Estimación Puntual

Estadística (M)

Muestra - Datos (Observaciones)

- Datos - Observaciones $x_1, \ldots, x_n$: Números.

Datos-Observaciones: son los resultados obtenidos al realizar el ''experimento''

Muestra - Datos (Observaciones)

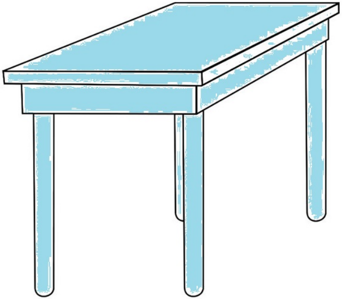
- Datos - Observaciones $x_1, \ldots, x_n$: Números.

  Datos-Observaciones: son los resultados obtenidos al realizar el "experimento"

- Muestra $X_1, \ldots, X_n$: Variables aleatorias.

  Datos-Observaciones: son realizaciones de las variables aleatorias

¿Cuánto mide la mesa?



La mesa es medida con un piolín por....

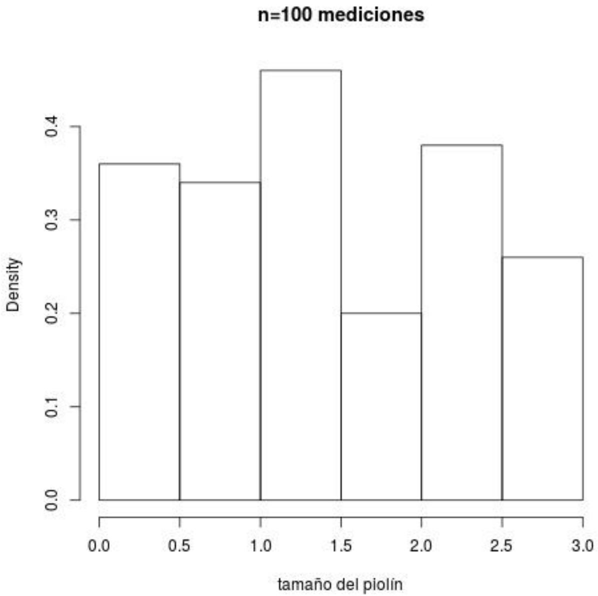
¿Cuánto mide la mesa?

Estas son las $n = 7$ primeras observaciones realizadas por Juani:

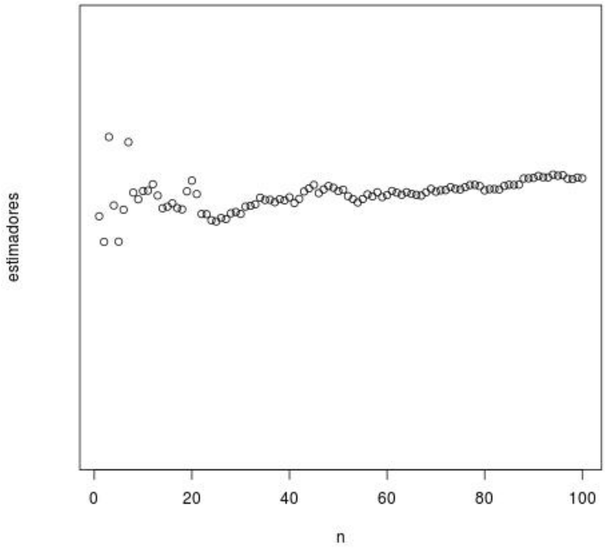
$$1.17, \quad 1.36, \quad 0.15, \quad 2.52, \quad 0.21, \quad 1.78, \quad 2.67$$

Propongamos un modelo para estas mediciones.

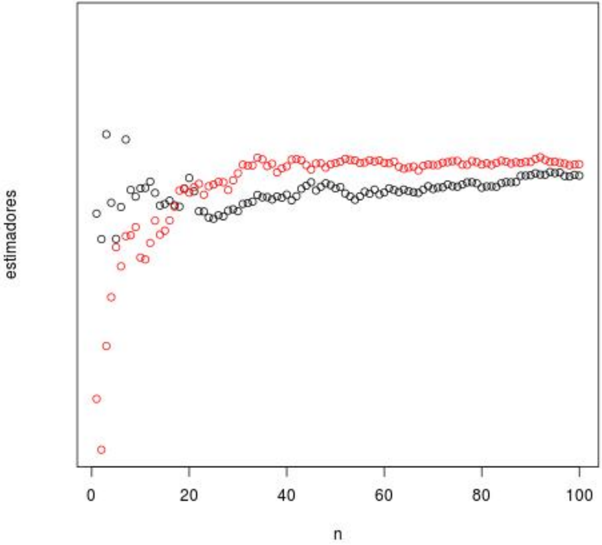
Histograma


n=100 mediciones
Density
0.0
0.1
0.2
0.3
0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
tamaño del piolín

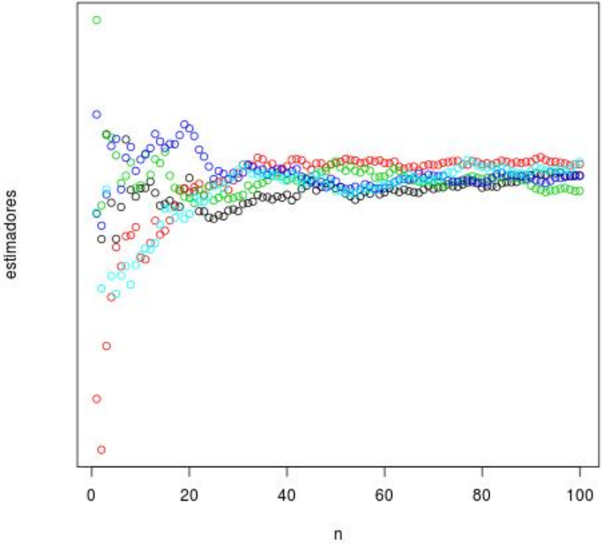
Juani cada vez con más datos. $\widehat{\theta}_n = 2\overline{X}_n$



Juani y Chita, cada vez con más datos. $\widehat{\theta}_n = 2\overline{X}_n$



Varios, cada vez con más datos. $\widehat{\theta}_n = 2\overline{X}_n$

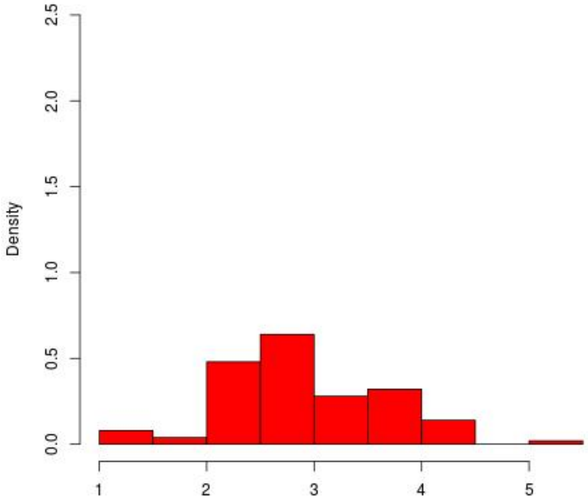


Cada uno mide lo suyo. $\widehat{\theta}_n = 2\overline{X}_n$

	Nombre	n=5	n=30	n=50
1	Juani	1.08	3.2	2.96
2	Chita	2.87	2.95	2.88
3	Perico	3.47	3.2	3.18
4	Bonita	3.88	3.23	3.18
5	Milo	3.79	2.93	2.81
6	Poli	3.01	2.9	2.59
7	Peperina	3.55	3.03	3.01
8	Tuco	2.09	2.79	3.1
9	Babá	4.14	3.41	3.01
10	Pepa	2.65	3.29	3.11
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

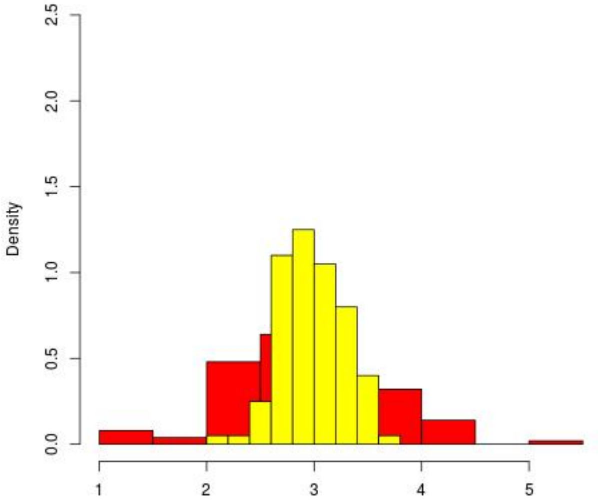
Histogramas de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$

Distribución (muestral) empírica de $\widehat{\theta}_n$



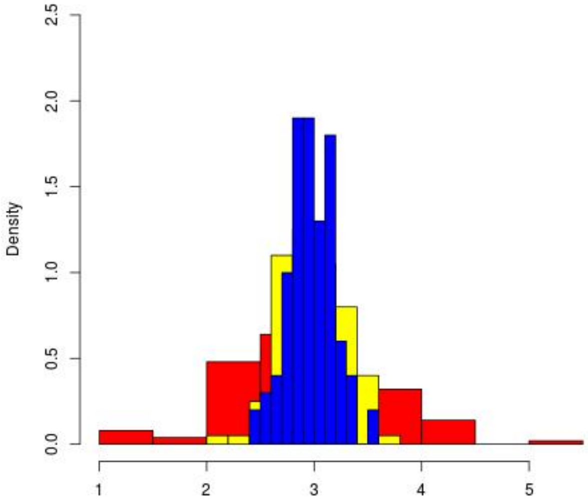
Histogramas de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$ y 30

Distribución (muestral) empírica de $\widehat{\theta}_n$



Histogramas de $\widehat{\theta}_n = 2\overline{X}_n : n = 5, 30$ y 50

Distribución (muestral) empírica de $\widehat{\theta}_n$



Estimación

Estimación puntual se refiere a proporcionar una única "mejor conjetura" de alguna cantidad de interés.

All of statistics. Wasserman

Estimación

Estimación puntual se refiere a proporcionar una única "mejor conjetura" de alguna cantidad de interés
All of statistics. Wasserman

- Alguna cantidad de interés : largo de la mesa (θ)
- "mejor conjetura": Estimador: Función de la muestra

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

Estimación

Estimación puntual se refiere a proporcionar una única "mejor conjetura" de alguna cantidad de interés
All of statistics. Wasserman

- Alguna cantidad de interés : largo de la mesa (θ)
- "mejor conjetura": Estimador: Función de la muestra

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

- Estimación: Valor del estimador en un conjunto de datos:

$$\widehat{\theta}_n(x_1, \ldots, x_n)$$

Notemos que el estimador ...

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

- $\widehat{\theta}_n$ es una variable aleatoria.
- $\widehat{\theta}_n$ tiene distribución (siempre).

  Distribución muestral o empírica de $\widehat{\theta}_n$: $f_{\widehat{\theta}_n}$

- $\widehat{\theta}_n$ tiene (en general) esperanza: $\mathbb{E}(\widehat{\theta}_n)$
- $\widehat{\theta}_n$ tiene (en general) varianza: $\mathbb{V}(\widehat{\theta}_n)$
- $\widehat{\theta}_n$ tiene (en general) desvío standard.

$$\mathsf{se} = \mathsf{se}(\widehat{\theta}_n) = \sqrt{\mathbb{V}(\widehat{\theta}_n)} \quad \textit{Error Standard de } \widehat{\theta}_n.$$

Muestra

- Muestra aleatoria: $X_1, \cdots, X_n$ variables aleatorias i.i.d. (independientes e idénticamente distribuidas)
- Datos u observaciones: $\mathbf{x} = x_1, \ldots, x_n$ constituyen una realización de la muestra aleatoria.

Inferencia Estadística

- Datos: $(X_i)_{i\geq 1}$ i.i.d. $X_i \sim F$, $F \in \mathcal{F}$ familia de distribuciones posibles para nuestro problema

- Objetivo: inferir *algo relacionado* con el mecanismo que genera los datos:
  - $\mathbb{E}_F[X]$
  - $\mathbb{V}_F(X)$
  - $\mathbb{P}_F(X > 1)$
  - F.

- Propongamos un estimador para cada uno de los *algo* planteados.

- Bernoulli: $X \sim \mathcal{B}(1, \theta)$, $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$. $p \in \Theta = [0, 1] \subset \mathbb{R}$.

- Poisson: $X \sim P(\lambda)$, $p(x, \lambda) = e^{-\lambda} \lambda^x / x!$, $\lambda \in \Theta = \mathbb{R}_{>0}$.

Modelos Paramétricos

Ejemplos discretos

- Bernoulli: $X \sim \mathcal{B}(1, \theta)$, $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$.
  $p \in \Theta = [0, 1] \subset \mathbb{R}$.

- Poisson: $X \sim P(\lambda)$, $p(x, \lambda) = e^{-\lambda} \lambda^x / x!$, $\lambda \in \Theta = \mathbb{R}_{>0}$.

$$\mathcal{M} = \{p(\cdot, \theta)\ , \theta \in \Theta\}\ ,$$

Modelos Paramétricos

Ejemplos continuos

- Uniforme: $X \sim \mathcal{U}(0,\theta)$, $f(x,\theta) = \frac{1}{\theta}\,\mathcal{I}_{[0,\theta]}(x)$, $\theta \in \Theta = \mathbb{R}_{>0}$.
- Normal: $X \sim \mathcal{N}(\mu,\sigma^2)$, $f(x,\theta) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$, $\theta = (\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$
- Exponencial: $X \sim \mathcal{E}(\lambda)$, $f(x,\lambda) = \lambda e^{-\lambda x}\mathcal{I}_{[0,\infty)}(x)$, $\lambda \in \Theta = \mathbb{R}_{>0}$.

Modelos Paramétricos

Ejemplos continuos

- Uniforme: $X \sim \mathcal{U}(0,\theta)$, $f(x,\theta) = \frac{1}{\theta}\, \mathcal{I}_{[0,\theta]}(x)$, $\theta \in \Theta = \mathbb{R}_{>0}$.
- Normal: $X \sim \mathcal{N}(\mu,\sigma^2)$, $f(x,\theta) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$, $\theta = (\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$
- Exponencial: $X \sim \mathcal{E}(\lambda)$, $f(x,\lambda) = \lambda e^{-\lambda x}\mathcal{I}_{[0,\infty)}(x)$, $\lambda \in \Theta = \mathbb{R}_{>0}$.

$$\mathcal{M} = \{f(\cdot,\theta)\ , \theta \in \Theta\}\ .$$

Modelos Paramétricos: $\mathcal{M}$ Ejemplos

Asumimos que la función de distribución F que genera los datos pertenece a una familia conocida, salvo por el valor de cierto parámetro:

$$\mathcal{M} = \{F(\cdot, \theta) \,, \theta \in \Theta\} \,,$$

siendo $\Theta \subset \mathbb{R}^k$, para algún k.

- Caso discreto: $\mathcal{M} = \{p(\cdot, \theta) \,, \theta \in \Theta\} \,,$
- Caso continuo: $\mathcal{M} = \{f(\cdot, \theta) \,, \theta \in \Theta\} \,.$

Si $F \in \mathcal{M}$, para conocerla necesitamos identificar su θ.

Estimadores de Máxima Verosimilitud

Ejemplo

- Monedas, cara$=1$, ceca$=0$.
- Moneda bolsillo derecho: equilibrada
- Moneda bolsillo izquierdo: probabilidad de cara es 0.8.
- Objetivo: identificar la moneda a partir de una muestra. En $n = 100$ lanzamientos se observa la muestra

$$\mathbf{x} = \underbrace{1,\ldots,1}_{\mathbf{12veces}}\underbrace{0,\ldots,0}_{\mathbf{5veces}}\underbrace{1,\ldots,1}_{\mathbf{23veces}}\underbrace{0,\ldots,0}_{\mathbf{8veces}}\underbrace{1,\ldots,1}_{\mathbf{15veces}}\underbrace{0,\ldots,0}_{\mathbf{3veces}}$$

$$\underbrace{1,\ldots,1}_{11veces}\underbrace{0,\ldots,0}_{4veces}\underbrace{1,\ldots,1}_{13veces}\underbrace{0,\ldots,0}_{6veces}$$

- ¿Cuál de las dos monedas diría que está utilizando?

Función de Verosimilitud

- $L(0.8; \mathbf{x})$ = probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.8$.

- $L(0.5; \mathbf{x})$ = probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.5$ (equilibrada).

Función de Verosimilitud

- $L(0.8; \mathbf{x})$ = probabilidad de observar la muestra $\mathbf{x}$ con moneda de $p = 0.8$.

$$
\begin{aligned}
L(0.8; \mathbf{x}) \ = \ & \underbrace{0.8 \ldots 0.8}_{12 veces}\underbrace{0.2 \ldots 0.2}_{5 veces}\underbrace{0.8 \ldots 0.8}_{23 veces}\underbrace{0.2 \ldots 0.2}_{8 veces}\underbrace{0.8 \ldots 0.8}_{15 veces}\underbrace{0.22 \ldots 0.2}_{3 veces} \\
& \underbrace{0.8 \ldots 0.8}_{11 veces}\underbrace{0.2 \ldots 0.2}_{4 veces}\underbrace{0.8 \ldots 0.8}_{13 veces}\underbrace{0.2 \ldots 0.2}_{6 veces} = \\
& = (0.8)^{74}(0.2)^{26} = 4.523128\ 10^{-26}
\end{aligned}
$$

siendo 74 el número de caras observadas en las $n = 100$ repeticiones.

¿Y ahora?

- $\mathbf{x}$ con 74 caras, 26 cecas.
- $L(0.8; \mathbf{x}) = 0.8^{74} 0.2^{26} = 4.523128\ 10^{-26}$
- $L(0.5; \mathbf{x}) \ = \ (1/2)^{74}(1/2)^{26} = 7.888609\ 10^{-31}$
- ¿cuál de las dos monedas diríamos que se está utilizando?

Propuesta de Máxima Verosimilitud

La propuesta de máxima verosimilitud consiste en pensar que la moneda que estamos utilizando es aquella para la cual los valores observados resultan más probables. Es decir, elegimos la moneda que maximiza la probabilidad de los valores observados. Siendo que

$$L(0.8; \mathbf{x}) \; > \; L(0.5; \mathbf{x})$$

<u>concluimos que se está utilizando la moneda no equilibrada</u>.

Verosimilitud

En términos generales

- Antes de realizar el *experimento* el **resultado** es **desconocido**.
- Las probabilidades nos permiten predecir un resultado **desconocido** a partir de parámetros **conocidos**: por ejemplo

$$P(resultado|\theta) \overset{\text{por ej. caso Binomial}}{=} P(x|n,p) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

Ahora se invierte el paradigma:

- Al realizar el experimento el **resultado** se hace **conocido**: **dato**.
- Nos interesa cuán **verosímil** es que un determinado parámetro haya generado el dato.

Entonces, veamos nuestro ejemplo un poco más en general

- $X_i \sim \mathcal{B}(1, \theta)$, $\theta \in [0, 1]$.
- 74 caras en $n = 100$ repeticiones.
- ¿θ?
- vamos a elegir aquel valor del parámetro para el cuál los valores observados tienen mayor probabilidad de ocurrir.