

# Guía de Ejercicios de R

## Manejo de Datos, Gráficos y Cálculo de Probabilidades

EYP1113 - Probabilidad y Estadística

01/2025

### Instrucciones Generales

Resuelva los siguientes ejercicios utilizando únicamente R base. Se recomienda documentar cada paso con comentarios en el código.

### Parte I: Manejo de Bases de Datos y gráficos

1. Utiliza la base `mtcars` de R, es una base de datos con formato `data.frame` que fueron extraídos de la revista estadounidense Motor Trend de 1974. Incluyen información del consumo de combustible y 10 aspectos del diseño y rendimiento de automóviles de 32 vehículos (modelos 1973-74). Para acceder a la base de datos escriba el siguiente código `data(mtcars)` en R. Estos datos tienen 32 observaciones y 11 variables, las cuales son:

`mpg` Millas por galón (US),  
`cyl` Número de cilindros,  
`disp` Cilindrada (pulgadas cúbicas),  
`hp` Caballos de fuerza brutos,  
`drat` Relación del eje trasero,  
`wt` Peso (en 1000 libras),  
`qsec` Tiempo en recorrer 1/4 de milla,  
`vs` Motor (0 = en V, 1 = recto),  
`am` Transmisión (0 = automática, 1 = manual),  
`gear` Número de marchas adelante,  
`carb` Número de carburadores.

Primero seleccionar las columnas `mpg`, `cyl`, `hp` y `am`, y guardar este nuevo conjunto de datos un objeto llamado `mtcars1`.

- a) Con la base de datos `mtcars1`, obtener los estadísticos mínimo, máximo, primer cuartil, segundo cuartil o mediana, tercer cuartil y la media de la variable `mpg` que se refiere a las millas por galón. ¿Qué se puede decir sobre esta variable con estos estadísticos? Interprete los resultados.

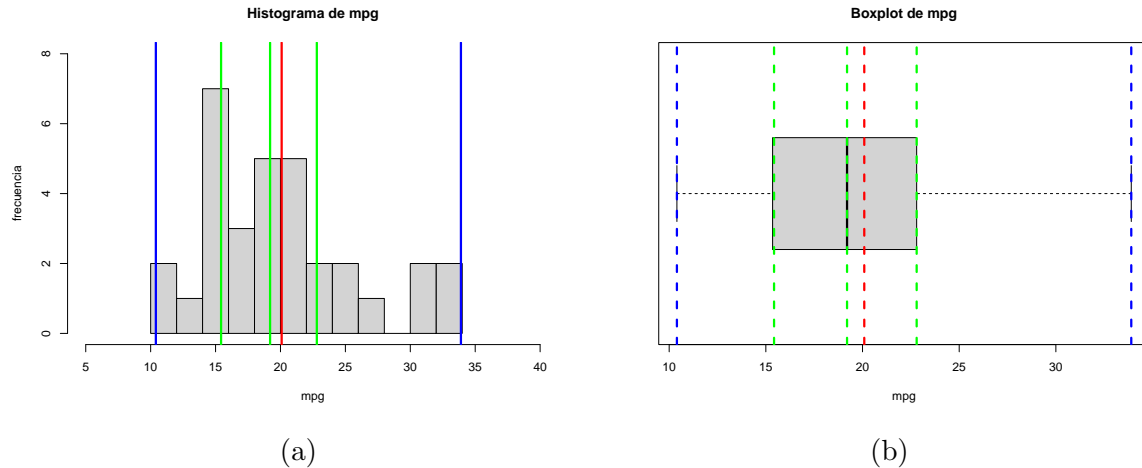


Figura 1: (a) Histograma y (b) boxplot de `mpg` millas por galon de los autos de la base de datos `mtcars`.

- b) Con la información del ítem anterior graficar el histograma de `mpg` insertando líneas horizontales de todos los estadísticos encontrados, como se ve en la figura 1a. También graficar el boxplot con la misma información y trazar las líneas de los estadísticos, como se ve en la figura 1b. La línea roja representa a la media, las líneas verdes a los cuartiles 1, 2 y 3, y las líneas azules al mínimo y máximo de `mpg`. Interprete la información de estos gráficos.
  - c) Calcular el coeficiente de asimetría y curtosis de `mpg`. Interprete estos resultados. ¿Estos coeficientes tienen relación con la interpretación de los gráficos del ítem anterior?
  - d) Con la base de datos `mtcars1`, se pide filtrar los autos con más de 20 millas por galón (`mpg`), es decir `mpg > 20` y calcular el promedio de `hp` agrupado por número de cilindros.
  - e) Con la base de datos `mtcars1`, se pide filtrar los autos con menos de 25 millas por galón (`mpg`) y los autos con menos de 180 caballos de fuerza brutos (`hp`), es decir `mpg < 25 & hp < 180`. Después calcular el número de autos, el promedio, mediana y desviación estándar de la cilindrada en pulgadas cúbicas `disp`, agrupado por número de cilindros (`cyl`) y por la transmisión (`am`).
2. El conjunto de datos de `iris` que está disponible en R base. Es una base de datos que contiene las medidas en centímetros de las variables largo y ancho del sépal (`Sepal.Length` y `Sepal.Width`) y largo y ancho del pétalo (`Petal.Length` y `Petal.Width`), respectivamente, de 50 flores, cada una de 3 especies de iris. Las especies son Iris setosa, versicolor y virginica.
    - a) Con la base de datos `iris`, obtener los estadísticos mínimo, máximo, primer cuartil, segundo cuartil o mediana, tercer cuartil, la media, el rango intercuartil, coeficiente de asimetría y curtosis de la variable `Sepal.Length` que se refiere al largo de sépal. ¿Qué se puede decir sobre esta variable con estos estadísticos? Interprete los resultados.
    - b) Filtrar las observaciones con el tipo de especie `setosa` y `versicolor` y guardar este nuevo conjunto de datos un objeto llamado `iris1`. Con esta nueva base

de datos, obtener los estadísticos mínimo, máximo, primer cuartil, segundo cuartil o mediana, tercer cuartil, la media, el rango intercuartil, coeficiente de asimetría y curtosis de la variable `Sepal.Length`. Interprete los resultados. ¿Existe diferencia con los resultados del ítem anterior?

- c) ¿Si comparamos los resultados del ítem a) y b) se puede decir que la especie *virginica* influye o aporta información a la variable largo de sépalo?
- d) Con la información de la base de datos `iris1` graficar el histograma de `Sepal.Length` pero diferenciar por el tipo de especie, como se ve en la figura 2. Interprete el gráfico.
- e) Con la información de la base de datos `iris1`, obtener los estadísticos mínimo, máximo, primer cuartil, segundo cuartil o mediana, tercer cuartil, la media, el rango intercuartil, coeficiente de asimetría y curtosis de la variable `Sepal.Length`, pero agrupado por tipo de especie (*setosa* y *versicolor*). Interprete los resultados. ¿Estos estadísticos complementan la interpretación del gráfico del ítem anterior?
- f) Ahora se pide trabajar con la base de datos `iris` completa y graficar un diagrama de dispersión entre el largo `Sepal.Length` y ancho `Sepal.Width` del sépalo de todas las especies, pero diferenciar la especie con diferente color, como se muestra en la figura 3a. Además se pide reportar la covarianza y el coeficiente de correlación entre ambas variables (`Sepal.Length` y `Sepal.Width`). Interprete.
- g) Con la base de datos `iris` graficar un diagrama de dispersión entre el largo `Sepal.Length` y ancho `Sepal.Width` del sépalo de dos especies (*versicolor* y *virginica*) y diferenciar la especie con diferente color, como se muestra en la figura 3b. Además se pide reportar la covarianza y el coeficiente de correlación entre ambas variables (`Sepal.Length` y `Sepal.Width`) de solo estas dos especies. Interprete.
- h) Se pide explicar porque se debe el cambio de covarianza cuando se elimina una especie en este caso la *setosa*. Los gráficos mostrados en el ítem f) y g) ¿ayudan a algún tipo de interpretación?

## Parte II: Cálculo de Probabilidades y gráficos

3. De la base de datos `iris` graficar un histograma de la variable ancho de sépalo `Sepal.Width` y añadir una curva al mismo grafico con la densidad con la función de R `curve()`, suponiendo que esta proviene de una distribución Normal  $\mathcal{N}(\mu, \sigma^2)$ . Reemplazar los parámetros de la distribución  $(\mu, \sigma^2)$  por el promedio muestral  $\bar{X}$  y su varianza muestral  $S_X^2$ , respectivamente. Recordar que en R el promedio muestral se obtiene con la función `mean()`, la varianza muestral con la función `var()` y la desviación estándar muestral con la función `sd()`. Se desea obtener el gráfico como la figura 4.
4. A partir de la distribución del ítem anterior, vale decir de la variable  $X = \text{Sepal.Width}$ , asumiendo una Distribución Normal  $X \sim \mathcal{N}(\mu, \sigma^2)$  y reemplazando los parámetros de la distribución  $(\mu, \sigma^2)$  por el promedio muestral  $\bar{X}$  y su varianza muestral  $S_X^2$ , respectivamente. Calcular:

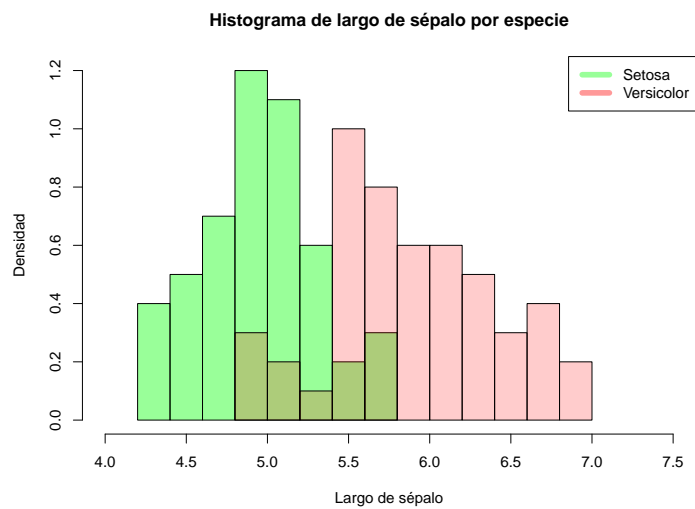


Figura 2: Histogramas del largo de sépalo por tipo de especie.

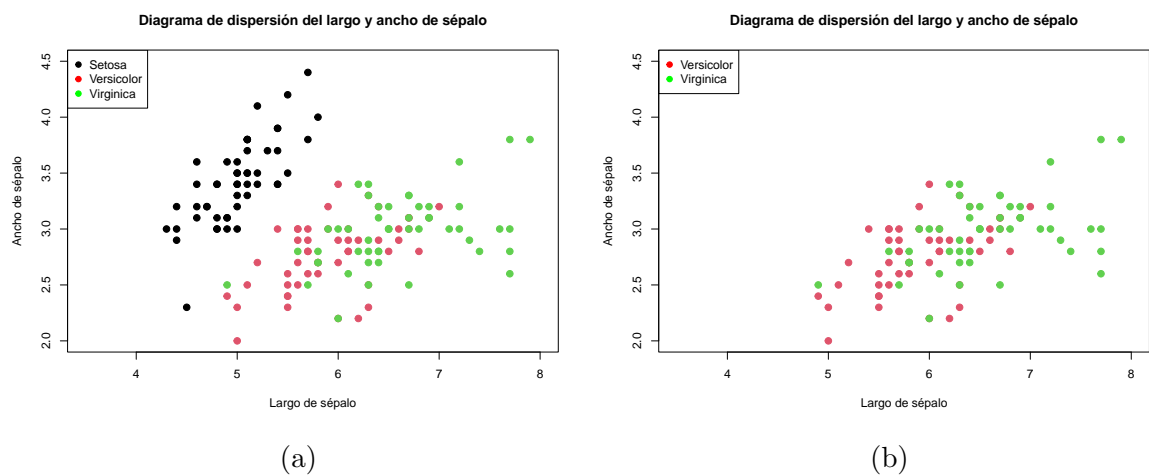


Figura 3: Diagrama de dispersión entre el largo y ancho de sépalo de (a) las tres especies (setosa, versicolor y virginica) y (b) de las dos especies (versicolor y virginica).

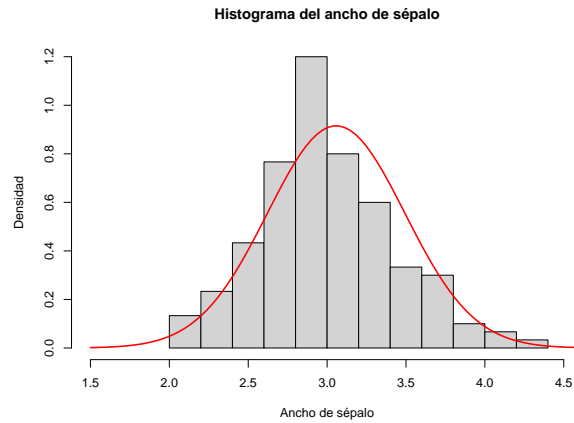


Figura 4: Histograma del ancho de sépalo

- a)  $P(X \leq 3,5)$
  - b)  $P(X > 3,5)$
  - c)  $P(2,5 \leq X \leq 3,5)$
  - d)  $P(X \leq k) = 0,8$
  - e) El cuantil correspondiente al 0,75.
  - f)  $f_X(3)$ ,  $f_X(2)$ ,  $f_X(4)$
5. Simula 100 observaciones de una variable Binomial con tamaño igual a 15 y probabilidad de éxito 0,6, es decir  $X \sim \text{Binomial}(\text{size} = 15, \text{prob} = 0,6)$ . Puedes seguir el siguiente código,

```
set.seed(1010)
n <- 15
p <- 0.6
X <- rbinom(100, size = n, prob = p)
```

- a) Se pide realizar el gráfico de la distribución Binomial con las simulación realizada, como se muestra en la figura 5.
- b) Calcular:
  - 1)  $P(X \leq 6)$
  - 2)  $P(X > 8)$
  - 3)  $P(7 < X \leq 10)$
  - 4)  $P(X \leq k) = 0,05$

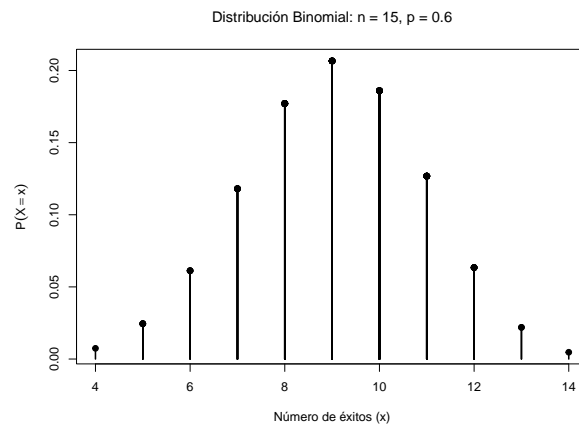


Figura 5: Densidad de la distribución Binomial con  $n=15$  y probabilidad de éxito igual a 0.6.