UNIVERSIDAD
AUSTRAL

STVDIORVM - AVSTRALIS - VNIVERSITAS -

INGENIERÍA

#SOMOSAUSTRAL

#SOMOSAUSTRAL

Estadística

Clase 5

UNIVERSIDAD AUSTRAL

INGENIERÍA

Agenda

- Estimación por Intervalos
- Distintos tipos de Intervalos
- Tamaño de muestra en IC

Intervalos de Confianza

UNIVERSIDAD AUSTRAL

INGENIERÍA

Intervalos de Confianza Aditivos

Se denomina así aquel intervalo que permite que la probabilidad de que la estimación difiera del parámetro en a lo sumo h veces el desvío estándar del estimador sea igual al nivel de confianza. En términos formales,

$$P\left(\left|\hat{\theta}-\theta\right| \leq h\sqrt{Var(\hat{\theta})}\right) = 1-\alpha$$

O expresado como intervalo,

$$P\left(\hat{\theta}-h\sqrt{Var(\hat{\theta})} \leq \theta \leq \hat{\theta}+h\sqrt{Var(\hat{\theta})}\right) = 1-\alpha$$

El factor h se denomina Factor de Confianza y es el valor del percentil de orden $1-\frac{\alpha}{2}$ de la distribución de probabilidad del estadígrafo de transformación del estimador $\hat{\theta}$.

Error de Muestreo

El error de muestreo es la máxima diferencia que podría haber entre el estimador y el parámetro

$$e = \hat{\theta} - \theta$$

En los intervalos aditivos esta cantidad que se suma y se resta al estimador representa el error de muestreo,

$$e = \hat{\theta} - \theta = h\sqrt{Var(\hat{\theta})}$$

Si se conocen los límites de un intervalo aditivo, la estimación puntual se puede calcular mediante la semisuma de los límites del intervalo y el error de muestreo se puede calcular mediante la semidiferencia de los límites del intervalo.

$$\hat{\theta} = \frac{L_s(\hat{\theta}) + L_i(\hat{\theta})}{2}$$

$$e = \frac{L_s(\hat{\theta}) - L_i(\hat{\theta})}{2}$$

Intervalo de Confianza para μ de Poblaciones Normales

Existen distintas variantes según si la población es infinita o finita, según si se conoce la varianza poblacional o no. Las variantes son 4 (cuatro),

1. σ^2 **conocida y Población Infinita**

2. σ^2 **conocida y Población Finita**

3. σ^2 **desconocida y Población Infinita**

4. σ^2 **desconocida y Población Finita**

IC para μ de Poblaciones Normales e Infinita y σ^2 conocida

Sea X una Población que se distribuye normalmente con Esperanza Matemática y Varianza Poblacional igual a μ y σ^2 respectivamente. Bajo estas condiciones se demuestra que,

$$\boldsymbol{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N}$$

Por otro lado considerando que la probabilidad de encontrar un valor de la variable estandarizada entre dos percentiles equidistantes del origen sea $1 - \alpha$

$$\boldsymbol{P\left(Z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha}$$

IC para μ de Poblaciones Normales e Infinita y σ^2 conocida

$$P\left(Z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(-\bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{X}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

IC para μ de Poblaciones Normales e Infinita y σ^2 conocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Ahora bien,

$$Z_{1-\frac{\alpha}{2}} > 0 \qquad Z_{\frac{\alpha}{2}} < 0$$

Se cumple por la simetría de la Distribución Normal que,

$$\left|Z_{1-\frac{\alpha}{2}}\right| = \left|Z_{\frac{\alpha}{2}}\right|$$

Se suele escribir al IC de la forma siguiente,

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

IC para μ de Poblaciones Normales e Finita y σ^2 conocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$

IC para μ de Poblaciones Normales e Infinita y σ^2 desconocida

$$P\left(\bar{X} - t_{n-1;1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1;1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right) = 1-\alpha$$

IC para μ de Poblaciones Normales e Finita y σ^2 desconocida

$$P\left(\bar{X} - t_{n-1;1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + t_{n-1;1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$

$$\hat{\sigma} = S_X = \sqrt{\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}}$$

UNIVERSIDAD AUSTRAL
INGENIERÍA

Intervalo de Confianza para μ de Poblaciones Desconocidas – No Normales -

Para la construcción hay que dividir en muestras grandes y pequeñas. Para las primeras se aplica el TCL (CCL) y para la segunda se aplica el Teorema de Tchebycheff. Luego también existen distintas variantes según si la población es infinita o finita, según si se conoce la varianza poblacional o no. Las variantes son 4 (cuatro),

Muestras grandes (Teorema Central del Límite) y Muestras Pequeñas (Teorema de Tchebycheff)

1. σ^2 **conocida y Población Infinita**

2. σ^2 **conocida y Población Finita**

3. σ^2 **desconocida y Población Infinita**

4. σ^2 **desconocida y Población Finita**

A continuación se exhiben los IC para Muestras Grandes (suponemos mayores a 30 para aplicar TCL)

IC para μ de Poblaciones NO Normales e Infinita y σ^2 conocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Finita y σ^2 conocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Infinita y σ^2 desconocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right) = 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Finita y σ^2 desconocida

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Infinita y σ^2 conocida

$$P\left(\bar{X} - k_0\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + k_0\frac{\sigma}{\sqrt{n}}\right) \geq 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Finita y σ^2 conocida

$$P\left(\bar{X} - k_0\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + k_0\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) \geq 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Infinita y σ^2 desconocida

$$\left(\bar{X} - k_0\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + k_0\frac{\hat{\sigma}}{\sqrt{n}}\right) \geq 1 - \alpha$$

IC para μ de Poblaciones NO Normales e Finita y σ^2 desconocida

$$P\left(\bar{X} - k_0\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + k_0\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) \geq 1 - \alpha$$

$$k_0 = \sqrt{\frac{1}{\alpha}}$$

UNIVERSIDAD AUSTRAL

INGENIERÍA

Intervalo de Confianza para p

Estimador: Proporción Muestral $\hat{p}$

- ***Población Normal y Finita***

$$\mathbf{Z} = \frac{\overline{p} - p}{\sqrt{\frac{\mathbf{p(1-p)}}{n}}} \longrightarrow \mathbf{N(0,1)}$$

Al construir el IC se llega a la siguiente expresión,

$$P\left(\overline{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \overline{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

¿Qué inconsistencia detectan en la fórmula anterior?

Intervalo de Confianza para p

- ***Población Normal y Finita***

$$P\left(\overline{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \overline{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

¿Qué inconsistencia detectan en la fórmula anterior? Que los límites dependen de la proporción poblacional que es desconocido. Por lo tanto, se aproxima mediante la utilización del estimador de dicho parámetro. En términos formales,

$$P\left(\overline{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leq p \leq \overline{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right) = 1-\alpha$$

- ***Población Normal e Infinita***

$$P\left(\overline{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}\left(\frac{N-n}{N-1}\right)} \leq p \leq \overline{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}\left(\frac{N-n}{N-1}\right)}\right) = 1-\alpha$$

Intervalo de Confianza para σ^2

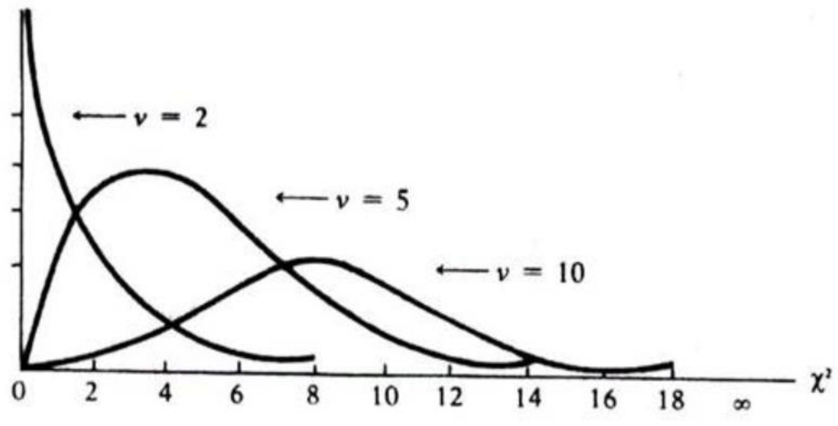
Estimador: Varianza Muestral S_X^2

$$S_X^2 = \hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

Se determinó que la siguiente expresión se distribuye como una variable Chi-Cuadrado con $n-1$ grados de libertad,

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

El siguiente gráfico refleja la forma funcional de dicha distribución en función de los grados de libertad que asuma.



Intervalo de Confianza para σ^2

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

Se intenta buscar dos valores de esta distribución para determinar con un cierto grado de confianza los posibles valores que puede asumir el parámetro desconocido,

$$P\left(A \leq \chi^2_{n-1} \leq B\right) = 1-\alpha$$

$$P\left(A \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq B\right) = 1-\alpha$$

$$P\left(\frac{1}{A} \geq \frac{\sigma^2}{(n-1)\hat{\sigma}^2} \geq \frac{1}{B}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)\hat{\sigma}^2}{B} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{A}\right) = 1-\alpha$$

Intervalo de Confianza para σ^2

$$P\left(\frac{(n-1)\hat{\sigma}^2}{B} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{A}\right) = 1-\alpha$$

Donde A y B son dos percentiles de la Distribución Chi-Cuadrado con $n-1$ grados de libertad y se determinan de la forma siguiente,

$$B = \chi^2_{n-1;1-\frac{\alpha}{2}}$$

$$A = \chi^2_{n-1;\frac{\alpha}{2}}$$

Tamaño de Muestra en IC

UNIVERSIDAD AUSTRAL

INGENIERÍA

El tamaño de la muestra para estimar los parámetros poblacionales es fundamental.

Para estimar la media poblacional μ hay que diferenciar si se conoce o no la varianza poblacional,

- **σ^2 conocida**

El cálculo se realiza teniendo en cuenta los siguientes factores,

1. El error de muestreo
2. La confianza de la estimación
3. La varianza de la población
4. El tamaño de la población

- **σ^2 desconocida**

El cálculo se realiza mediante un proceso iterativo.

Tamaño de la muestra para estimar μ con σ^2 conocida

A partir del error de muestreo y mediante un despeje se obtiene que el tamaño de la muestra se determina de la forma siguiente,

$$e = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} * \sigma}{e} \right)^2$$

Si la población es finita, el error de muestreo es,

$$e = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} \ \sigma^2 \ N}{e^2(N-1) + Z^2_{1-\frac{\alpha}{2}} \ \sigma^2}$$

Tamaño de la muestra para estimar μ con σ^2 desconocida

A partir del error de muestreo y mediante un despeje se obtiene que el tamaño de la muestra se determina de la forma siguiente,

$$\boldsymbol{e = t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}}$$

$$n = \left(\frac{\boldsymbol{t_{n-1;1-\frac{\alpha}{2}} * S}}{e}\right)^2$$

Tal como se observa el tamaño de la muestra depende del factor de confianza, que a su vez depende de n.

Tamaño de la muestra para estimar μ con σ^2 desconocida

Por lo tanto se procederá de la siguiente forma,

1. Se toma una muestra piloto de tamaño arbitrario n_0 (generalmente es igual a 10) y se calcula el valor de la varianza muestral mediante la fórmula insesgada vista en clase.
2. Se calcula un tamaño de muestra inicial mediante la siguiente fórmula,

$$\boldsymbol{n_1} = \left(\frac{\boldsymbol{t_{n_0-1;1-\frac{\alpha}{2}}} * \boldsymbol{S}}{e} \right)^2$$

3. Se obtiene un segundo tamaño de muestra $\boldsymbol{n_2}$, mediante la fórmula anterior empleando el tamaño $\boldsymbol{n_1}$, con una $\boldsymbol{t_{n_1-1;1-\frac{\alpha}{2}}}$.
4. Se obtiene un segundo tamaño de muestra $\boldsymbol{n_3}$, mediante la fórmula anterior empleando el tamaño $\boldsymbol{n_2}$, con una $\boldsymbol{t_{n_2-1;1-\frac{\alpha}{2}}}$.
5. Este proceso iterativo se repite tantas veces hasta que dos tamaños de muestra consecutivos sean iguales. En este caso, se asume que ese es el tamaño de muestra adecuado. $\boldsymbol{n_{i+1}} = \boldsymbol{n_i}$.

Tamaño de la muestra para estimar p

El cálculo se realiza teniendo en cuenta los siguientes factores,

1. El error de muestreo
2. La confianza de la estimación
3. La varianza de la población
4. El tamaño de la población

Hay que distinguir entre Población Finita e Infinita.

UNIVERSIDAD AUSTRAL
INGENIERÍA

Tamaño de la muestra para estimar p - Población Infinita

$$e = Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Dado que el valor de $\overline{p}$ es desconocido porque se determina con la muestra, no puede formar parte de la fórmula utilizada para calcular el tamaño de la muestra. Por lo tanto se utiliza un valor alternativo que se simbolizará $\widehat{p}$.

$$n = \frac{Z^2_{1-\frac{\alpha}{2}}\widehat{p}\widehat{q}}{e^2}$$

El valor de $\widehat{p}$ se podrá obtener de alguna de las formas siguientes,

1. Mediante datos que surjan de trabajos anteriores.
2. Mediante una muestra piloto (generalmente de tamaño 50)
3. Si es imposible alguna de las formas citadas, se utiliza $\widehat{p} = 0,50$, dado que cuando hay dicotomía, la mayor dispersión se alcanza si cada grupo tiene el 50%.

Tamaño de la muestra para estimar p - Población Infinita

$$e = Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{p}(1-\overline{p})}{n} \left(\frac{N-n}{N-1}\right)}$$

Dado que el valor de $\overline{p}$ es desconocido porque se determina con la muestra, no puede formar parte de la fórmula utilizada para calcular el tamaño de la muestra. Por lo tanto se utiliza un valor alternativo que se simbolizará $\widehat{\overline{p}}$.

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} \widehat{\overline{p}}\widehat{\overline{q}}\, N}{e^2(N-1) + Z^2_{1-\frac{\alpha}{2}} \widehat{\overline{p}}\widehat{\overline{q}}}$$

El valor de $\widehat{\overline{p}}$ se podrá obtener de alguna de las formas siguientes,

1. Mediante datos que surjan de trabajos anteriores.
2. Mediante una muestra piloto (generalmente de tamaño 50)
3. Si es imposible alguna de las formas citadas, se utiliza $\widehat{\overline{p}} = 0,50$, dado que cuando hay dicotomía, la mayor dispersión se alcanza si cada grupo tiene el 50%.

Muchas gracias.

www.austral.edu.ar

UNIVERSIDAD AUSTRAL

INGENIERÍA