Intervalos de Confianza

Tema 6

Introducción

Para estudiar parámetros **poblacionales** (como la media o la proporción) de una población cualquiera, debemos tomar una muestra aleatoria de dicha población y a través de ella, calcular una aproximación a dichos parámetros **poblacionales** que desconocemos y queremos estimar. A esa aproximación se le llama estimación.

Además, junto a esa estimación, y dado que muy probablemente no coincida con el valor real del parámetro, acompañaremos el error aproximado que se comete al realizarla.

El objetivo de la estimación paramétrica pues, es determinar características de las poblaciones a partir de las observaciones realizadas a las muestras.

La estadística descriptiva se encarga de estimar los parámetros (media, varianza…) de las muestras. El problema que se plantea es cómo a partir de esos parámetros estimados para una muestra, podemos averiguar los parámetros poblacionales.

La extrapolación de los datos muestrales a los poblacionales constituye el objetivo de la inferencia paramétrica.

Introducción

- La inferencia paramétrica se divide en dos categorías generales:
  - Estimación: puede ser
    - Estimación puntual
    - Estimación por intervalos
  - Contraste de hipótesis

- Definiciones:
  - Un ESTADÍSTICO es una función definida sobre una o varias variables aleatorias.
  - Un ESTIMADOR de un parámetro es un estadístico utilizado para aproximar un parámetro de una población.

Contenidos

- Introducción
- Estimación
  - Estimación puntual
    - Estimación de parámetros en la distribución binomial
    - Estimación de parámetros en la distribución normal
  - Estimación por intervalos
    - Intervalo de confianza para la media de una población de varianza σ conocida
    - Intervalo de confianza para la media de una distribución normal de varianza desconocida
    - Intervalo de confianza para la varianza poblacional de una distribución normal
    - Intervalo de confianza para la proporción

Distribución binomial

Consideremos el suceso A, cuya probabilidad de ocurrencia es p, como parte de un experimento dicotómico. Si realizamos n veces el experimento, un estimador puntual de p es la proporción:

$$\hat{p} = \frac{número\ de\ veces\ que\ aparece\ el\ suceso\ A}{n}$$

Este estimador ha de cumplir que:

1. Su varianza sea mínima e igual a:

$$\sigma^2 = \frac{p(1-p)}{n}$$

2. Sea insesgado, es decir, la distribución en el muestreo de $\hat{p}$ tiene de media p
3. Si el tamaño de la muestra es suficientemente grande sigue una distribución normal:

$$\mathrm{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Distribución binomial (Ejemplo)

Un ayuntamiento desea conocer si en su ciudad se cumple la ley sobre obligatoriedad del uso del casco reglamentario en motoristas. Para ello, encarga un estudio en el que se observa que, de 464 motoristas, 131 de ellos no llevan el casco reglamentario.

Hallar un estimador puntual de la proporción de motoristas que no llevan casco en esa ciudad.

SOLUCIÓN:

Se trata de una población binomial, que se divide entre motoristas con casco (fracaso) y sin casco (éxito). Contamos (éxito) cuando van sin casco ya que queremos estimar la proporción poblacional (p) de motoristas sin casco. Para ello usamos como estimador la proporción muestral $\hat{p}$ que nos dan en el enunciado:

$$\hat{p} = \frac{\text{número de veces que aparece el suceso } A}{n} = \frac{\text{número de motoristas sin casco}}{\text{número total motoristas}} = \frac{131}{464} = 0.2823 \Rightarrow 28.23\%$$

Distribución normal

Consideremos una población cuya característica en estudio siga una distribución normal $N(\mu, \sigma)$, en la que μ es la media y σ la desviación típica.

Extraemos al azar una muestra representativa de dicha población , un estimador puntual de μ, (media poblacional) , será $\bar{x}$ (media muestral), dónde:

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

De igual manera, un estimador puntual de σ^2 (varianza poblacional) será s^2(cuasivarianza muestral), donde:

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

Y para la desviación típica σ, obtenemos como estimador s.

Distribución normal

Ambos estimadores deben ser insesgados y de varianza mínima y además:

- $\bar{X}$ sigue una distribución normal de parámetros:

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

- S^2 sigue una distribución χ^2 con n-1 grados de libertad:

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

Resumen estimación puntual

Parámetro poblacional	Estimador	Estimación
μ Media poblacional	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ Media muestral	$\overline{x}$
σ Varianza poblacional	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ Cuasivarianza muestral	s^2
p Proporción poblacional	$\hat{P}$ Proporción muestral	$\hat{p}$

Estimación por intervalos

Sea X una v.a. cuya función de densidad (pdf) depende de parámetros desconocidos, y sea θ uno de esos parámetros.

La Estimación por intervalos consiste en dar un intervalo I en el interior del cual se encontrará el parámetro θ que queremos estimar, con una probabilidad de acierto (o confianza) previamente fijada $(1-\alpha)$ y que se tratará sea lo mayor posible, es decir, lo más próxima a 1 posible.

$$P(\theta \in I) \geq 1 - \alpha$$

Estimación por intervalos

Para el cálculo de I se utilizan dos estadísticos, al primero de ellos se le llama limite inferior del intervalo y al segundo limite superior del intervalo.

El *nivel de confianza* $(1-\alpha)$ representa la probabilidad de que el intervalo contenga al parámetro poblacional que estamos estudiando. Siendo lo más habitual que $(1-\alpha) = 0.95$ o $(1-\alpha) = 0.99$.

A mayor valor de $(1-\alpha)$, más probabilidad de acierto en nuestra estimación. Esto implica que α tendrá que ser pequeño, próximo a 0.

α se denomina *nivel de significación*. Y es la probabilidad de que I no contenga al parámetro θ (probabilidad de error)

Tipificaciones asociadas a la distribución en el muestreo de los estimadores $\bar{X}$ y S para intervalos de confianza

a) Si σ es conocido, Z sigue una distribución $N(0{,}1)$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

b) Si σ es desconocido y $n > 30, Z$ sigue una distribución aproximadamente normal, $N(0{,}1)$,

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

c) Si σ es desconocido y $n < 30$, t_{n-1}, sigue una distribución t de Student con n-1 grados de libertad:

$$t_{n-1} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Intervalo de confianza para la media de una población de varianza σ^2 conocida

Para un nivel de confianza $1-\alpha$ queda:

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha \;\Rightarrow\; P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$$

El intervalo de confianza viene dado por la expresión:

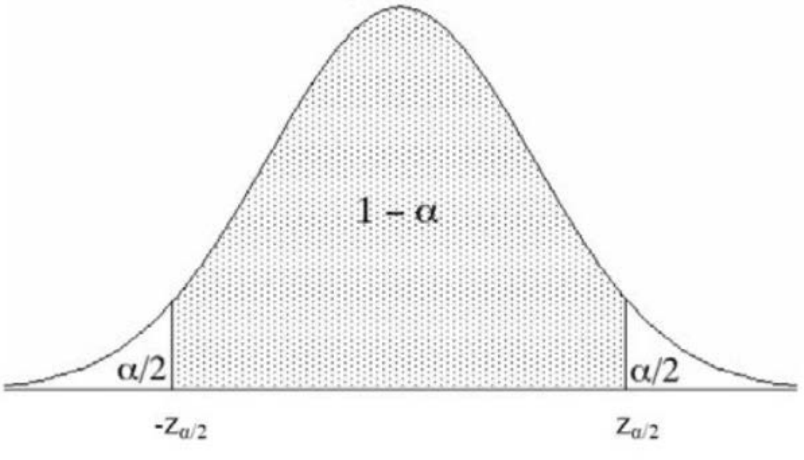
$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

Dónde: $z_{\alpha/2}$: valor de la N(0,1) a su derecha un área de $\alpha/2$

$\bar{X}$: Media muestral

σ: Desviación típica poblacional

n: Tamaño muestral



Ejemplo

Una muestra aleatoria extraída de una población normal de varianza σ^2 igual a 100, presenta una media muestral de 160. Con n=144, se pide:

a) Calcular un intervalo de confianza al 95% para la media poblacional.
b) Calcular un intervalo de confianza al 90% para la media poblacional.
c) Comparar ambos intervalos desde el punto de vista de la información que genera.

Solución

Apartado a)

$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

Conocido α determinamos el valor crítico $z_{\alpha/2}$

$$1 - \alpha = 0{,}95 \ \Rightarrow \ \alpha = 0{,}05 \Rightarrow {}^{\alpha}/_{2} = 0{,}025$$

$$1 - {}^{\alpha}/_{2} = 0{,}975$$

De la tabla de la normal estándar $z_{0,025} = 1{,}96$

luego

$$I_{0,95} = \left[160 - 1{,}96 \frac{\sqrt{100}}{\sqrt{144}}; 160 + 1{,}96 \frac{\sqrt{100}}{\sqrt{144}}\right]$$

$$\boxed{I_{0,95} = [158{,}37; 161{,}63]}$$

SOLUCIÓN

Apartado b) mismo intervalo con diferente confianza

$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

Conocido α determinamos el valor crítico $z_{\alpha/2}$

$$1 - \alpha = 0,90 \Rightarrow \alpha = 0,10 \Rightarrow {\alpha}/{2} = 0,05$$

$$1 - {\alpha}/{2} = 0,95$$

De la tabla de la normal estándar $z_{0,05} = 1,64$

Luego

$$I_{0,90} = \left[160 - 1,64 \frac{\sqrt{100}}{\sqrt{144}}; 160 + 1,64 \frac{\sqrt{100}}{\sqrt{144}}\right]$$

$$\boxed{I_{0,90} = [158,63; 161,37]}$$

SOLUCIÓN

Calculamos la longitud de cada uno de los intervalos:

$$I_{0,95} = [161{,}63 - 158{,}37] = 3{,}26$$

$$I_{0,90} = [161{,}37 - 158{,}63] = 2{,}74$$

El segundo intervalo es más corto, pero su *confianza*, es decir, la probabilidad con la que contiene en su interior a la media poblacional μ, es menor:

$$P(\mu \in I_{0,95}) \geq 0{,}95 \text{ en el apartado a)}$$

$$P(\mu \in I_{0,90}) \geq 0{,}90 \text{ en el apartado b)}$$

Intervalo de confianza para la media de una distribución normal de varianza σ^2 desconocida

El estadístico pivote es: Utilizando la s (cuasi-desviación típica muestral):

$$t_{n-1} \sim \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

t_{n-1} es la distribución t de Student con n-1 grados de libertad

Intervalo de confianza para la media de una población, con varianza desconocida y n<30

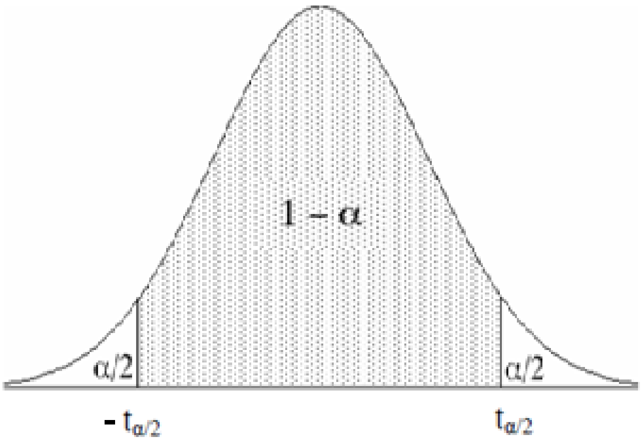
$$P\left(-t_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} < t_{\alpha/2}\right) = 1-\alpha \;\Rightarrow\; P\left(\bar{X} - \frac{s}{\sqrt{n}} t_{\alpha/2} < \mu < \bar{X} + \frac{s}{\sqrt{n}} t_{\alpha/2}\right) = 1-\alpha$$

El intervalo de confianza en este caso es:

$$I = \left[\bar{X} \pm t_{n-1;\alpha/2} \frac{s}{\sqrt{n}}\right]$$

Dónde: $t_{n-1;\alpha/2}$: valor de la t de Student de n-1 grados de libertad, que deja a su derecha un área de $^{\alpha}/_{2}$

- $\bar{X}$: Media muestral
- s: cuasi-desviación típica muestral
- n: Tamaño muestral



gráfica de la pdf de la distribución t de Student

Ejemplo

Se ha recogido una muestra aleatoria para prever la inflación anual en siete países diferentes. Las previsiones han sido:

1.5 ; 2.1 ; 1.9 ; 2.3 ; 2.5 ; 3.2 ; 3.0

a) Utilizando estos datos construir un intervalo de confianza al 99% para la media de la previsión de inflación en estos siete países.

b) Si los expertos opinan que el intervalo de confianza calculado para estimar la media es demasiado amplio y desean que su longitud sea de 1 punto, hallar el nivel de confianza para este nuevo intervalo.

SOLUCIÓN

Apartado a)

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$\bar{x} = \frac{1{,}5 + 2{,}1 + 1{,}9 + 2{,}3 + 2{,}5 + 3{,}2 + 3{,}0}{7} = 2{,}357 \cong 2{,}4$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0{,}36 \Rightarrow s = \sqrt{0{,}36} = 0{,}6$$

Como σ (desviación típica poblacional) es desconocida, usamos la s (cuasi-desviación típica muestral), junto con el valor crítico de la t de Student:

$$I = \left[\bar{x} \pm t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right]$$

$$1 - \alpha = 0{,}99 \Rightarrow \alpha = 0{,}01 \Rightarrow \alpha/2 = 0{,}05$$

$1 - \alpha/2 = 0{,}995$ para mirar en la tabla de la t de Student

Solución

Buscamos en la tabla de la t de Student con $n-1$ grados de libertad ($7-1=6$) el valor crítico $t_{6;0,05}=3,7$

$$I=\left[2,4-3,7\frac{0,6}{\sqrt{7}};2,4+3,7\frac{0,6}{\sqrt{7}}\right]$$

$$I=[1,6;3,2]\text{ con P}=99\%$$

Apartado b) el ancho del intervalo anterior es $L=2t_{n-1;\alpha/2}\frac{0,6}{\sqrt{7}}$.

Si queremos que ahora sea igual a 1, hacemos

$$1=2t_{\alpha/2}\frac{0,6}{\sqrt{7}}$$

$\Rightarrow 1=0,4535t_{\alpha/2}\Rightarrow t_{\alpha/2}=\frac{1}{0,4535}=2,226$ lo que, mirando en la tabla (o en una calculadora de valores críticos de la t de Student, como MATLAB), corresponde a un valor de $\alpha=0,068$, o lo que es lo mismo, el valor de confianza para el nuevo intervalo es de:

$$1-\alpha=1-0,068=0,932\ \Rightarrow 93,2\%$$

Intervalo de confianza para la varianza poblacional

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

El intervalo de confianza para la varianza poblacional es:

$$I = \left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right]$$

Dónde:

- $\chi^2_{n-1;\alpha/2}$: Chi cuadrado con n-1 grados de libertad que deja a su derecha una probabilidad (un área) de $\alpha/2$
- s^2: cuasi-varianza muestral
- n: Tamaño de la muestra

Ejemplo

La puntuación media de 20 alumnos de un curso de estadística elegidos al azar presentó una cuasidesviación típica muestral $s = 0{,}0965$. Suponiendo que la variable sigue una distribución normal, calcular un intervalo de confianza al 95% de probabilidad para la varianza poblacional.

$s = 0{,}0965; n = 20; \; n - 1 = 19$

$1 - \alpha = 0{,}95; \; \alpha = 0{,}05;$

${}^{\alpha}/_{2} = 0{,}025; \; 1 - {}^{\alpha}/_{2} = 0{,}975$

$$I = \left[\frac{(n-1)s^2}{\chi^2_{n-1,{}^{\alpha}/_{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-{}^{\alpha}/_{2}}} \right]$$

$s^2 = 0{,}00931225$

$\chi^2_{19;0.025} = 32{,}85; \; \chi^2_{19;0.975} = 8{,}91$

$$I = \left[\frac{19 \cdot 0{,}00931225}{32{,}85}, \frac{19 \cdot 0{,}00931225}{8{,}91} \right]$$

$$= (0{,}0053; 0{,}019 \;)$$

INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

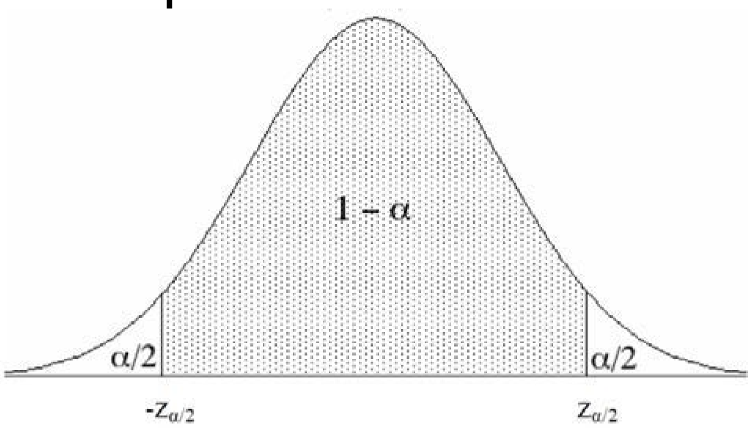
$$P\left(\hat{P} - \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\, z_{1-\alpha/2} < p < \hat{P} + \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\, z_{1-\alpha/2}\right) = 1-\alpha$$

El intervalo queda:

$$I = \left[\hat{P} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right]$$

Dónde:

- $\hat{P}$: proporción muestral
- n: Tamaño de la muestra



Ejemplo

- Una conocida marca de detergentes realiza una encuesta a 100 amas de casa elegidas al azar, para consultarles sobre si están o no satisfechas con el detergente que usan para lavar la ropa. El 55% de ellas afirma estar muy satisfechas con los resultados obtenidos, y que volverían a comprar la misma marca del citado detergente. Con estos datos calcular un intervalo del confianza al 99% para la proporción de amas de casa satisfechas con los resultados obtenidos en la colada con el citado detergente.

$\hat{p} = 0{,}55; n = 100; 1 - \alpha = 0{,}99;\ \alpha = 0{,}01;$
${}^{\alpha}/_{2} = 0{,}005; 1 - {}^{\alpha}/_{2} = 0{,}995$
$Z_{0{,}995} = 2{,}57$

$$I = \left[\hat{p} \pm z_{1-{}^{\alpha}/_{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$I = \left[0{,}55 - \sqrt{\frac{0{,}55 \cdot 0{,}45}{1000}}\,2{,}57; 0{,}55 + \sqrt{\frac{0{,}55 \cdot 0{,}45}{1000}}\,2{,}57\right]$$

$$I = [0{,}4222; 0{,}6778]$$