

Intro a la Probabilidad y estadística



Martes y Jueves Aula B17
Dra Ana Georgina Flesia

Ejemplo

La calibración de una balanza debe ser revisada al pesar 25 veces un espécimen de 10 Kg. Suponga que los resultados de los diferentes pesos son independientes entre si y que la variable peso esta normalmente distribuida con un desvío estándar $\sigma = 0.20$ Kg. Sea μ el verdadero valor medio de lectura de peso de la balanza.

- a) ¿Cual es el nivel de confianza para el intervalo $\bar{x} \pm 2.81 \frac{\sigma}{\sqrt{n}}$ para μ ?
- b) ¿Cual es el valor de $z_{\alpha/2}$ para un IC del 99,7 % para μ ?
- c) ¿Que tan grande debería ser el tamaño de muestra tal que la longitud del IC del 95 % para μ sea a lo sumo de 0.05?
- d) Si de la muestra observada se obtuvo un promedio y desvío estándar muestrales de $\bar{x} = 10.30$ Kg y $s_{n-1} = 0.19$ Kg respectivamente, obtenga un IC del 95 % para μ . Compare el intervalo obtenido con el intervalo del inciso anterior.

Resolución

- a) Si tengo un intervalo $\bar{x} \pm 2.81 \frac{\sigma}{\sqrt{n}}$ para μ con una muestra normal, entonces

$$2.81 = z_{\alpha/2} = \text{percentil}(1 - \alpha/2)$$

Por lo cual, buscando 2.81 en la tabla normal patrón resulta

$$(1 - \alpha/2) = 0.9975 \Rightarrow \alpha = 2(1 - 0.9975) = 0.005$$

y el nivel del intervalo es $(1 - \alpha) = 0.995$.

- b) Si el nivel del IC es $(1 - \alpha) = 0.997$, entonces $\alpha = 0.003$ y $\alpha/2 = 0.0015$. Buscando el percentil en la tabla resulta

$$z_{\alpha/2} = \text{percentil}(1 - \alpha/2) = \text{percentil}(0.9985) = 2.97$$

- c) Para que la longitud del IC del 95 % para μ sea a lo sumo de 0.05?

$$n \geq \left(2 \frac{z_{0.025} \sigma}{L} \right)^2 = \left(2 \frac{1.960.2}{0.05} \right)^2 = 245.86$$

Por lo cual se necesitan mas de 246 especímenes para obtener esa longitud y el mismo nivel del intervalo.

Resolución

- d) Si de la muestra observada se obtuvo un promedio y desvío estándar muestrales de $\bar{x} = 10.30$ Kg y $s_{n-1} = 0.19$ Kg respectivamente, el IC del 95 % para μ con varianza estimada implica el uso de la tabla t de student.

$$10.30 \pm t_{0.05/2, 24} \frac{s}{\sqrt{25}} = 10.30 \pm 2.0639 \frac{0.19}{\sqrt{25}} = 10.30 \pm 0,0784282$$

Si la distribución es normal con $\sigma = 0.2$ el intervalo de nivel 95% es

$$10.30 \pm z_{0.05/2} \frac{\sigma}{\sqrt{25}} = 10.30 \pm 1.96 \frac{0.2}{\sqrt{25}} = 10.30 \pm 0,0784$$

Y son bastante parecidos.

Intervalos de confianza para la varianza

CASO C: muestra $N(\mu, \sigma^2)$ con $\mu, \sigma > 0$ desconocida

Sea X_1, X_2, \dots, X_n una m.a. con distribución $N(\mu, \sigma^2)$, con μ y σ^2 desconocidos.

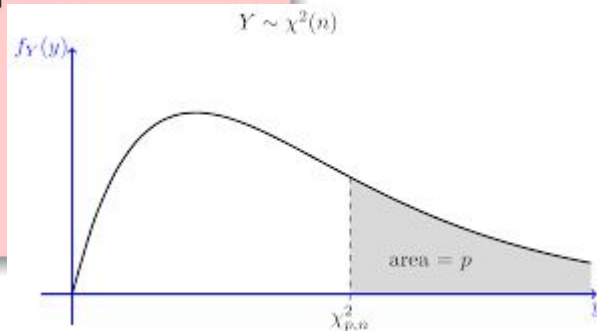
- 1ro) Sabemos que

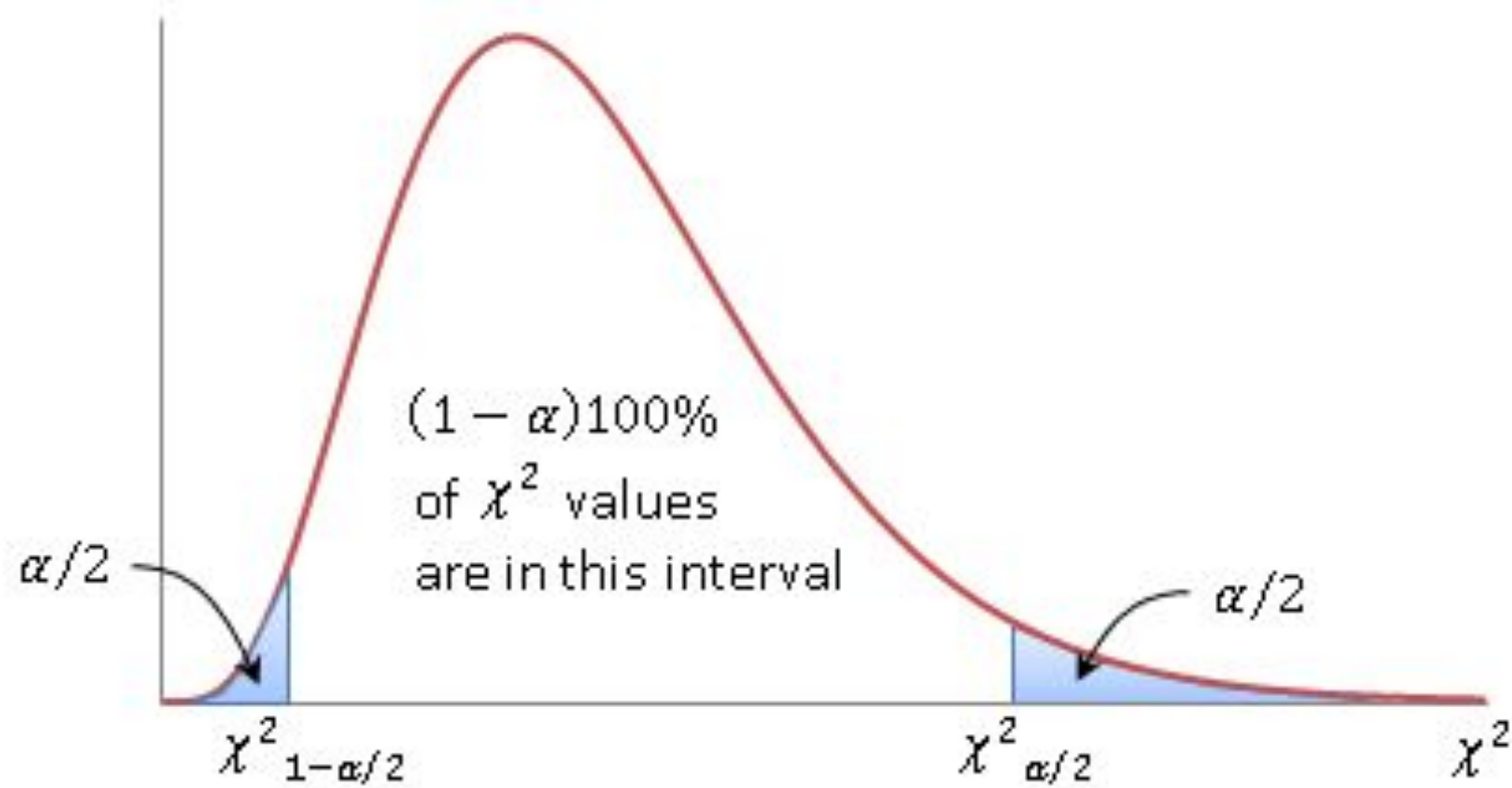
$$h(X_1, X_2, \dots, X_n; \theta) = (n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

- 2do) Fijado un nivel de confianza $(1 - \alpha)$, hay que buscar en la tabla de la distribución χ_{n-1}^2 los valores de a y b tales que cumplan

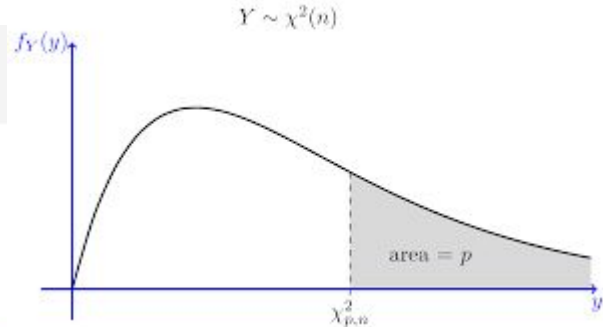
$$P(a \leq (n-1) \frac{S^2}{\sigma^2} \leq b) = 1 - \alpha$$

Estos valores son $a = \chi_{1-\alpha/2, n-1}^2$ y $b = \chi_{\alpha/2, n-1}^2$.





Intervalos de confianza para la varianza



CASO C: muestra $N(\mu, \sigma^2)$ con μ y σ^2 desconocida

Sea X_1, X_2, \dots, X_n una m.a. con distribución $N(\mu, \sigma^2)$, con μ y σ^2 desconocidos.

► 3ro) Despejando σ^2 resulta

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

Intervalos de confianza para la $N(\mu, \sigma^2)$

Intervalos de confianza % 100 $(1 - \alpha)$

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\bar{X} = \sum_{i=1}^n X_i/n, \quad S = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)}$$

| Hipótesis | Parámetro | Intervalo de Confianza |
|------------------------|------------|---|
| σ^2 conocido | μ | $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ |
| σ^2 desconocido | μ | $\bar{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$ |
| μ desconocido | σ^2 | $\left[\frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2} \right]$ |

Ejemplo

Se efectuaron las siguientes observaciones de resistencia a la fractura de placas base de 18 % de acero maragizado al níquel:

69.571.972.673.173.373.575.575.775.876.176.2

76.277.077.978.179.679.779.980.182.283.783.7

Asuma que la muestra proviene de una distribución $N(\mu; \sigma^2)$.

- a) Construir un intervalo de confianza del 99 % para la resistencia media a la fractura.
- b) Construir un intervalo de confianza del 99 % para la desviación estándar poblacional de la resistencia a la fractura.

Ejemplo

Asuma que la muestra proviene de una distribución $N(\mu; \sigma^2)$. Como no tenemos datos de la media o la varianza, debemos estimarlos de la muestra

a) Se tiene que

$$\bar{x} = 76.88 \quad s_{21} = 3.784$$

Si $\alpha = 0.01$, el valor crítico con 21 grados de libertad es

$$t_{0.01/2, 21} = 2.8314$$

Por lo cual el intervalo de confianza del 99 % para la resistencia media a la fractura es.

$$\bar{x} \pm t_{0.01/2, 21} s_{n-1} / \sqrt{n} = 76.88 \pm 3.784 \times 2.8314 / 4.69 = 76.88 \pm 2.284$$

b) Un intervalo de confianza del 99 % para la desviación estándar poblacional de la resistencia a la fractura es

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

Ejemplo

Esto es

$$\left[\frac{21 \times 14.324}{\chi_{0.005,21}^2}, \frac{21 \times 14.324}{\chi_{0.995,21}^2} \right]$$

De la tabla resulta $\chi_{0.995,21}^2 = 8.034$ y $\chi_{0.005,21}^2 = 41.401$

$$\left[\frac{21 \times 14.324}{41.401}, \frac{21 \times 14.324}{8.034} \right] = [7.265, 37.43]$$

Intervalos de confianza asintóticos

CASO D: muestra aleatoria de tamaño grande σ conocido

Sea X_1, X_2, \dots, X_n una m.a. con $E(X_k) = \mu$ y $Var(X_k) = \sigma^2$ y n suficientemente grande para que valga el Teorema Central del Límite. Entonces se tiene que $\bar{X} = \sum_{i=1}^n X_i/n$ tiene distribución aproximadamente $N(\mu, \sigma^2/n)$.

- ▶ 1ro) Sabemos que si $n \geq 30$ y σ es conocido

$$h(X_1, X_2, \dots, X_n; \theta) = \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

- ▶ 2do) Fijado un nivel de confianza $(1 - \alpha)$,

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

es un intervalo de confianza aproximado para μ

Intervalos de confianza aproximados

CASO D: muestra aleatoria de tamaño grande σ desconocido

Sea X_1, X_2, \dots, X_n una m.a. con $E(X_k) = \mu$ y $Var(X_k) = \sigma^2$ ambos desconocidos

- 1ro) Sabemos que si $n \geq 40$ y n suficientemente grande para que valga el Teorema Central del Límite, por el teorema de Slutsky, como S^2 converge a σ^2 , resulta

$$h(X_1, X_2, \dots, X_n; \theta) = \sqrt{n} \frac{(\bar{X} - \mu)}{S_{n-1}} \sim N(0, 1)$$

- 2do) Fijado un nivel de confianza $(1 - \alpha)$,

$$\bar{X} \pm z_{\alpha/2} \frac{S_{n-1}}{\sqrt{n}}$$

es un intervalo de confianza aproximado para μ

Ejemplo

Se tiene una muestra de la duración de eco de radar de 110 relámpagos en cierta región. El promedio muestral fue de 0.81 segundos y el desviación estándar muestral (s_{n-1}) de 0.34 segundos. Calcular un IC de nivel aproximado 0.99 para la media de duración de eco μ

Fijado un nivel de confianza $(1 - \alpha) = 0.99$,

$$\bar{X} \pm z_{\alpha/2} \frac{s_{n-1}}{\sqrt{n}} = 0.81 \pm 2.575 \times 0.34/\sqrt{110} = 0.81 \pm 0.0835$$

es un intervalo de confianza aproximado para μ

Ejemplo: Muestra Bernoulli

Sea X_1, X_2, \dots, X_n una m.a.con distribución Bernoulli(p) y tamaño de muestra suficientemente grande. Entonces por TCL resulta que

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

donde $p = E(X)$ y $V(X) = p(1 - p)$. Denotaremos con $\hat{p} = \bar{X}$, entonces

$$\frac{(\hat{p} - p)}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

Ejemplo: Muestra Bernoulli

Planteando la ecuación:

$$P \left(\left| \frac{(\hat{p} - p)}{\sqrt{\frac{p(1-p)}{n}}} \right| \leq z_{\alpha/2} \right) \simeq (1 - \alpha)$$

y trabajando con el evento elevado al cuadrado se puede obtener una ecuación cuadrática en p donde las raíces de esa ecuación son la cota inferior y superior de un intervalo de confianza aproximado $(1 - \alpha)$ para p , resultando:

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

Ejemplo: Muestra Bernoulli

Para n suficientemente grande:

- ▶ $\frac{z_{\alpha/2}^2}{2n}$ es insignificante en comparación con \hat{p} ;
- ▶ $\frac{z_{\alpha/2}^2}{4n}$ es insignificante en comparación con $\frac{\hat{p}(1-\hat{p})}{n}$ y
- ▶ $\frac{z_{\alpha/2}^2}{n}$ es insignificante en comparación con 1.

Luego desechando esos términos insignificantes resulta este IC aleatorio de nivel aproximado $(1 - \alpha)$ para p tradicional

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

siempre que $n\hat{p} \leq 10$ y $n(1 - \hat{p}) \geq 10$

Ejemplo: Muestra Bernoulli

Si se quiere calcular el tamaño de muestra necesario para que un IC para p tenga nivel $(1 - \alpha)$ y longitud L , entonces

$$n \geq \left(\frac{z_{\alpha/2}}{L} \right)^2$$

independendientemente del valor de la proporción observada \hat{p}

Ejemplo: Muestra Bernoulli

En un artículo sobre estimación de fuentes de defectos visuales, se reporta que se estudiaron con un sensor de inspección 356 matrices de silicio de las cuales 201 pasaron la prueba.

- a) Construir un intervalo de confianza de nivel aproximado 0.98 para la proporción poblacional de matrices que pasan la inspección.
- b) ¿Que tamaño de muestra seria necesario para que la longitud un intervalo de confianza de nivel aproximado 0.98 para la proporción poblacional de matrices que pasan la inspección sea a lo sumo 0.05, independientemente del valor \hat{p} ?

resolución

- a) Un intervalo de confianza de nivel aproximado 0.98 para la proporción poblacional es

$$\begin{aligned}\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.565 \pm \text{percentil}(1 - \alpha/2) \sqrt{\frac{(0.565)(1 - 0.565)}{356}} \\ &= 0.565 \pm 2.325 \times \sqrt{\frac{0.565 \times 0.435}{356}} \\ &= 0.565 \pm 0.061\end{aligned}$$

Pues si $\alpha = 0.02$, entonces $\alpha/2 = 0.01$ y $\text{percentil}(0.99) = 2.325$.

- b) El tamaño de muestra necesario para que $L = 0.05$ es

$$n \geq \left(\frac{z_{0.01}}{0.05}\right)^2 = (2.325/0.05)^2 = 2162.25$$