

# Estadística I

## Tema 1: Análisis de datos univariantes

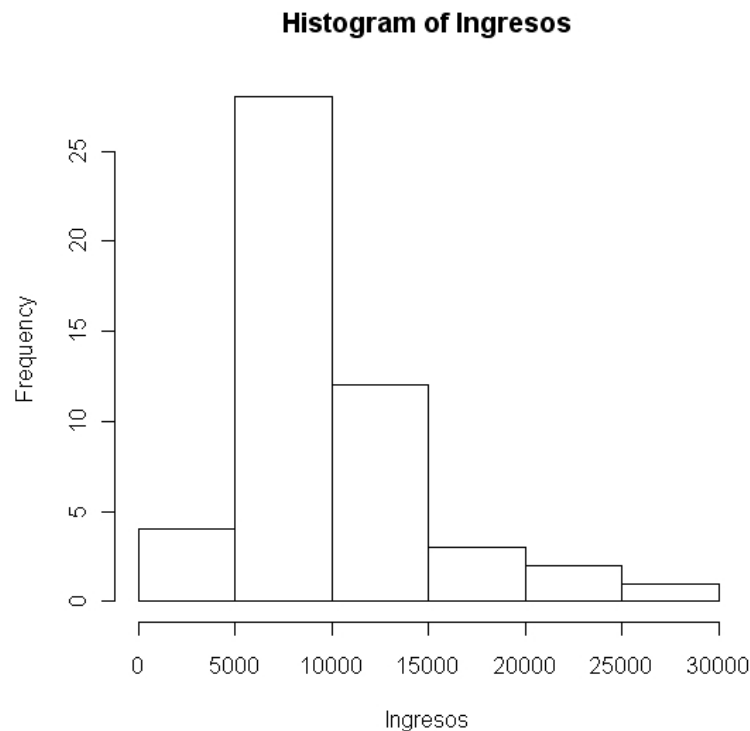
### SOLUCIONES

Grado en Administración de Empresas 08/09

1. a) Tabla de distribución de frecuencias absolutas con seis intervalos de igual amplitud:

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia ni |
|-------|-------------|-------------|----------------|---------------|
| 1     | 0           | 5000        | 2500           | 4             |
| 2     | 5000        | 10000       | 7500           | 28            |
| 3     | 10000       | 15000       | 12500          | 12            |
| 4     | 15000       | 20000       | 17500          | 3             |
| 5     | 20000       | 25000       | 22500          | 2             |
| 6     | 25000       | 30000       | 27500          | 1             |

b) La distribución es unimodal asimétrica a la derecha.



c) Por la presencia de datos atípicos en la muestra (valores muy grandes en comparación a la mayoría de los valores), es preferible el uso de la mediana (9163) al de la media (10040.36). Para medir la dispersión podemos utilizar la desviación típica (4591.89) o el rango intercuartílico (4095.5), pero no el rango (22494) ya que es muy sensible a la presencia de datos atípicos.

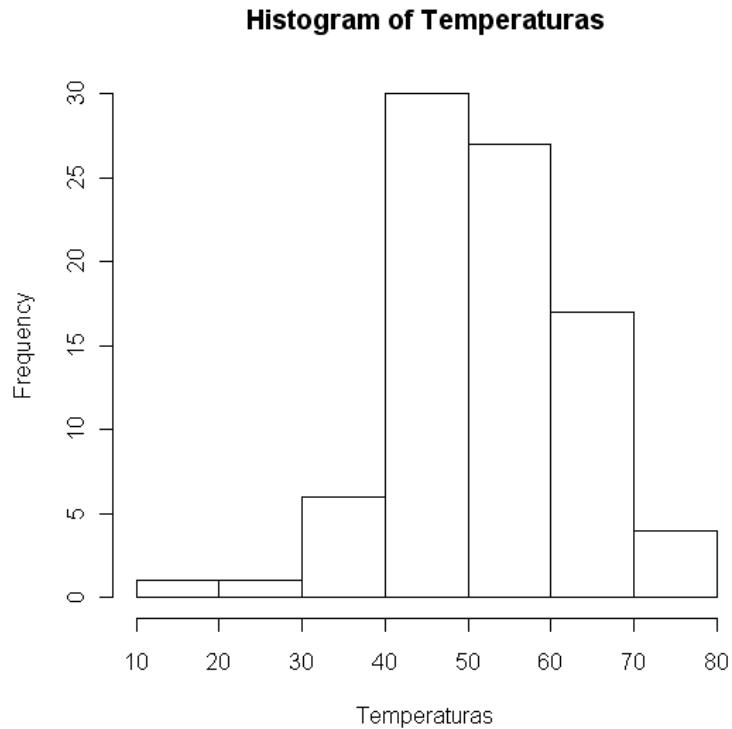
2. a) Distribución de frecuencias absolutas, usando intervalos de longitud 10, el primero de ellos empezando en 10.

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia ni |
|-------|-------------|-------------|----------------|---------------|
| 1     | 10          | 20          | 15             | 1             |
| 2     | 20          | 30          | 25             | 1             |

|   |    |    |    |    |
|---|----|----|----|----|
| 3 | 30 | 40 | 35 | 6  |
| 4 | 40 | 50 | 45 | 30 |
| 5 | 50 | 60 | 55 | 27 |
| 6 | 60 | 70 | 65 | 17 |
| 7 | 70 | 80 | 75 | 4  |

---

b) La distribución es unimodal asimétrica a la izquierda.



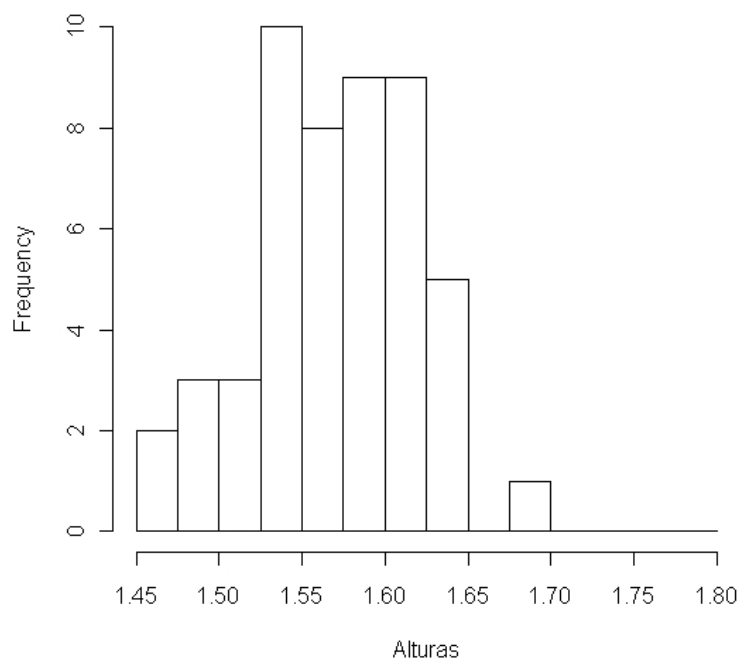
3. a) La variable “altura.<sup>es</sup> cuantitativa puesto que toma valores numéricos. Es continua porque puede tomar cualquier valor dentro de un cierto intervalo (aunque luego se redondee a dos decimales) y no sólo un número finito o numerable de valores.
- b) Distribución de frecuencias absolutas, usando intervalos de longitud 2.5 **cm**, el primero de ellos empezando en 1.45 **m**.

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia ni |
|-------|-------------|-------------|----------------|---------------|
| 1     | 1.450       | 1.475       | 1.463          | 2             |
| 2     | 1.475       | 1.500       | 1.487          | 3             |
| 3     | 1.500       | 1.525       | 1.513          | 3             |
| 4     | 1.525       | 1.550       | 1.538          | 10            |
| 5     | 1.550       | 1.575       | 1.563          | 8             |
| 6     | 1.575       | 1.600       | 1.588          | 9             |
| 7     | 1.600       | 1.625       | 1.613          | 9             |
| 8     | 1.625       | 1.650       | 1.638          | 5             |
| 9     | 1.650       | 1.675       | 1.663          | 0             |
| 10    | 1.675       | 1.700       | 1.688          | 1             |

---

c) La distribución es unimodal ligeramente asimétrica a la derecha.

**Histogram of Alturas**



4. a) Tabla de distribución de frecuencias relativas:

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia $f_i$ |
|-------|-------------|-------------|----------------|------------------|
| 1     | 0           | 3           | 1.5            | 0.42             |
| 2     | 3           | 6           | 4.5            | 0.26             |
| 3     | 6           | 9           | 7.5            | 0.15             |
| 4     | 9           | 12          | 10.5           | 0.08             |
| 5     | 12          | 15          | 13.5           | 0.04             |
| 6     | 15          | 18          | 16.5           | 0.03             |
| 7     | 18          | 21          | 19.5           | 0.02             |

b) Tabla de distribución de frecuencias relativas acumuladas:

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia $F_i$ |
|-------|-------------|-------------|----------------|------------------|
| 1     | 0           | 3           | 1.5            | 0.42             |
| 2     | 3           | 6           | 4.5            | 0.68             |
| 3     | 6           | 9           | 7.5            | 0.83             |
| 4     | 9           | 12          | 10.5           | 0.91             |
| 5     | 12          | 15          | 13.5           | 0.95             |
| 6     | 15          | 18          | 16.5           | 0.98             |
| 7     | 18          | 21          | 19.5           | 1                |

5. a) Distribución de frecuencias absolutas, usando intervalos de longitud 5, el primero de ellos empezando en 15:

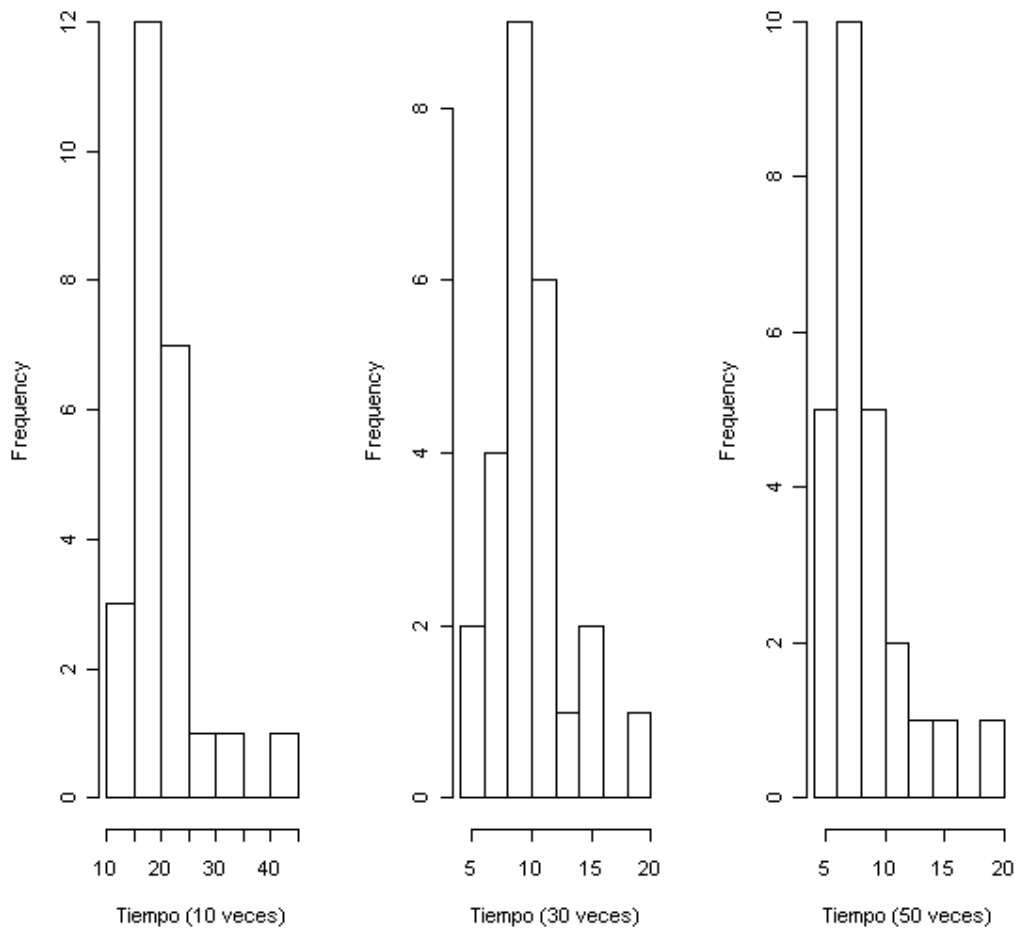
| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frecuencia $n_i$ |
|-------|-------------|-------------|----------------|------------------|
| 1     | 15          | 20          | 17.5           | 3                |
| 2     | 20          | 25          | 22.5           | 6                |

|   |    |    |      |   |
|---|----|----|------|---|
| 3 | 25 | 30 | 27.5 | 7 |
| 4 | 30 | 35 | 32.5 | 5 |
| 5 | 35 | 40 | 37.5 | 2 |
| 6 | 40 | 45 | 42.5 | 2 |

b) y c) Distribución de frecuencias relativas y frecuencias relativas acumuladas:

| Clase | Límite inf. | Límite Sup. | Marca de Clase | Frec. fi | Frec. Fi |
|-------|-------------|-------------|----------------|----------|----------|
| 1     | 15          | 20          | 17.5           | 0.12     | 0.12     |
| 2     | 20          | 25          | 22.5           | 0.24     | 0.36     |
| 3     | 25          | 30          | 27.5           | 0.28     | 0.64     |
| 4     | 30          | 35          | 32.5           | 0.20     | 0.84     |
| 5     | 35          | 40          | 37.5           | 0.08     | 0.92     |
| 6     | 40          | 45          | 42.5           | 0.08     | 1        |

6. a) Histogramas de frecuencias para cada uno de los conjuntos de datos anteriores.



b) Parece existir una relación entre el número de veces que se realiza una tarea y el tiempo que se necesita para realizarla, pues notamos en los histogramas que los intervalos más frecuentes son: 18-20, 8-10, y 6-8, respectivamente, señalándonos que el tiempo necesario disminuye a medida que se incrementa el número de veces que se realiza la operación. También, en cuanto a la tasa de disminución decreciente en el tiempo necesario, observamos que de 10 a

30 veces disminuimos aproximadamente 10 unidades, y de 30 a 50 veces, sólo disminuimos en 2 unidades. Estas dos observaciones son congruentes con el enunciado.

- c)  $t_{10}$  = Tiempo de realización después de 10 repeticiones.  
 $t_{30}$  = Tiempo de realización después de 30 repeticiones.  
 $t_{50}$  = Tiempo de realización después de 50 repeticiones.

Medias y medianas:

$$\begin{aligned} \overline{t_{10}} &= 20.92 & \overline{t_{30}} &= 10.36 & \overline{t_{50}} &= 8.88 \\ Me(t_{10}) &= 20.00 & Me(t_{30}) &= 10.00 & Me(t_{50}) &= 8.00 \end{aligned}$$

- d) El coeficiente de variación de una variable estadística se define como el cociente entre su desviación típica y su media, y por tanto no tiene unidades.

$$CV(t_{10}) = 0.3026 \quad CV(t_{30}) = 0.3114 \quad CV(t_{50}) = 0.3892$$

7. a) La variable “número mensual de accidentes.” es cuantitativa ya que toma valores numéricos. Es discreta ya que los únicos valores que puede tomar son los naturales  $(0, 1, 2, \dots)$ .

b)

$$\bar{x} = 3 \quad Me(x) = 2 \quad Mo(x) = 2$$

8.

$$\bar{x} = 1.5718 \quad Me(x) = 1.57 \quad Mo(x) = 1.62 \quad Q_1 = 1.5325 \quad Q_3 = 1.62$$

9. a)

$$\bar{x} = 28.875 \quad Me(x) = 29.5 \quad Mo(x) = 19$$

b) Calcula el rango y el rango intercuartílico.

$$Q_1 = 19 \quad Q_3 = 39.5$$

$$R = 42 \quad RI = 20.5$$

c) Calcula la varianza y la desviación típica.

$$s^2 = 143.856 \quad s = 11.99$$

d) Diagrama de caja:

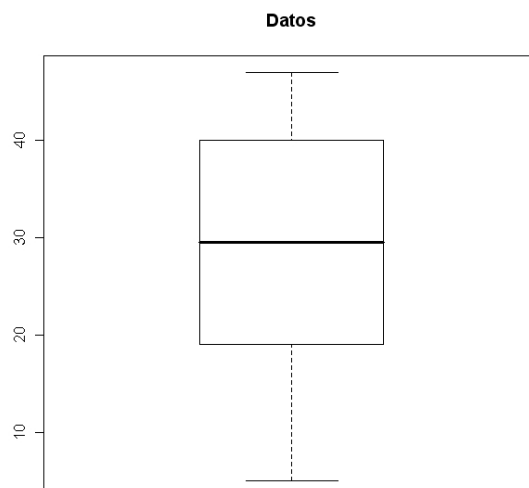


Diagrama de caja

e)

$$Q_1 - 1.5RI = -11.75 \quad Q_3 + 1.5RI = 70.25$$

No existen datos atípicos ya que todos los datos se encuentran entre esos dos valores.

e) La media es ligeramente inferior a la mediana, por lo que los datos presentan una leve asimetría a la izquierda (esto se puede apreciar en el diagrama de caja).

10. a)

$$\bar{x} = 7.55 \quad Me(x) = 7.8 \quad Mo(x) = 7.7$$

b) Calcula el rango y el rango intercuartílico.

$$Q_1 = 7.3 \quad Q_3 = 8.85$$

$$R = 7.3 \quad RI = 1.55$$

c) Calcula la varianza y la desviación típica.

$$s^2 = 3.07 \quad s = 1.75$$

d) Diagrama de caja:

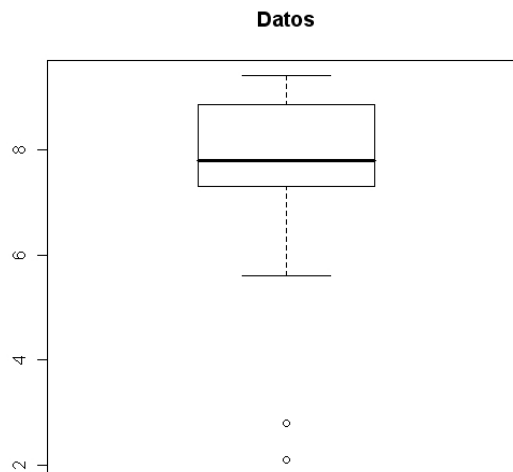


Diagrama de caja

e)

$$Q_1 - 1.5RI = 4.975 \quad Q_3 + 1.5RI = 11.175$$

Hay dos notas atípicas, 2.1 y 2.8, ya que se encuentran por debajo del valor 5.475. Los podemos apreciar en el diagrama de caja.

f) La media es inferior a la mediana, por lo que los datos presentan una asimetría a la izquierda (esto se puede apreciar en el diagrama de caja).

11. a)

$$\bar{x} = 180 \quad Me(x) = 162 \quad Mo(x) = 162$$

b) Puesto que la mediana es inferior a la media, hay asimetría a la derecha.

12. a)

$$R = 1325.49 \quad s^2 = 240952 \quad s = 490.869$$

b) El rango y la desviación típica en millones de dólares, y la varianza en millones de dólares al cuadrado.

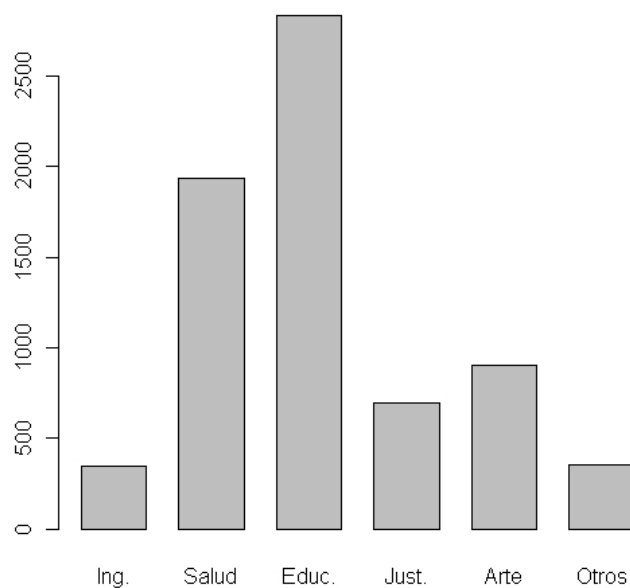
- c) Notemos que la observación correspondiente a Eastern (-852.32) es el valor inferior, y por tanto, al modificarse varía el rango y disminuye (no se habría modificado si no fuese uno de los extremos inferior o superior).

Veamos qué ocurre en el caso de la varianza muestral,  $s^2$ . La varianza muestral es la media de las distancias al cuadrado de los valores de la muestra a la media, es decir,

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Puesto que el nuevo valor de Eastern va a estar más cerca de la media que el anterior (se puede calcular la media, y se obtiene  $\bar{x} = -53.71857$ ), el término  $(x_i - \bar{x})^2$  correspondiente a Eastern será ahora menor, y por lo tanto la varianza disminuirá.

13. Diagrama de barras:



14. Diagrama de caja:

