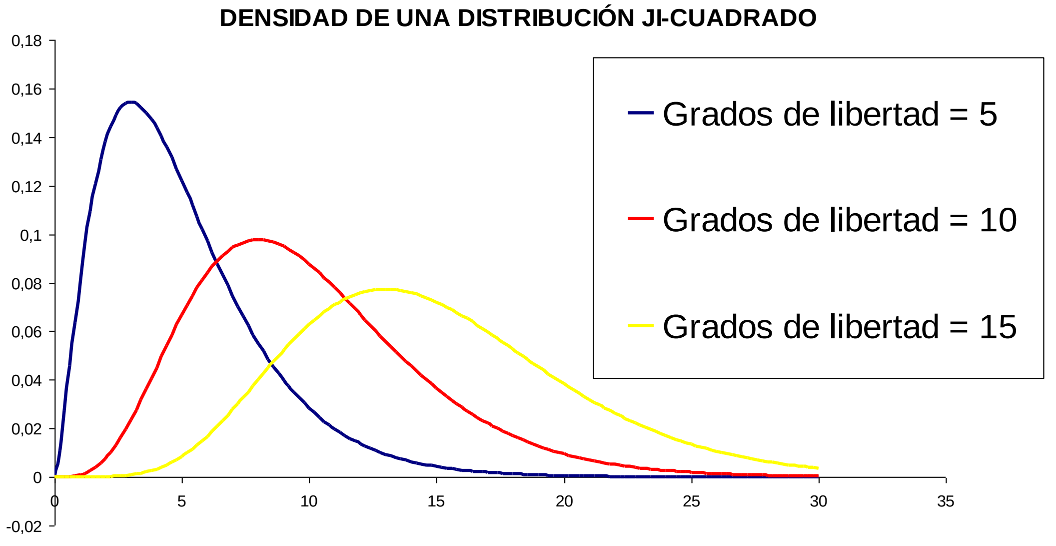
NOTA.- Si X no es normal pero el tamaño muestral n es elevado, el resultado es válido gracias al T.C.L que nos aproxima la distribución de la media muestral.

DISTRIBUCIONES RELACIONADAS

Distribución ji-cuadrado:

Una v.a. se distribuye según una ji-cuadrado con n grados de libertad y se denota χ_n^2 si su función de densidad es:

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \qquad x > 0$$



DISTRIBUCIONES RELACIONADAS

PROPIEDADES:

1) Características

$$E(X) = n$$

$$Var(X) = 2n$$

2) Reproductividad

$$X_1 \sim \chi^2_{n_1} \text{ y } X_2 \sim \chi^2_{n_2} \text{ independientes} \Rightarrow Y = X_1 + X_2 \sim \chi^2_{n_1+n_2}$$

3) Relación con la Normal

$$\boxed{\text{Si } X_1, X_2, \ldots, X_n \sim \text{i.i.d. } N(0,1) \text{ entonces } X_1^2 + X_2^2 + \ldots + X_n^2 \sim \chi^2_n}$$

DISTRIBUCIONES RELACIONADAS

Distribución t-Student

Sean X e Y v.a. independientes

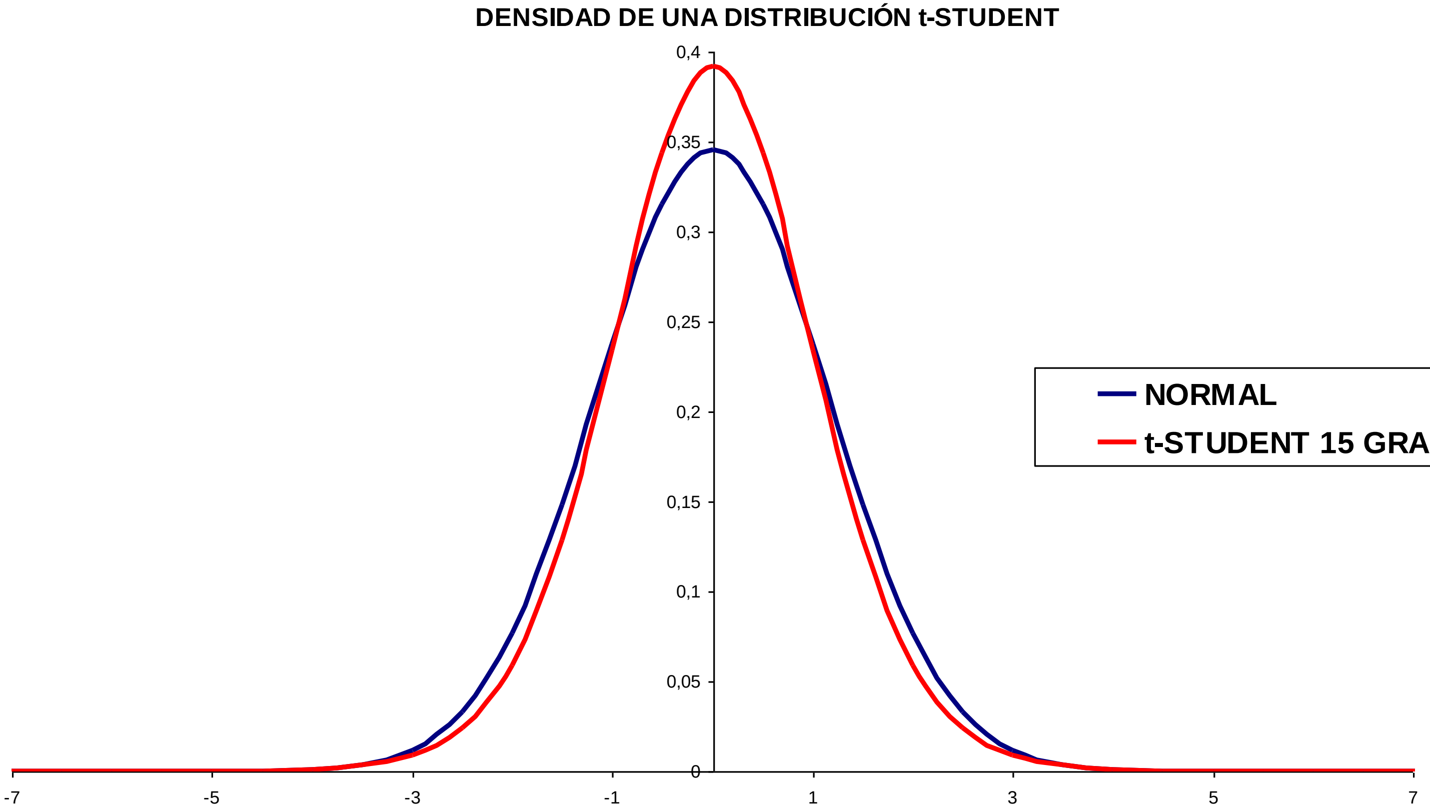
$$\left.\begin{array}{l} X \sim N(0,1) \\ Y \sim \chi^2_n \end{array}\right. \Rightarrow T = \frac{X}{\sqrt{Y/n}} \sim t_n$$

Diremos que la variable T se distribuye según una *t-Student con* n grados de libertad y su función de densidad es:

$$f(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Características: $E(T) = 0 \qquad Var(T) = \frac{n}{n-2}$

DISTRIBUCIONES RELACIONADAS



PROPIEDADES:

a) Asintóticamente tiene una distribución NORMAL

$$\text{Cuando } n \rightarrow \infty \text{ entonces } t_n \rightarrow N(0,1)$$

DISTRIBUCIONES RELACIONADAS

Teorema de Fisher

Sea $X_1, \ldots, X_n$ una m.a.s. a partir de $X \sim N(\mu, \sigma)$.

Se verifica que: i) $\overline{X}$ y S_1^2 son v.a. independientes

$$\text{ii) } \frac{(n\text{-}1)S_1^2}{\sigma^2} \sim \chi^2_{n-1}$$

Observaciones:

1) El estadístico $(n-1)S^2_1 / \sigma^2$ será utilizado como pivote para ontener el I.C para la varianza.

2) Las distribuciones de los estadísticos $\bar{X}$, media muestral, y S_1^2, cuasivarianza muestral, se combinan para calcular el estadístico pivote cuando la varianza es desconocida.

INTERVALOS NOTABLES

A.2.i) I.C. para la media μ de una N(μ,σ) con σ desconocida y n pequeño

Cuando el tamaño de la muestra es moderado o bajo, la estimación de la varianza poblacional mediante la varianza muestral no es buena.

Para resolver el inconveniente se utiliza el Teorema de Fisher.

Construcción del Estadístico Pivote

Sea $X_1,\ldots,X_n$ una m.a.s. a partir de $X \sim N(\mu,\sigma)$

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

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

$$\Rightarrow T = \frac{Z}{\sqrt{\frac{Y}{n-1}}} = \frac{\overline{X}-\mu}{S_1/\sqrt{n}} = \frac{\overline{X}-\mu}{S_1}\sqrt{n} \sim t_{n-1}$$

INTERVALOS NOTABLES

A.2.i) I.C. para la media μ de una N(μ,σ) con σ desconocida y n pequeño

Fijado α sean $t_{n-1;\alpha_1}$ y $t_{n-1;\alpha_2}$ tales que:

$$p\left\{t_{n-1;\alpha_1} \le T_{n-1} \le t_{n-1;\alpha_2}\right\} = 1-\alpha \Leftrightarrow$$

$$p\left\{t_{n-1;\alpha_1} \le \frac{\bar{X}-\mu}{S_1/\sqrt{n}} \le t_{n-1;\alpha_2}\right\} = 1-\alpha$$

Siguiendo el argumento de minimización de la amplitud del intervalo se llega a un intervalo simétrico.

$$\boxed{I.C_{1-\alpha}(\mu) = \left(\bar{X} - t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}, \bar{X} + t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}\right)}$$

INTERVALOS NOTABLES

B) I.C. para el parámetro p en una Be(p)

Un ejemplo típico donde el interés se centra en la media de una población *no normal*, es el parámetro *p=probabilidad de éxito.*

Siempre que el tamaño muestral sea elevado, podemos recurrir al T.C.L para asegurar la normalidad asintótica de la media muestral a la que nos referiremos como **Proporción Muestral.** Además, su valor muestral se utilizará para estimar el poblacional en el cálculo de la varianza

Sea $(X_1,\ldots,X_n)$ una m.a.s. a partir de una Be(p)

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$\boxed{I.C._{1-\alpha}(p) = \left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)}$$

INTERVALOS NOTABLES

C) I.C. para la varianza de una normal

La obtención del I.C se sustenta en el teorema de Fisher que nos da la distribución del pivote basado en la cuasivarianza muestral.

Sea $X_1, \ldots, X_n$ una m.a.s. a partir de $X \sim N(\mu, \sigma)$

Se verifica que $\dfrac{(n\text{-}1)S_1^2}{\sigma^2} \sim \chi^2_{n-1}$

Fijado $1\text{-}\alpha$ tenemos que

$$P\left\{ \chi^2_{n-1;1-\alpha/2} \le \frac{(n\text{-}1)S_1^2}{\sigma^2} \le \chi^2_{n-1;\alpha/2} \right\} = 1-\alpha$$

$$\boxed{I.C_{1-\alpha}\left(\sigma^2\right) = \left(\frac{(\text{n-1})\text{S}_1^2}{\chi^2_{n-1;\alpha/2}}, \frac{(\text{n-1})\text{S}_1^2}{\chi^2_{n-1;1-\alpha/2}} \right)}$$