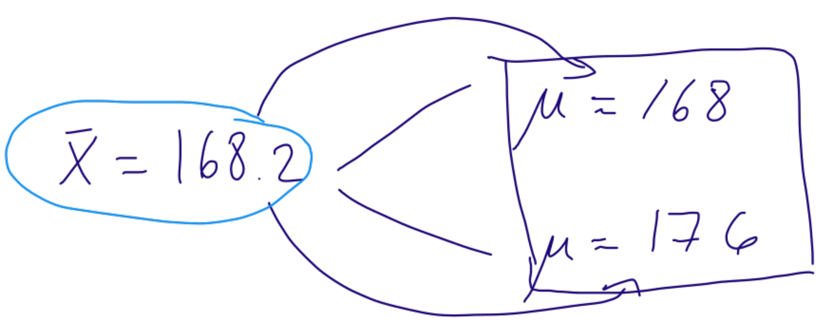
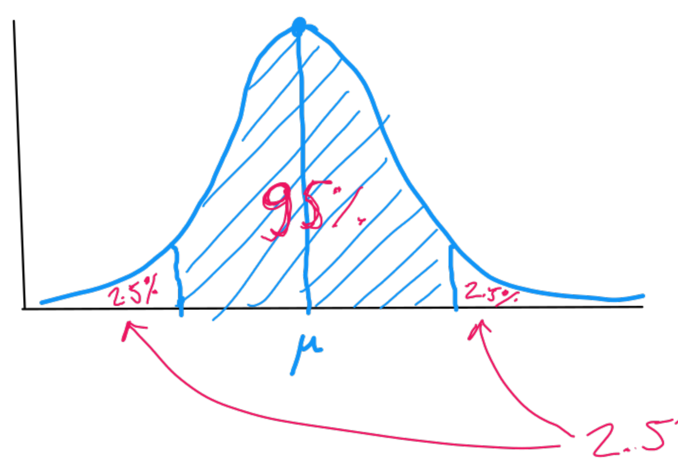
Estimación por intervalos



En la sección anterior, vimos cómo obtener estimaciones puntuales para el o los parámetros de interés, mismas que podemos entender como nuestra mejor opción (o "best guess"). Sin embargo, al tratarse de puntos fijos, son estimaciones poco flexibles.

Para resolver esto, existen métodos de estimación que proveen un conjunto de posibles valores para el parámetro. Uno de ellos es el de intervalos de confianza.

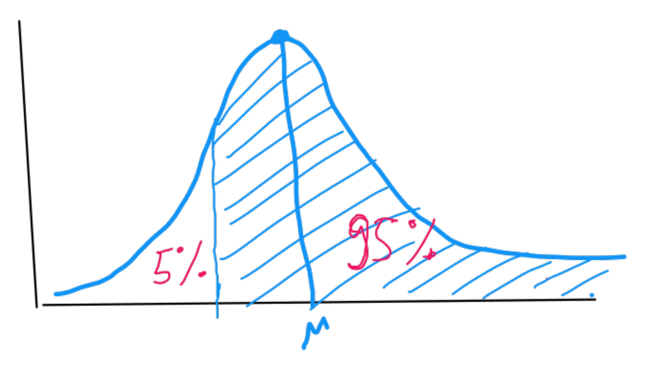


$X_i \sim N(\mu, \sigma^2) \quad i = 1, \dots, n$

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

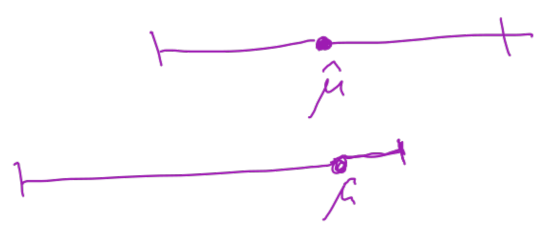
$\hat{\mu}_{MV} = \bar{X}$

$2.5\% + 2.5\% = 5\% = \alpha$



$\hat{\mu} = \bar{x} = 14.3$

$\mu \in [13.2, 14.7]$



Def. Sea $X_1, \dots, X_n$ una m.a. con f.d.p $f_X(x; \theta)$, con θ el parámetro de interés. Sean $T_1 = L(X_1, \dots, X_n)$ y $T_2 = U(X_1, \dots, X_n)$ dos estadísticos tales que $T_1 \le T_2$, de modo que para una muestra aleatoria observada $x_1, \dots, x_n$ ocurrirá idealmente que

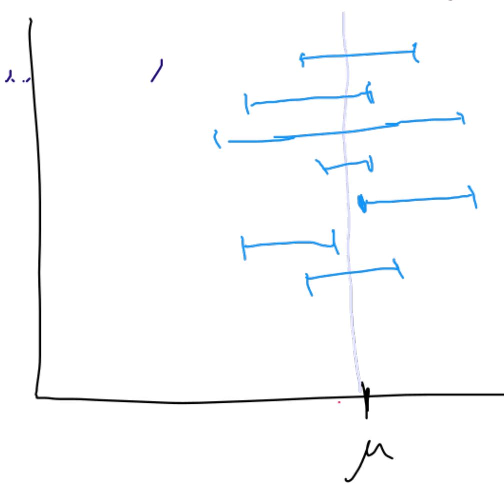
$$T_1 \le \theta \le T_2$$

$$\underbrace{L(\underline{X})}_{\text{lower bound}} \le \theta \le \underbrace{U(\underline{X})}_{\text{upper bound}}$$

- Al intervalo $[L(\underline{X}), U(\underline{X})]$ le llamamos intervalo estimador

- Notar que tanto $L(\underline{X})$ como $U(\underline{X})$ son variables aleatorias, por lo que podría ocurrir que el intervalo no contenga a θ.

- Lo que buscamos es determinar $L(\underline{X})$ y $U(\underline{X})$ de tal forma que el intervalo que obtengamos capture a θ con cierta frecuencia, llamada nivel o coeficiente de confianza.



- Así, al observar una muestra $\underline{x} = (x_1, \dots, x_n)$, tenemos $t_1 = L(\underline{x})$ y $t_2 = U(\underline{x})$, entonces el intervalo $[t_1, t_2]$ conforma un intervalo para θ al $(1-\alpha)100\%$ de confianza ($\alpha \in (0,1)$ es la probabilidad de quedar fuera del intervalo)

- ¿Por qué "confianza" y no "proba"?

$$P(T_1 \le \theta \le T_2) = 1 - \alpha \quad \longleftarrow \text{aleatorio}$$

pero al observar la muestra

$$P(t_1 \le \theta \le t_2) = \begin{cases} 1 \\ 0 \end{cases} \quad \longleftarrow \text{observado}$$

El $(1-\alpha)100\%$ de los intervalos calculados a partir de T_1 y T_2 contendrán al verdadero valor de θ (para distintas muestras)

- Existen varios métodos para obtener intervalos de confianza (IC), pero nosotros nos enfocaremos en el método pivotal.

Método pivotal

Ejemplo. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

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

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

pivote

Calcular un intervalo para μ al 95% de confianza.

$$0.95 = P\left(a \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le b\right)$$

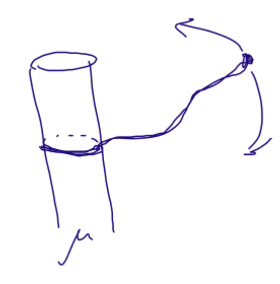
$$= P\left(a\sigma \le \sqrt{n}(\bar{X}-\mu) \le b\sigma\right)$$

$$= P\left(a\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le b\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(a\frac{\sigma}{\sqrt{n}} - \bar{X} \le -\mu \le b\frac{\sigma}{\sqrt{n}} - \bar{X}\right)$$

$$= P\left(\bar{X} - b\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - a\frac{\sigma}{\sqrt{n}}\right)$$

$$\mu \in \left[\bar{X} - b\frac{\sigma}{\sqrt{n}}, \bar{X} - a\frac{\sigma}{\sqrt{n}}\right]$$



Def. Sea $X_1, \dots, X_n$ m.a. con f.d.p. $f_X(x;\theta)$.
Sea $Q(X_1, \dots, X_n, \theta) = Q(\underset{\sim}{X}, \theta)$ una función de la muestra y del parámetro tal que su distribución <u>no</u> depende de θ. Decimos que $Q(\underset{\sim}{X}, \theta)$ es una cantidad pivotal.

→ Ejemplos. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

- $Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$ — no aparecen μ ni σ^2

  es función de $\underset{\sim}{X}$ y de μ y σ^2

  $\therefore Z = Q(\underset{\sim}{X}, \mu, \sigma^2)$ es una cantidad pivotal

- $T = \frac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}$

  $\therefore T = Q(\underset{\sim}{X}, \mu)$ es una c.p.

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

  $\therefore J = Q(\underset{\sim}{X}, \sigma^2)$ es c.p.

- $M = \frac{\bar{X}}{\mu} \sim N\left(1, \frac{\sigma^2}{\mu^2}\right)$ — aparecen μ y σ^2

  $\therefore M$ no es una c.p.

Pasos a seguir

1. Encontrar una cantidad pivotal $Q(\underset{\sim}{X}, \theta)$ para el parámetro a estimar)
2. Escoger a y b tales $a < b$ y
   $P(a \le Q(\underset{\sim}{X}, \theta) \le b) = 1-\alpha$
3. Despejar θ de forma que tengamos
   $P(L(\underset{\sim}{X}) \le \theta \le U(\underset{\sim}{X})) = 1-\alpha$
4. Definir el intervalo $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ con nivel de confianza $(1-\alpha)100\%$

<u>Observaciones:</u>

- Existe un número infinito de IC para un mismo parámetro (y mismo nivel de confianza)
- Variantes.
  - Intervalos bilaterales
    $P(a \le Q(\underset{\sim}{X}, \theta) \le b) = 1-\alpha$
  - Intervalos unilaterales
    $P(a \le Q(\underset{\sim}{X}, \theta)) = 1-\alpha$
    $P(Q(\underset{\sim}{X}, \theta) \le b) = 1-\alpha$

Ejemplo Sea $X_1, \dots, X_n \overset{iid}{\sim} U(0, \theta)$
Sea $Q(\underset{\sim}{X}, \theta) = \frac{1}{\theta} X_{(n)}$
Encontrar <u>un</u> IC para θ al 95%

(1) Q depende de la muestra (porque contiene a $X_{(n)} = \max\{X_1, \dots, X_n\}$ y del parámetro θ).
Falta ver que f_Q <u>no</u> depende de θ.

$$F_Q(x) = P(Q \le x) = P\left(\frac{X_{(n)}}{\theta} \le x\right)$$

$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^n$

$$= P(X_{(n)} \le \theta x) = F_{X_{(n)}}(\theta x)$$

$$= \left(\frac{\theta x}{\theta}\right)^n = x^n I_{(0,1)}(x)$$

$\to f_{X_{(n)}}(x) = n\left(\frac{x}{\theta}\right)^{n-1} I_{(0,\theta)}(x)$

no depende de θ

$Q = \frac{X_{(n)}}{\theta}$ $\qquad \frac{X_{(n)}}{\theta} \in \left(\frac{0}{\theta}, \frac{\theta}{\theta}\right) = (0,1)$

$$\therefore Q(\underset{\sim}{X}, \theta) = \frac{X_{(n)}}{\theta} \text{ es una c.p.}$$

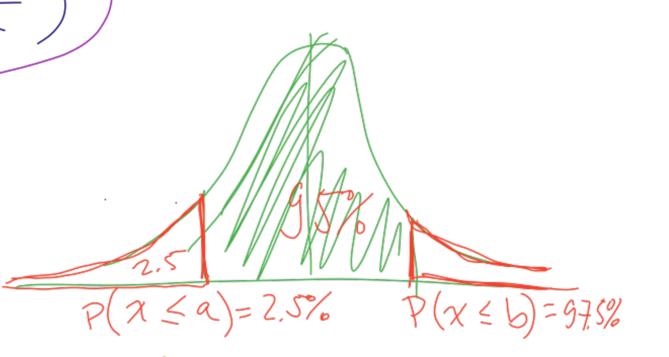
(2) y (3)

$$1-\alpha = 0.95 = P\left(a \le Q(\underset{\sim}{X}, \theta) \le b\right)$$

$$= P\left(a \le \frac{X_{(n)}}{\theta} \le b\right)$$

$$= P\left(\frac{X_{(n)}}{b} \le \theta \le \frac{X_{(n)}}{a}\right)$$

Para ver cuánto valen a y b:



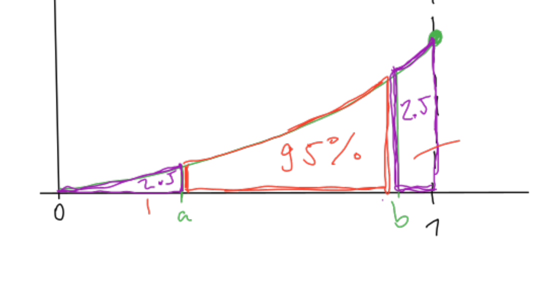
$$P\left(a \le \frac{X_{(n)}}{\theta} \le b\right)$$

$P\left(\frac{X_{(n)}}{\theta} \le a\right) = 0.025$
$\iff F_Q(a) = 0.025$
$\iff a^n = 0.025$
$\iff a = (0.025)^{\frac{1}{n}}$

$P\left(\frac{X_{(n)}}{\theta} \le b\right) = 0.975$
$\iff b = (0.975)^{\frac{1}{n}}$

(4) $$\left(\frac{X_{(n)}}{b}, \frac{X_{(n)}}{a}\right) = \left(\frac{X_{(n)}}{0.975^{\frac{1}{n}}}, \frac{X_{(n)}}{0.025^{\frac{1}{n}}}\right)$$

es un IC para θ al 95%



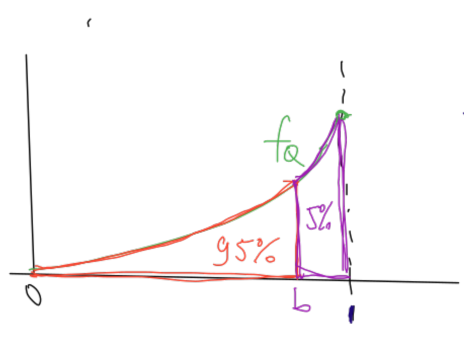
nx^{n-1}

Alternativa: intervalo unilateral

$$1-\alpha = 0.95 = P\left(Q(\underset{\sim}{X}, \theta) \le b\right)$$

$$= P\left(\frac{X_{(n)}}{\theta} \le b\right)$$

$$= P\left(\theta \ge \frac{X_{(n)}}{b}\right)$$



$$\Rightarrow F_Q(b) = 0.95 \iff b^n = 0.95 \iff b = 0.95^{\frac{1}{n}}$$

$$\therefore \left(\frac{X_{(n)}}{0.95^{\frac{1}{n}}}, 1\right) \text{ es un IC para } \theta \text{ al } 95\%$$