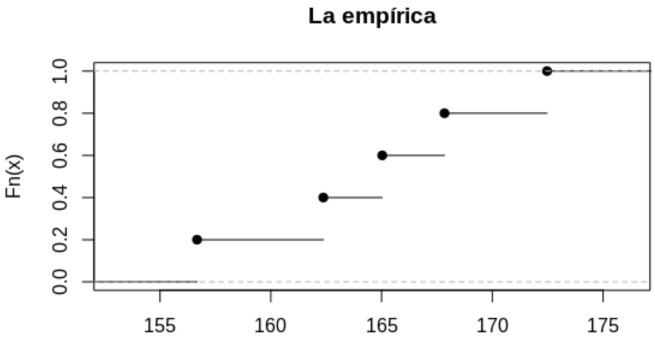
Probabilidades y Estadística (C) – 2019

La última – repasando algunas cositas

▸ música maestro

La empírica



valores	156.67	162.37	165.03	167.84	172.47
puntual	1/5	1/5	1/5	1/5	1/5

Distribución Empírica

- $X_1, \ldots, X_n$ iid, $X_i \sim F$. $X \sim F$.
- Acumulada F. $F(t) = \mathbb{P}(X \leq t)$.
- Estimación de la acumulada:

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{X_i \leq t}$$

- $\widehat{F}_n$ ES una función de distribución acumulada (de discreta).
- $\widehat{F}_n$ asigna peso $1/n$ a cada valor $X_1, \ldots, X_n$.

valores	X_1	$\ldots$	$\ldots$	$\ldots$	X_n
puntual	$1/n$	$1/n$	$1/n$	$1/n$	$1/n$

- Atenti a los repetidos!

Medidas de resumén - *Posición* - Mongo *muestral*

- Media
- Mediana
- Percentil
- Cuartiles:
- Media α- podada

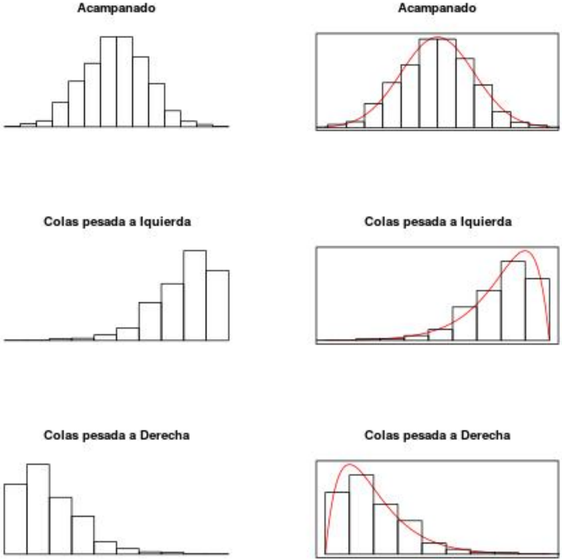
Medidas de resumén - *Dispersión*

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

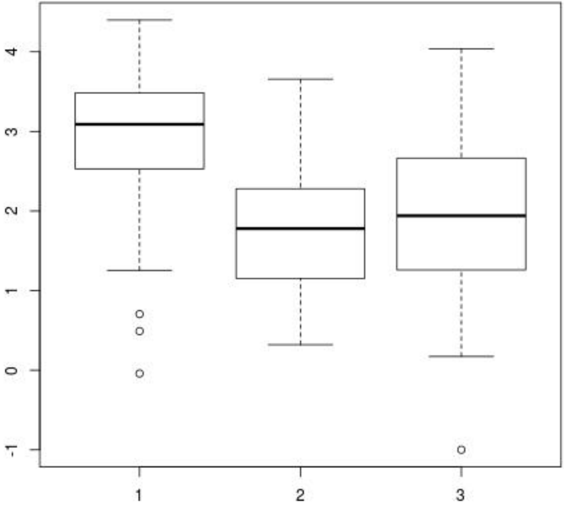
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- Distancia intercuartil. $Q_3 - Q_1$
- MAD: medianda$\{|x_i - \widetilde{x}|\}$

Histogramas- Media & Mediana



Boxplot - en R boxplot(datos)



Estadística

POBLACION $\leftrightarrow F$	MUESTRA $X_1, \ldots X_n$ i.i.d. $X_i \sim F$
Parámetro: Valor asociado de F	Estimador:estadístico para estimar θ
$\theta = \theta(F)$	$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$
θ: valor poblacional	$\widehat{\theta}_n$ NUEVA VARIABLE ALEATORIA

Muestra - Datos (Observaciones)

- Muestra (aleatoria simple):

$$X_1, \ldots, X_n \quad \text{Variables aleatorias iid.}$$

- Datos - Observaciones - Valores observados

$$x_1, \ldots, x_n \quad \text{Números.}$$

Parámetro, estimador, estimación – Estimación Puntual

Un parámetro es un número FIJO (somos frecuentistias) que describe algun aspecto de *"la población"*: F.

- Parámetro: $\theta = \theta(F)$
- Muestra: $X_1, \ldots, X_n$ iid, $X_i, \sim F$.
- Estimador:

$$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$$

- Estimación:

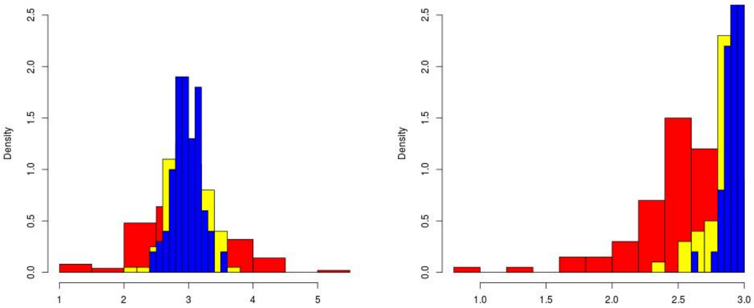
$$\widehat{\theta}_{n,\mathsf{obs}} = \widehat{\theta}_n(x_1, \ldots, x_n)$$

Sampling distribution

- $\widehat{\theta}_n$ es una variable aleatoria.
- $\widehat{\theta}_n$ tiene distribución (siempre)
- $\widehat{\theta}_n$ tiene (en general) esperanza: $\mathbb{E}(\widehat{\theta}_n)$
- $\widehat{\theta}_n$ tiene (en general) varianza: $\mathbb{V}(\widehat{\theta}_n)$

Histogramas de $\widehat{\theta}_n = 2\overline{X}_n$ y de $\widetilde{\theta}_n = \max\{X_1, \dots, X_n\}$

$$X_i \sim \mathcal{U}(0, \theta)$$



Métodos de estimación

- Momentos

- Máxima verosimilitud – Propiedad de Invarianza
  $\widehat{g}(\theta)_{MV} = g(\widehat{\theta}_{MV})$

Propiedades

- Consistencia

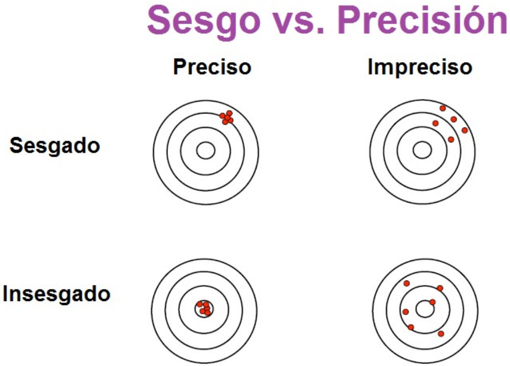
$$\widehat{\theta}_n(X_1,\ldots,X_n) \to^{\mathcal{P}} \theta(F) \text{ , cuando } X_i \sim F, \ \forall F \in \mathcal{M}$$

$$\text{abreviado:} \quad \widehat{\theta}_n \to \theta \ , \quad \forall\theta$$

- Error cuadratico medio: ECM$=\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\}$
- Lema: Si $\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\} \to 0$, entonces $\widehat{\theta}_n \to^{\mathcal{P}} \theta$
- Sesgo: $\mathbb{E}(\widehat{\theta}_n) - \theta$.
- Estimador insesgado: Sesgo$=$0: $\mathbb{E}(\widehat{\theta}_n) - \theta$
- Lema: $\mathbb{E}\left\{(\widehat{\theta}_n - \theta)^2\right\} = \mathbb{V}(\widehat{\theta}_n) + \left\{\mathbb{E}(\widehat{\theta}_n) - \theta\right\}^2$
- Si $\mathbb{V}(\widehat{\theta}_n) \to 0$ y $\mathbb{E}(\widehat{\theta}_n) \to \theta$, entonces

$$\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\} \to 0$$

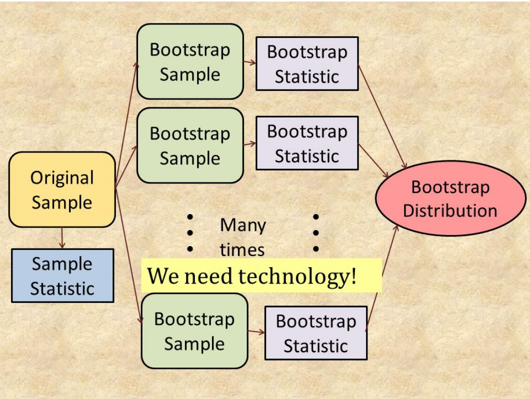
Exactitud (sesgo) - Precisión (varianza)



Sobre la distribución del estimador

- Distribución exacta (distribución de la suma, etc (proba))
- Distribución aproximada (TCL)
- Aproximación Bootstrap. Necesitas una compu.

Esquema Bootstrap



Error de Estimación

"Toda estimación relevante conlleva un error" – Walter Sosa

Definición: llamamos error de una estimación a la estimación del desvío (exacto o aproximado) del estimador con el cual estimamos.

- Estimador: $\widehat{\theta}_n$, estimación: $\widehat{\theta}_{n,\mathrm{obs}}$
- se $:= \sqrt{V(\widehat{\theta}_n)}$ o se $:\approx \sqrt{V(\widehat{\theta}_n)}$. Sea $\widehat{\mathrm{se}}$ un estimador de se. Error de estimación: $\widehat{\mathrm{se}}_{\mathrm{obs}}$

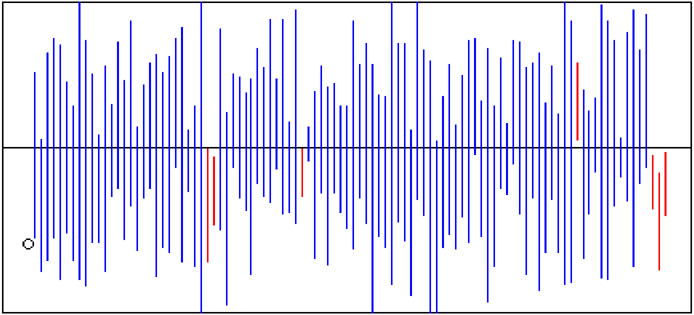
Ejemplo: Si $\widehat{\mu}_n = \overline{X}_n$, tenemos que

$$\widehat{\mathrm{se}} := \sqrt{\frac{\widehat{V(X)}}{n}}$$

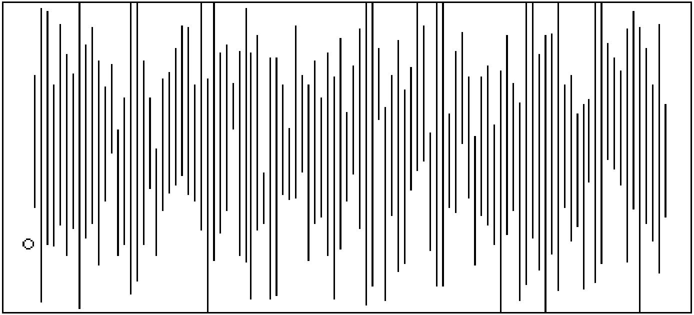
Intervalos de confianza – Estimación por intervalos

- Intervalo de confianza: rango de valores plausibles para el parámetro de interés.
- Los intervalos se contruyen usando un método que que tiene una probabilidad predeterminada de capturar el verdadero parámetro.

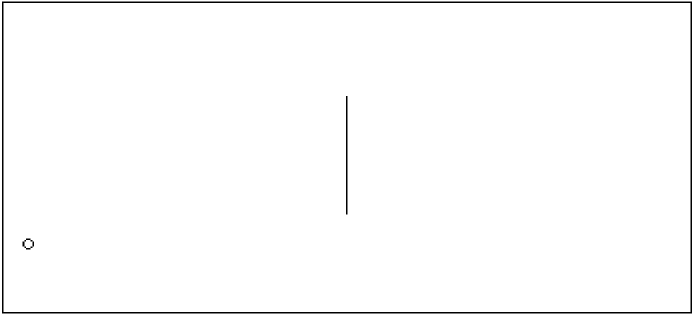
Muchos intervalos y la verdad



Muchos intervalos



Mi intervalo y yo, buena suerte! (confianza)



Intervalos – Jerga

- Modelo y parámetro de interés
- Nivel (de cubrimiento) $(1-\alpha)$
- El Pivot y su distribución – Proba. Normales, χ^2, t- Student, etc...
- Mundo Asintótico (TCL $+$ Slutzky)

Test de hipótesis

Determinar si los datos obtenidos resulta suficientemente convincentes para sacar alguna conclusión algo sobre el parámetro de interés.

Hipótesis nula y alternativa

- Hipótesis alternativa H_1: Escenario para el cual buscamos evidencia significativa.

- La hipótesis alternativa (H_1) se establece mediante la observación de evidencia (en los datos) que contradice la hipótesis nula y apoya la hipótesis alternativa.

- Los datos son *raros* bajo la hipótesis nula (H_0), y ADEMAS sugieren que sea rechazada en favor de la hipótesis alternativa.

Significatividad Estadísica

Cuando los resultados son poco probables suponiendo que la hipótesis nula H_0 es cierta, indicando además evidencias en favor de H_1, decimos que los resultados son estadísticamente significativos.

Si nuestra muestra es estadísticamente significativa, tenemos evidencia convincente contra H_0 y en favor de H_1.

Hay evicencia significativa? $\equiv$ Puede rechazar H_0 en favor de H_1?

Test de Hipótesis – Jerga

- Modelo y parámetro de interés
- H_0, H_1 y $\mathcal{R}$, la región de rechazo de H_0 en favor de H_1.
- Función de Potencia
- Error Tipo I y Error Tipo II y sus probabilidades.
- Nivel de significación (α)
- El estadístico del Test y su distribución (cuando?)
- p- valor
- Mundo Asintórtico

Relación entre Test e Intervalos

Si el valor del parámetro bajo la hipótesis nula ($H_0 : \theta = \theta_0$) queda fuera del intervalo de confianza observado, rechazamos H_0 en favor de $H_1 : \theta \neq \theta_0$.

Por otro lado, si el valor del parámetro bajo la hipótesis nulas (θ_0) cae dentro del intervalo de confianza, entendemos que θ_0 es un valor plausible y por lo tanto no tenemos evidencia para rechazar la hipótesis nula.

Gracias por venir!