TEMA 7. Estimación

Alicia Nieto Reyes

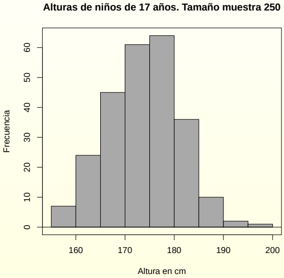
BIOESTADÍSTICA

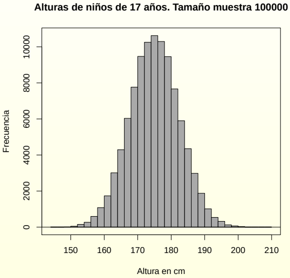
1. Estimación Puntual

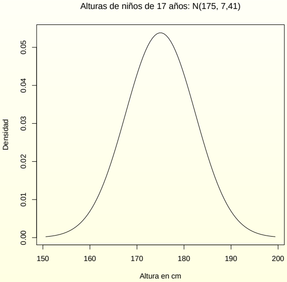
1. Estimación por intervalos
   - Estimación por intervalos de la **Media** de una Variable Normal de **Varianza Conocida**
   - Estimación por intervalos de la **Media** de una Variable Normal de **Varianza Desconocida**
   - Elección del **Tamaño Muestral** con Varianza Conocida
   - Estimación por intervalos de la **Varianza** de una Variable Normal

μ=media poblacional $\qquad$ σ=desviación típica poblacional

$\bar{X}=\hat{\mu}$=media muestral $\qquad$ $S=\hat{\sigma}$=desviación típica muestral







$\bar{X} = 173.85$	$\bar{X} = 175$	$\mu = 175$
$S = 7.38$	$S = 7.41$	$\sigma = 7.41$

Sin redondear por dos decimales:

$\bar{x} = 173.8511$	$\bar{x} = 174.9979$
$S = 7.375297$	$S = 7.4139$

A partir de aquí queremos estimar μ y σ por lo que supondremos que son desconocidos

Estimación por intervalos de la Media de una Variable Normal de Varianza Conocida

Dado una muestra podemos calcular $\bar{X}$. Pero $\bar{X}$ no es necesariamente μ

No se puede predecir el valor de $\bar{X}$ antes de tomar la muestra, entonces $\bar{X}$ es una variable aleatoria

- $\bar{X}$ y S son variables
- μ y σ son constantes

Resultado

Si tomamos una muestra al azar de tamaño n y X sigue una distribución $N(\mu, \sigma)$, entonces $\bar{X}$ sigue una distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$

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

Estimación por intervalos de la Media de una Variable Normal de Varianza Conocida

Dada $Y \sim N(0,1)$, $P(-1.96 \leq Y \leq 1.96) = 0.95$

Entonces

$$
\begin{aligned}
P(-1.96 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq 1.96) &= P(\frac{-1.96\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq \frac{1.96\sigma}{\sqrt{n}}) \\
&= P(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96\sigma}{\sqrt{n}}) \\
&= 0,95
\end{aligned}
$$

Ejemplo: Conocemos $\sigma = 7.41$.
Si la muestra obtenida ha sido la anterior ($n = 100000$ y $\bar{X} = 174.9979$):

entonces, el intervalo de confianza al nivel 0.95 para μ es

$$(174.9520, 175.0438)$$

Estimación por intervalos de la Media de una Variable Normal de Varianza Conocida

El intervalo de confianza al nivel 0.95 para μ es $(174.9520, 175.0438)$

No quiere decir: La probabilidad de que μ esté entre 174.9520 y 175.0438 es 0.95

Porque: μ es un valor desconocido pero fijo, no es una variable

Quiere decir: Si cogemos muchas veces una muestra de tamaño 100000 y calculamos $\bar{X}$ cada vez, en el 95% de las veces μ está comprendido entre $\bar{X} - 0.0459$ y $\bar{X} + 0.0459$

Nota que $\dfrac{1.96 \cdot 7.41}{\sqrt{100000}} = 0.0459$

Los intervalos de confianza al nivel $1 - \alpha$ suelen construirse para un nivel de confianza

$$1 - \alpha = 0.95 \text{ ó } 0.975 \text{ ó } 0.99 \text{ ó } 0.999$$

Estimación por intervalos de la Media de una Variable Normal de Varianza Desconocida

Resultado

Si tomamos una muestra al azar de tamaño n y X sigue una distribución $N(\mu, \sigma)$, *entonces la variable* $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$ *sigue una distribución t de Student con* $n-1$ *grados de libertad*

Dada $Z \sim t_{99999}$, $P(-1.959988 \leq Z \leq 1.959988) = 0.95$

Entonces,

$$P(-1.959988 \leq Z \leq 1.959988) =$$
$$P(-1.959988 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq 1.959988) =$$
$$P(\bar{X} - \frac{1.959988S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.959988S}{\sqrt{n}}) = 0.95$$

Estimación por intervalos de la Media de una Variable Normal de Varianza Desconocida

Ejemplo: Si la muestra obtenida ha sido la anterior

$$(n = 100000, \bar{X} = 174.9979 \text{ y } S = 7.4139)$$

Como

$$P(\bar{X} - \frac{1.959988S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.959988S}{\sqrt{n}}) = 0.95$$

El intervalo de confianza al nivel 0.95 para μ es

$$(174.9519, 175.0439)$$

- Nota que este intervalo de confianza es algo mayor que el que construimos cuando la varianza era conocida,

  $$(174.9520, 175.0438),$$

  debido a la perdida de información que aquí hemos tenido

Estimación por intervalos de la Media de una Variable Normal de Varianza Desconocida

Tenemos que $P(-1.969537 \leq Z \leq 1.969537) = 0.95$ con $Z \sim t_{249}$, Como

$$P(\bar{X} - \frac{1.969537S}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.969537S}{\sqrt{n}}) = 0.95$$

Si la muestra es la dada anteriormente con $n = 250$, el intervalo de confianza al nivel 0.95 para μ es $(172.9324, 174.7698)$

- Nota que el intervalo de confianza es mayor que con $n = 100000$,

  $$(174.9519, 175.0439),$$

  porque al tener una muestra menor tenemos menos información
- Si repetimos el procedimiento con diferentes muestras de tamaño $n = 250$, este es uno del 5% de casos en que μ

  (su verdadero valor es 175)

  no está entre $\bar{X} - 0.1245645 \cdot S$ y $\bar{X} + 0.1245645 \cdot S$

  Nota que $\dfrac{1.969537}{\sqrt{250}} = 0.1245645$

Estimación por intervalos de la Media de una Variable Normal de Varianza Desconocida

1. Utilizando la muestra de $n = 100000$:

   Tenemos que $P(-2.575878 \leq Z \leq 2.575878) = 0,99$ con $Z \sim t_{99999}$,
   El intervalo de confianza al nivel 0.99 para μ es $(174.9375, 175.0583)$

   - Para exigir un nivel de confianza mayor necesitamos un intervalo mayor (Con la misma muestra, el intervalo de confianza al nivel 0.95 para μ es $(174.9519, 175.0439)$)

2. Utilizando la muestra de $n = 250$

   Tenemos que $P(-2.595718 \leq Z \leq 2.595718) = 0.99$ con $Z \sim t_{249}$,

   El intervalo de confianza al nivel 0.99 para μ es $(172.6403, 175.0619)$

   - Nota que aunque con esta misma muestra y un intervalo de confianza a un nivel menor μ no esta en el intervalo;
     (intervalo de confianza al nivel 0.95 para μ es $(172.9324, 174.7698)$)
     si lo esta cuando el nivel sube a 0.99

Elección del Tamaño Muestral con Varianza Conocida

Denotamos por d_α al nº tal que $P(-d_\alpha \leq Y \leq d_\alpha) = \alpha$ con $Y \sim N(0,1)$

Entonces $P(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96\sigma}{\sqrt{n}}) = 0,95$

¿n necesario para que la longitud del intervalo de confianza sea $\leq L$?

$$L \geq \bar{X} + \frac{d_\alpha \sigma}{\sqrt{n}} - (\bar{X} - \frac{d_\alpha \sigma}{\sqrt{n}}) = 2\frac{d_\alpha \sigma}{\sqrt{n}}$$

Es decir, $n \geq (\frac{2 d_\alpha \sigma}{L})^2$

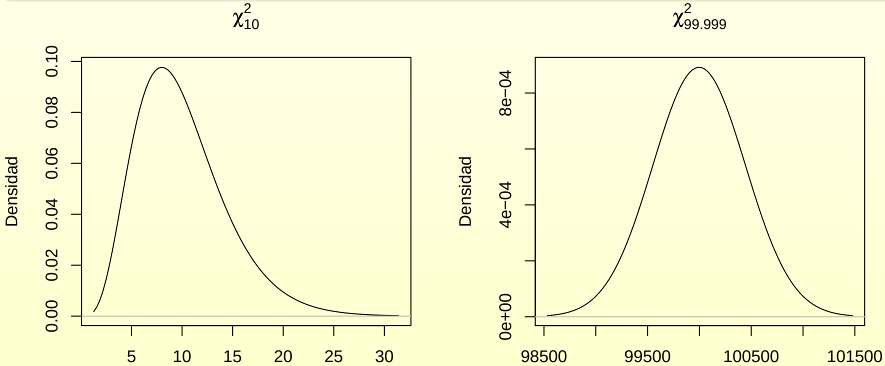
Ejemplo: Si seguimos en el problema de las alturas $\sigma = 7.41$

- Para obtener un intervalo de confianza de longitud menor o igual a 2 al nivel 0.95 basta con $n \geq 216$ porque $n \geq (1.96 \cdot 7.41)^2 = 215.2617$
- Para obtener un intervalo de confianza de longitud menor o igual a 1 al nivel 0.95 basta con $n \geq 862$ porque $n \geq (2 \cdot 1.96 \cdot 7.41)^2 = 861.0469$

Estimación por intervalos de la Varianza de una Variable Normal

Resultado

Si tomamos una muestra al azar de tamaño n y X sigue una distribución $N(\mu, \sigma)$, *entonces la variable* $\dfrac{(n-1)S^2}{\sigma^2}$ *sigue una distribución ji-cuadrado con* $n-1$ *grados de libertad*



Estimación por intervalos de la Varianza de una Variable Normal

Dada $Y \sim \chi^2_{99999}$, $P(99124.38 \leq Y \leq 100877.4) = 0.95$

Entonces,

$$P(99124.38 \leq Y \leq 100877.4) =$$
$$P(99124.38 \leq \frac{(n-1)S^2}{\sigma^2} \leq 100877.4) =$$
$$P(\frac{1}{100877.4} \leq \frac{\sigma^2}{(n-1)S^2} \leq \frac{1}{99124.38}) =$$
$$P(\frac{(n-1)S^2}{100877.4} \leq \sigma^2 \leq \frac{(n-1)S^2}{99124.38}) = 0.95$$

Ejemplo: Si la muestra ha sido la anterior ($n = 100000$, y $S = 7.4139$):

El intervalo de confianza al nivel 0.95 para σ es $(54.48729, 55.4509)$