

Taller de práctica Prueba 1

Operaciones en R (Taller Introductorio)

Conceptos básicos de estadística (sección 1.1)

Medidas de Localización (sección 1.2)

Medidas de escala (sección 1.3)

Gráficos descriptivos: Histograma, Caja y Violín (secciones 1.5.1, 1.5.2 y 1.5.3)

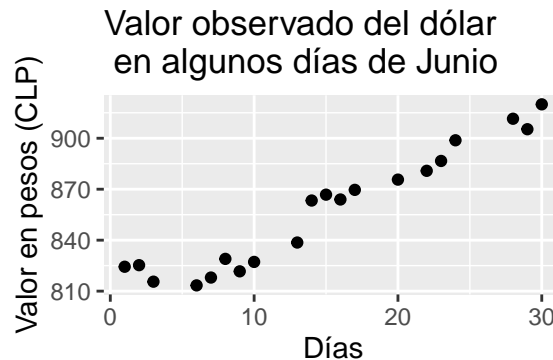
Parte 1: Operaciones en R

Para los siguientes enunciados, escriba el código asociado.

1. Defina el vector $x = (10, 20, 30, 40, 10, 20, 30, 40, 10, 20, 30, 40)$ de dos formas distintas.
2. Asigne los valores 550, 340, 920 y 615 a las variables $V1$, $V2$, $V3$ y $V4$. Luego, calcule el promedio de estas mediante el comando `mean()`.
3. Genere una matriz de 4 filas y 4 columnas que contenga los números pares desde el 2 hasta el 32, sin que se repita alguno.

Parte 2: Conceptos básicos de estadística

1. Observe el siguiente gráfico, y explique conceptualmente, cuál es el dato y cuál sería la información. ¿Qué se necesita para transitar de un concepto a otro (¿y en qué dirección se transita?)?



2. Realice un esquema o mapa conceptual que relacione los conceptos de población, muestra, parámetro, estadístico, estimador y estimación.
3. Suponga, que está interesado en predecir la variación porcentual del IPC para el siguiente mes considerando un rezago a un mes (es decir, respecto al mes anterior), para ello, usted sabe que este valor será similar al promedio de las variaciones porcentuales del IPC a rezago de un mes de los meses anteriores (históricamente), sin embargo, utiliza los últimos doce meses, dando como resultado un 3%. A continuación, especifique, la población, la muestra, el parámetro, el estadístico, el estimador y la estimación del estudio. ¿Qué sucede si se considerarán para el cálculo los últimos 24 meses o los últimos 36 meses?, explique el concepto estadístico detrás de esta situación.
4. ¿Por qué es incorrecto hablar de una muestra representativa? Justifique.

Parte 3: Medidas de Localización y Escala

1. ¿Cuál es la desventaja del promedio como medida de resumen? Explique.
2. Mencione las ventajas y desventajas de utilizar el rango, la varianza (o desviación estándar) y el coeficiente de variación para comparar la variabilidad de los datos de dos más muestras.
3. La siguiente base de datos contiene registros del Índice de Confianza del Consumidor (ICC). Este indicador de confianza del consumidor proporciona una indicación de la evolución futura del consumo y el ahorro de los hogares. Un indicador por encima de 100 señala un aumento en la confianza de los consumidores hacia la situación económica futura, como consecuencia de la cual son menos propensos a ahorrar y más inclinados a gastar dinero en compras importantes en los próximos 12 meses. Los valores por debajo de 100 indican una actitud pesimista hacia la evolución futura de la economía, lo que posiblemente resulte en una tendencia a ahorrar más y consumir menos.

Las variables que contiene la base de datos son las siguientes:

- Locacion: lugar en donde se mide el ICC (FRA = Francia, POL = Polonia, OECD = OCDE, ESP = España, BEL = Bélgica, ITA = Italia, DEU = Alemania).
- Mes: corresponde al mes en el que se realiza la medición del índice.
- Ano: corresponde al año en el que se realiza la medición del índice.
- ICC: valor del índice de confianza del consumidor.

La base de datos al cargarse en R se ha guardado con el nombre de `datos`.

- a) Interprete las siguientes salidas de código.

```
mean(datos$ICC[datos$Locacion == "ESP"])
## [1] 100.342

mean(datos$ICC[datos$Locacion == "DEU"])
## [1] 100.7433

median(datos$ICC[datos$Ano %in% c(2017:2020) & datos$Mes %in% c(2:5)])
## [1] 100.8683

median(datos$ICC[datos$Locacion == "OECD" & datos$Mes %in% c(3,6,7)])
## [1] 100.5179
```

- b) Interprete las siguientes salidas, y compárelas según sea posible.

```
var(datos$ICC[datos$Locacion == "ESP"])
## [1] 9.03605

var(datos$ICC[datos$Locacion == "DEU"])
## [1] 1.605925

sd(datos$ICC[datos$Locacion == "ESP"])/abs(mean(datos$ICC[datos$Locacion == "ESP"]))*100
## [1] 2.995757

sd(datos$ICC[datos$Locacion == "DEU"])/abs(mean(datos$ICC[datos$Locacion == "DEU"]))*100
## [1] 1.257901
```

Parte 4: Gráficos descriptivos

La siguiente base contiene datos respecto a los ataques al corazón de distintos pacientes hospitalarios. Algunas de las columnas de la base de datos son las siguientes:

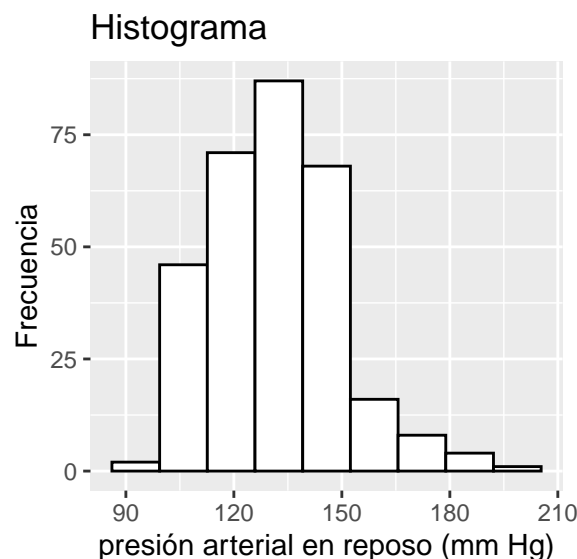
- sex: sexo del paciente (Hombre “H” y Mujer “M”).
- cp: Tipo de dolor en el pecho, Valor 1: angina típica, Valor 2: angina atípica, Valor 3: dolor no anginoso, Valor 4: asintomático.
- trtbps: presión arterial en reposo (en mm Hg).
- fbs: azúcar en sangre en ayunas > 120 mg/dl (V = verdadero; F = falso).
- thalachh: frecuencia cardíaca máxima alcanzada (en latidos por minuto).
- oldpeak: tiempo de duración del último ataque al corazón (en minutos).

La base de datos al cargarse en R se ha guardado con el nombre de `datos`.

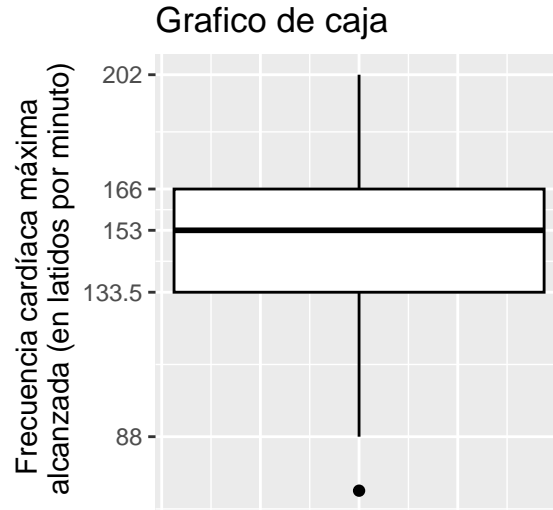
```
str(datos)

## 'data.frame': 303 obs. of 14 variables:
## $ age      : int  63 37 41 56 57 57 56 44 52 57 ...
## $ sex      : int  1 1 0 1 0 1 0 1 1 1 ...
## $ cp       : int  4 3 2 2 1 1 2 2 3 3 ...
## $ trtbps   : int  145 130 130 120 120 140 140 120 172 150 ...
## $ chol     : int  233 250 204 236 354 192 294 263 199 168 ...
## $ fbs      : int  1 0 0 0 0 0 0 0 1 0 ...
## $ restecg  : int  0 1 0 1 1 1 0 1 1 1 ...
## $ thalachh : int  150 187 172 178 163 148 153 173 162 174 ...
## $ exng     : int  0 0 0 0 1 0 0 0 0 0 ...
## $ oldpeak  : num  2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
## $ slp      : int  0 0 2 2 2 1 1 2 2 2 ...
## $ caa      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ thall    : int  1 2 2 2 2 1 2 3 3 2 ...
## $ output   : int  1 1 1 1 1 1 1 1 1 1 ...
```

1. Interprete el siguiente histograma haciendo énfasis en las frecuencias más altas.



2. Interprete el siguiente gráfico de caja haciendo énfasis en las medidas de posición.

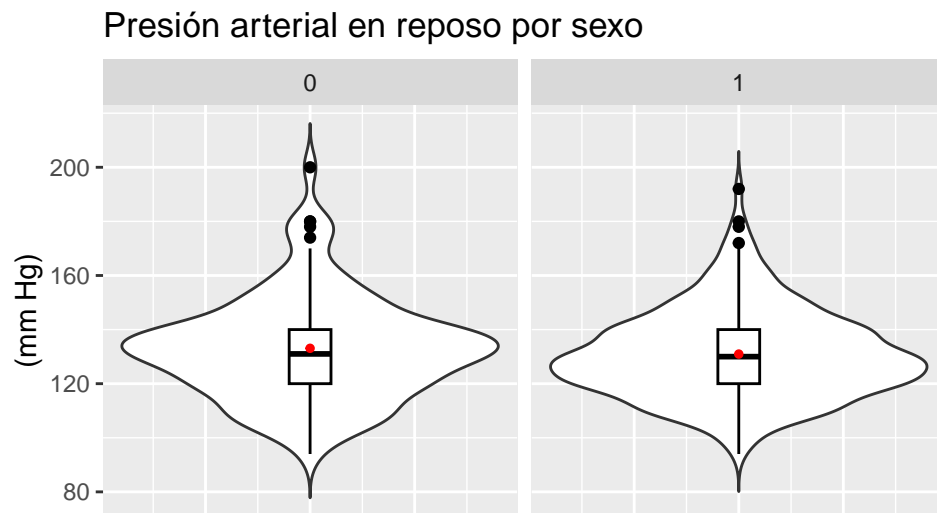


3. Interprete cada uno de los gráficos de violín por separado, haciendo énfasis en los siguientes aspectos:

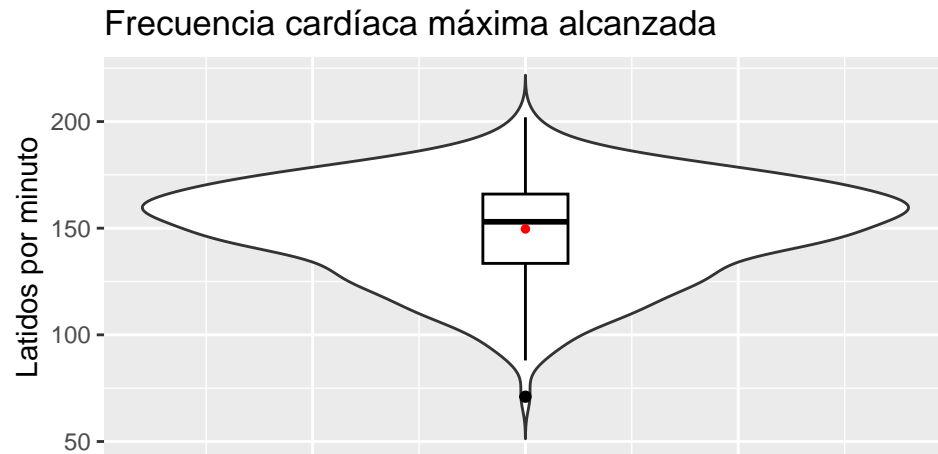
- La ubicación de las mayores concentraciones de datos.
- Posición del promedio respecto a la mediana.
- Posibles razones de los ocurrido en *b*).

En caso de ser un gráfico segregado por una segunda variable, compare los violines respecto a la ubicación del promedio y la longitud vertical del gráfico (interprete esto como dispersión de los datos).

a) Gráficos de violín de la presión arterial por sexo.



b) Gráfico de violín de la frecuencia cardíaca.



c) Gráficos de violín del tiempo de duración del último ataque al corazón, según el tipo de dolor de pecho.

