

Práctica 1

Introducción a la Estadística - Estadística descriptiva - Sucesiones

Aclaración: Los ejercicios marcados con (*) son optativos.

Ejercicio 1. Se quiere estimar la proporción de usuarios de determinado navegador de internet que personalizan su home-page para recibir noticias de temas y sitios que les resulten de interés. Para ello, se realiza un estudio en el que se recolectan datos de $n = 1675$ usuarios de dicho navegador. En éste, se revela que 496 de ellos personalizaron su home-page.

- a) ¿Cuál es la población y cuál es el parámetro de interés?
- b) Utilizando la información de la muestra calcule su mejor estimación del valor del parámetro.
- c) ¿Que habría que hacer para calcular el valor del parámetro exactamente?

Ejercicio 2. Las historias clínicas del último año de un hospital muestran que el colesterol LDL promedio de todos sus pacientes es de 120 mg/dL. Se selecciona una muestra de 75 pacientes del hospital en la que se observa un colesterol LDL medio de 118,5 mg/dL. Identifique cuál de los datos presentados (120 ó 118,5) es el parámetro y cuál es su estimación.

Ejercicio 3. El pueblo Costa Esperanza tiene 100 habitantes de los cuales 15 trabajan 4 hs diarias, 30 trabajan 6 hs diarias, 20 trabajan 7 hs diarias y 35 trabajan 8 hs diarias.

- a) ¿Cuál es la proporción (poblacional) de habitantes que trabajan 4 hs diarias?
- b) ¿Cuál es el promedio (poblacional) de la cantidad de horas de trabajo diario de los habitantes de Costa Esperanza?
- c) ¿Cuál es la probabilidad de que un habitante de Costa Esperanza (elegido al azar) trabaje 4 hs diarias? ¿Qué relación tiene este valor con el calculado en el ítem II.?
- d) Se define la v.a. X = cantidad de horas de trabajo diario de un individuo elegido al azar de Costa Esperanza. Calcule $\mathbb{E}(X)$. ¿Qué relación tiene este valor con el calculado en el ítem b)?

Ejercicio 4. Cierta población tiene N habitantes de los cuales N_1 trabajan t_1 hs diarias, N_2 trabajan t_2 hs diarias, N_3 trabajan t_3 hs diarias y N_4 trabajan t_4 hs diarias. Sea

X = cantidad de horas de trabajo diario de un habitante elegido al azar de dicho pueblo.

- a) Pruebe que $\mathbb{P}(X = t_1)$ es igual a la proporción de habitantes de dicho pueblo que trabajan t_1 horas diarias. Es decir, que hay una equivalencia entre *probabilidad* y *proporción poblacional*.
- b) Pruebe que $\mathbb{E}(X)$ es igual a la cantidad de horas promedio de trabajo diario de los habitantes de dicho pueblo. Es decir, que hay una equivalencia entre *esperanza* y *media (o promedio) poblacional*.

Ejercicio 5. Considere la función $f(x) = x(1 - x)$ con dominio en el intervalo $(0, 1)$. Grafique dicha función y halle su máximo. ¿Dónde lo alcanza?

(Nota: este ejercicio proporciona una cota que usaremos mucho a lo largo de la materia.)

Ejercicio 6. Considerar el conjunto de datos `airquality`, de la librería `datasets` de R. Ejecute el comando `data(airquality)` para cargar el conjunto en el panel del environment, ábralo para visualizarlo y pida ayuda con el comando `?airquality` para obtener información sobre el significado de sus variables.

- a) ¿Cuántas observaciones y cuántas variables tiene el conjunto de datos? (Hint: Explore los comandos `dim`, `nrow` y `ncol`.)
- b) ¿Cuáles son los nombres de las variables? (Hint: explore el comando `names`.)
- c) ¿De qué tipo es cada variable? Utilice el comando `str` (aplicado al conjunto de datos) y el comando `class` aplicado a cada variable.
- d) ¿Qué variables tienen datos faltantes y cuántos son? (Hint: explore el comando `summary` aplicado al conjunto de datos. Recuerde que los datos faltantes en R se codifican con NA).
- e) Realizar un histograma de densidad de las variables `Wind` y `Temp`.

- I. ¿Diría que cada distribución es simétrica? Si hay asimetrías, ¿hacia qué lado son?

En base al histograma de `Temp`, calcule aproximadamente...

- II. qué proporción de datos tienen valor de `Temp` entre 65 y 70 .
- III. cuántos datos tienen valor de `Temp` entre 65 y 70.

- f) Realizar un boxplot para las variables `Wind` y `Temp`. ¿Identifica outliers? En caso afirmativo, ¿a qué valores corresponden? Explore el comando `sort`.

- g) A partir del boxplot de **Wind** calcule aproximadamente las 5 medidas resumen. Luego, calcúlelas con **R** y compare.
- h) Calcule media y mediana de la variable **Temp**, ¿cuál es mayor? ¿a qué cree que se debe?
- i) Calcule dos medidas de dispersión de la variable **Wind**

Ejercicio 7. (*) Sea x_1, \dots, x_n un conjunto de datos.

- a) Sea $f(a) = \sum_{i=1}^n (x_i - a)^2$. Probar que $f(a)$ alcanza su mínimo en $a = \bar{x}$.
- b) Probar que $\sum_{i=1}^n (x_i - \bar{x}) = 0$
- c) Probar que $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$.

Ejercicio 8. Para cada una de las siguientes sucesiones $(a_n)_n$ calcula los valores de los cuatro primeros términos (a_1, a_2, a_3, a_4) y calcula $\lim_{n \rightarrow +\infty} a_n$, si existe.

- a) $a_n = \frac{1}{n^2}$
- b) $a_n = \frac{5n}{3n^2 + 1}$
- c) $a_n = \frac{-3n^5 + n + 2}{n^5 + 2n^2}$
- d) $a_n = \frac{n - n^2 + 3}{2n - 1}$
- e) $a_n = \frac{4n + 1}{\sqrt{16n^2 + n}}$

Respuestas numéricas

Ejercicio 1

b) Estimación: 0,2961

Ejercicio 3

a) 0,15

b) 6,6

c) 0,15

d) 6,6

Ejercicio 5

El máximo es $\frac{1}{4}$ y lo alcanza en $x = \frac{1}{2}$.

Ejercicio 8

a) 0

b) 0

c) -3

d) $-\infty$

e) 1