Análisis e Interpretación de Datos

MÁSTER UNIVERSITARIO EN ANÁLISIS Y VISUALIZACIÓN DE DATOS MASIVOS / VISUAL ANALYTICS AND BIG DATA

Miller Janny Ariza Garzón

Tema 7. Intervalos de Confianza

unir LA UNIVERSIDAD EN INTERNET

Universidad Internacional de La Rioja

Tabla de contenido

❑ Tema 7: Intervalos de Confianza

- Introducción.
- Para la media de una población normal: varianza conocida y desconocida.
- Calculando el tamaño de la muestra.
- Intervalo de confianza para la proporción.
- Intervalo de confianza para la varianza de una población normal.
- Intervalo de confianza para la diferencia de medias y proporciones.
- Intervalos de confianza robustos.

Contenido

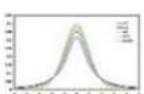
Intervalos de confianza

Introducción

- Nivel de confianza
  Ej.: 90%, 95%, 99%
- Nivel de significación α
- Valor crítico
  Zα/2
- Margen de error
  $E = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

IC para la media

- Varianza conocida
  $\mu \in \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]_{1-\alpha}$
- Tamaño de la muestra (media)
  $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$
- Varianza desconocida
  $\mu \in \left[\bar{x} - t_{\alpha,n-1\alpha/2}\frac{s}{\sqrt{n}}; \bar{x} + t_{\alpha,n-1\alpha/2}\frac{s}{\sqrt{n}}\right]_{1-\alpha}$
  → T-Student

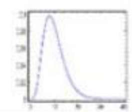


IC para la proporción

- $p \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]_{1-\alpha}$
- Tamaño de la muestra (p)
  $n = \frac{z_{\alpha/2}{}^2\hat{p}(1-\hat{p})}{E^2}$

IC para la varianza

- Chi Cuadrada



- $\sigma^2 \in \left[\frac{(n-1)s^2}{x_{1-\alpha/2}}; \frac{(n-1)s^2}{x_{1+\alpha/2}}\right]_{1-\alpha}$

IC para la diferencia de poblaciones

- Caso medias
  $(\mu_1 - \mu_2) \in \left[(\bar{x} - \bar{y}) \mp z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right]_{1-\alpha}$

Introducción

Parámetro

Poblacional

$X \sim f(\theta, \ldots)$

$\theta = ?$

Estimador

Muestra aleatoria

$X_1, X_2, \ldots, X_n$

$\hat{\theta} = h(X_1, X_2, \ldots, X_n)$

Estimador o Estadístico

Estimación Puntual de θ

Distribución Muestral

$\hat{\theta} \sim f^*$

Normal, t-student, Chi, F,…

Intervalo de confianza

$\left(LI\left(\hat{\theta}, f_q^*, \ldots\right), LS\left(\hat{\theta}, f_q^*, \ldots\right)\right)$

Estimación por intervalo de θ

Media:

$X \sim f(\mu, \ldots) = ?$

$\mu = ?$

Se conoce σ^2

$$\hat{\theta} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$n \geq 30$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

TLC

$$LI = \bar{X} - z_{1-\alpha/2} * \sigma/\sqrt{n}$$

$$LS = \bar{X} + z_{1-\alpha/2} * \sigma/\sqrt{n}$$

Ejemplo:

X: Estaturas de los españoles en cm

$\mu = ?$

Supongamos que el IC al 95% para la media de estatura de los españoles es el siguiente:

$$LI = 167cm$$

$$LS = 192cm$$

Con una herramienta para calcular intervalos que funciona en el 95% de las veces, la estatura media se encuentra entre 167cm y 192cm. Tenemos un nivel de confianza del 95% de que el intervalo (167;192) contenga a la media poblacional de la estatura de los españoles.

Intervalos de confianza para la media μ

Parámetro

Estimador

Distribución muestral

Intervalo de confianza

$$X \sim f(\mu, \dots) = ?$$

$$\mu = ?$$

Se conoce σ^2

$$\hat{\theta} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$n \geq 30$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

TLC

$$LI = \bar{X} - z_{1-\alpha/2} * \sigma/\sqrt{n}$$

$$LS = \bar{X} + \underbrace{z_{1-\alpha/2} * \sigma/\sqrt{n}}_{\text{Margen de Error}}$$

$$X \sim f(\mu, \dots) = N$$

$$\mu = ?$$

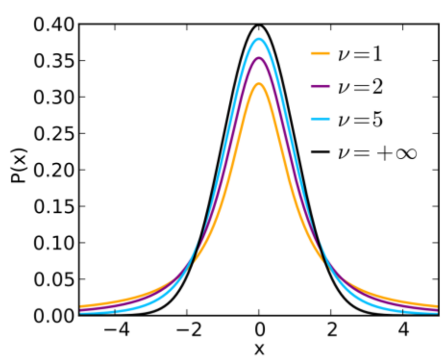
Se desconoce σ^2

$$\hat{\theta} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{\underbrace{n-1}_{\text{GL}}}$$

$$LI = \bar{X} - t_{1-\alpha/2} * s/\sqrt{n}$$

$$LS = \bar{X} + \underbrace{t_{1-\alpha/2} * s/\sqrt{n}}_{\text{Margen de Error}}$$



S es la raíz de

$$s_c^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Intervalos de confianza para la media μ

Ejemplo: Queremos calcular la media para la estatura de las mujeres policía con un IC al 95%. (Enunciado tomado del tema.pdf):

> Para determinar la estatura media de las policías del Cuerpo Nacional de cara establecer un intervalo para el examen de entrada al Cuerpo, se tomó una muestra aleatoria de 10 mujeres resultando:
> 152, 166, 159, 155, 161, 159, 162, 158, 157, y 165cm de estatura.

- No conocemos la varianza
- Se supone que la estatura será un v.a. normal. → Se usa la distribución $t - student$
- Hallamos la media y la cuasivarianza muestrales.

$$\bar{x} = \frac{166+159+\cdots+165}{10} = 159{,}40 \text{ cm}$$

$$s = \sqrt{\frac{\sum_{i=l}^{n}(x_i - \bar{x})^2}{n-1}} \rightarrow s = 4{,}30$$

$GL = n - 1$=10-1=9

Intervalos de confianza para la media μ

Ejemplo: Queremos calcular la media para la estatura de las mujeres policía con un IC al 95%. (Enunciado tomado del tema.pdf):

$$\bar{x} = \frac{166+159+\cdots+165}{10} = 159{,}40 \text{ cm} \qquad s = \sqrt{\frac{\sum_{i=l}^{n}(x_i - \bar{x})^2}{n-1}} \rightarrow s = 4{,}30$$

$$GL = n - 1 = 9$$

$$\left[159{,}40 - 2{,}26\,\frac{4{,}30}{\sqrt{10}};\ 159{,}40 + \underbrace{2{,}26}_{t_{0.975,9}}\,\frac{4{,}30}{\sqrt{10}}\right]_{0{,}95}$$

Tenemos una confianza del 95% de que la media de estatura de las mujeres policía se encuentra entre 156,32 y 162,47cm.

Intervalos de confianza para la media P

Parámetro

$X \sim f(P, \ldots)$

$\boldsymbol{P=?}$

Estimador

$$\hat{\theta} = \hat{P} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$n \geq 30$

$$X_i = \begin{cases} 1 \; si \; se \; da \; un \; éxito \\ 0 \; en \; otro \; caso \end{cases}$$

Distribución muestral

$$\hat{P} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

TLC

Intervalo de confianza

$$LI = \hat{P} - z_{1-\alpha/2} * \sqrt{\hat{P}(1-\hat{P})/_n}$$

$$LS = \hat{P} + \underbrace{z_{1-\alpha/2} * \sqrt{\hat{P}(1-\hat{P})/_n}}_{\text{Margen de Error}}$$

Ejemplo:

Se ha interrogado en un trabajo de estadístico escolar a 100 jóvenes sobre si fuman o no. 30 afirmaron fumar mientras que 70 se declararon no fumadores. ¿Qué porcentaje de fumadores habrá en este instituto con un nivel de confianza del 95%?

$\boldsymbol{P=?}$

$$\hat{P} = \frac{30}{100} = 0.3$$

$$LI = 0.3 - 1.96\sqrt{\frac{0.3 * 0.7}{100}} = 0.21$$

$$LS = 0.3 + \underset{\uparrow \atop z_{0.975}}{1.96}\sqrt{\frac{0.3 * 0.7}{100}} = 0.39$$

Con una confianza del 95% el porcentaje de jóvenes que fuman se encuentra entre 21% y 39%.

Intervalos de confianza para la varianza σ^2

Parámetro

$X \sim f(\sigma^2, \ldots) = N$

$\boldsymbol{\sigma^2 = ?}$

Estimador

$$\hat{\theta} = s^2 = \frac{\sum (x - \bar{x})^2}{\boldsymbol{n-1}}$$

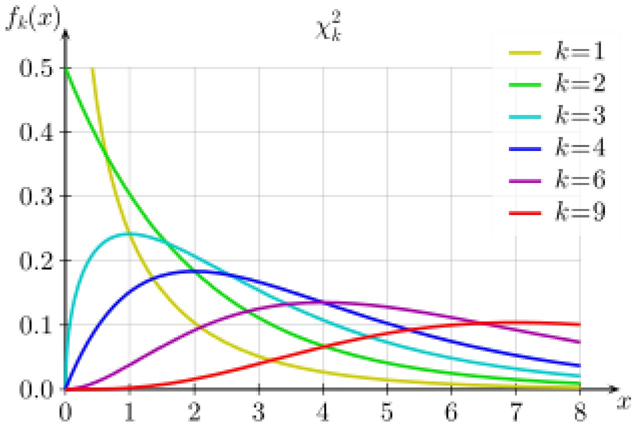
Distribución muestral

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{\underbrace{\boldsymbol{n-1}}_{\text{GL}}}$$

Intervalo de confianza

$$LI = \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}$$

$$LS = \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}$$



Tamaño de muestra n

Parámetro

$$X \sim f(\mu, \ldots) = N$$

$$\boldsymbol{\mu = ?}$$

$$LI = \bar{X} - z_{1-\alpha/2} * \sigma/\sqrt{n}$$

$$LS = \bar{X} + \underbrace{z_{1-\alpha/2} * \sigma/\sqrt{n}}$$

E: Margen de Error

$$n = \frac{Z_{1-\alpha/2}^2 * \sigma^2}{E^2}$$

POBLACIÓN FINITA:

$$n = \frac{\dfrac{Z_{1-\alpha/2}^2 * \sigma^2}{E^2}}{1 + \dfrac{\left(\dfrac{Z_{1-\alpha/2}^2 * \sigma^2}{E^2}\right)}{N}}$$

Factores:

- Confianza
- Varianza. Usar estudios previos o pilotos.

Parámetro

$$X \sim f(\sigma^2, \ldots) = N$$

$$\boldsymbol{P = ?}$$

$$LI = \hat{P} - z_{1-\alpha/2} * \sqrt{\hat{P}(1-\hat{P})/n}$$

$$LS = \hat{P} + \underbrace{z_{1-\alpha/2} * \sqrt{\hat{P}(1-\hat{P})/n}}$$

E: Margen de Error

$$n = \frac{Z_{1-\alpha/2}^2 * \hat{P}(1-\hat{P})}{E^2}$$

POBLACIÓN FINITA:

$$n = \frac{\dfrac{Z_{1-\alpha/2}^2 * \hat{P}(1-\hat{P})}{E^2}}{\dfrac{N-1}{N} + \dfrac{\dfrac{Z_{1-\alpha/2}^2 * \hat{P}(1-\hat{P})}{E^2}}{N}}$$

Factores:

- Proporción: Usar estudios previos o pilotos. Usar 0.5 y maximizar n.

Tamaño de muestra n

Ejemplo:

Cuál es el tamaño muestral mínimo necesario (con una confianza del 95%) para estimar la media de la estatura de mujeres policías con una precisión de 1cm (máximo error permisible):

$$n_0 = \frac{Z_{1-\alpha/2}^2 * \sigma^2}{E^2} = \frac{1.96^2 * 4.30^2}{1} = 71.03 \cong 72$$

Se aconseja redondear por derecha.

Si se conociera el tamaño poblacional y este fuera de $N = 1000$, por ejemplo:

$$n = \frac{n_o}{\frac{N-1}{N} + \frac{n_o}{N}} = \frac{71.03}{\frac{999}{1000} + \frac{71.03}{1000}} = 66.38 \cong 67$$

Intervalos de confianza para diferencia de medias

Bajo el supuesto de normalidad en dos grupos independientes:

- Si las varianzas se consideran iguales:

$$\bar{X}_1 - \bar{X}_2 \pm t_{1-\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

- Si las varianzas se consideran diferente:

$$\bar{X}_1 - \bar{X}_2 \pm t_{1-\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$$

Si $0 \epsilon (LI, LS)$ las dos medias se consideran iguales
Si $LI, LS > 0$ Media de la población 1 es mayor que la media de la población 2
Si $LI, LS < 0$ Media de la población 1 es menor que la media de la población 2

IC para diferencia de proporciones

$$LI = \hat{p}_1 - \hat{p}_2 - z\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)}$$

$$LS = \hat{p}_1 - \hat{p}_2 + z\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)}$$

Si $0 \epsilon (LI, LS)$ las dos proporciones se consideran iguales
Si $LI, LS > 0$ Proporción de la población 1 es mayor que la proporción de la población 2
Si $LI, LS < 0$ Proporción de la población 1 es menor que la proporción de la población 2

Próxima sesión

Tema 8_I: Contraste de hipótesis

- Introducción.
- Dos tipos de error en la significancia estadística.
- Pasos a seguir en un contraste de hipótesis.
- Contrastes de hipótesis para una media.

- Presentación de la última actividad. Estadística Inferencial (entrega 17 de febrero)

- *Laboratorio última actividad en la semana siguiente (060223-100223).*

LA UNIVERSIDAD
EN INTERNET

www.unir.net