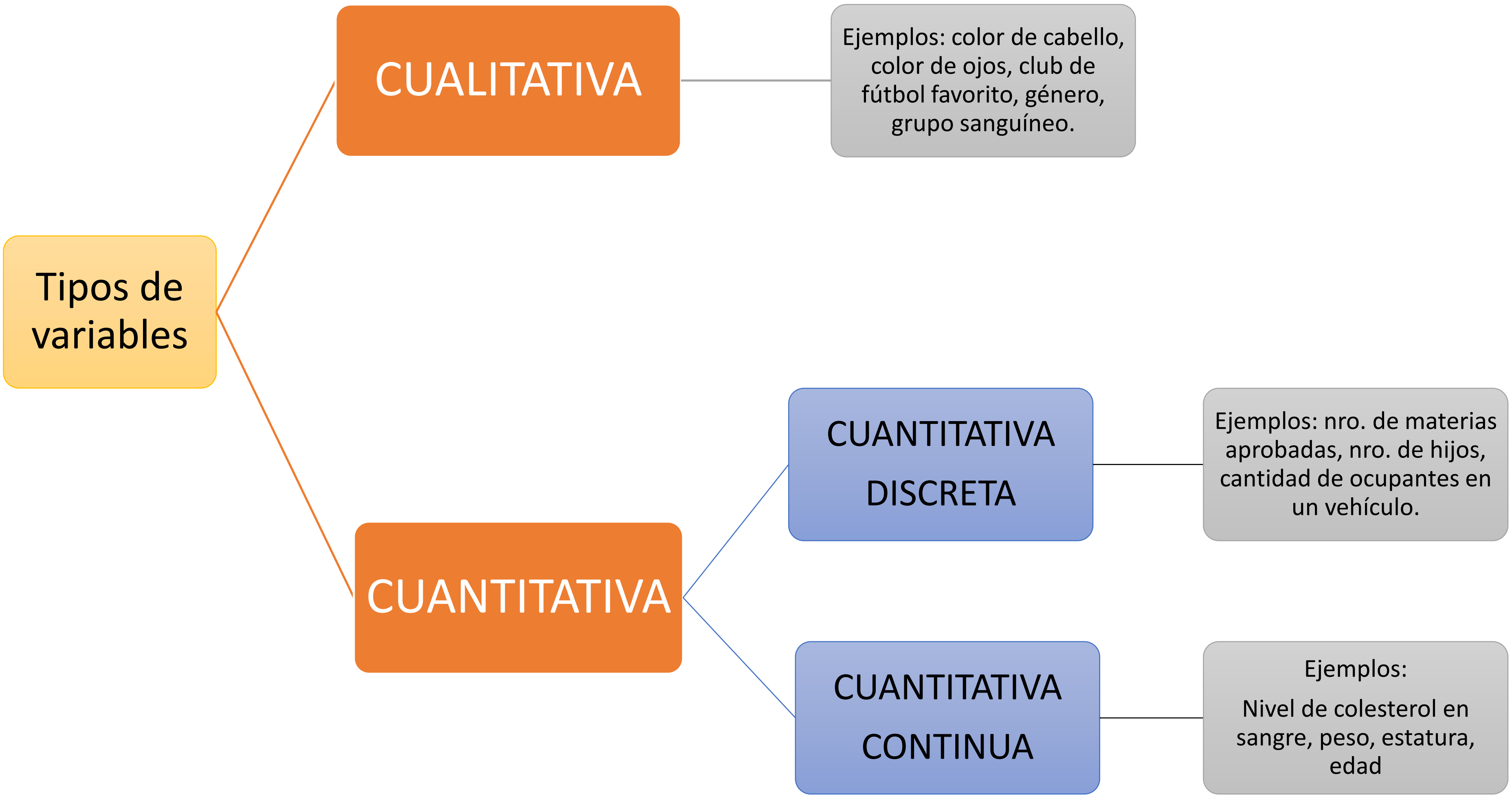
Introducción



Ejercicio 1

Los siguientes son los puntajes que obtuvieron 40 estudiantes de una comisión en un parcial de Probabilidad y Estadística:

75 – 89 – 66 – 52 – 90 – 68 – 83 – 94 – 77 – 60 – 63 – 77 – 31 – 88 – 74 – 37 – 85 – 76 – 74 – 63
38 – 47 – 87 – 65 – 97 – 49 – 65 – 72 – 73 – 81 – 69 - 72 – 91 – 87 – 76 – 58 – 63 – 70 – 72 – 65

a) Defina la variable en estudio. Realice el gráfico de tallo y hojas.

Variable: Puntaje que obtuvieron los estudiantes en el parcial de probabilidad y estadística.

Clasificación: Cuantitativa continua

Para construir el diagrama de tallo y hoja descomponemos las observaciones en decena y unidad, las vamos ordenando en diferentes filas y luego las reorganizamos dentro de cada fila.

3	1 7 8
4	7 9
5	2 8
6	6 8 0 3 3 5 5 9 3 5
7	5 7 7 4 6 4 2 3 2 6 0 2
8	9 3 8 5 7 1 7
9	0 4 7 1

→

3	1 7 8
4	7 9
5	2 8
6	0 3 3 3 5 5 5 6 8 9
7	0 2 2 2 3 4 4 5 6 6 7 7
8	1 3 5 7 7 8 9
9	0 1 4 7

Ejercicio 1

b) Agrupe los datos en intervalos de clase y construya la tabla de distribución de frecuencias completa.

Si n=40, el número de intervalos de clase podemos obtenerlo como:

$$k = \sqrt{n} = \sqrt{40} = 6{,}32 \ldots \cong 6$$

Luego, calculamos la amplitud de los intervalos de clase:

$$A_{IC} = \frac{R}{k}$$

Recordemos que el rango es: $R = x_{máx} - x_{mín} = 97 - 31 = 66$

$$A_{IC} = \frac{R}{k} = \frac{66}{6} = 11$$

Construimos 6 intervalos de clase de amplitud 11. Iniciamos con el dato mínimo y notemos que el último intervalo contenga al dato máximo. Los intervalos son todos semicerrados a izquierda y el último intervalo es cerrado en ambos extremos.

Ejercicio 1

3	1 7 8
4	7 9
5	2 8
6	0 3 3 3 5 5 5 6 8 9
7	0 2 2 2 3 4 4 5 6 6 7 7
8	1 3 5 7 7 8 9
9	0 1 4 7

Tabla de distribución de frecuencias

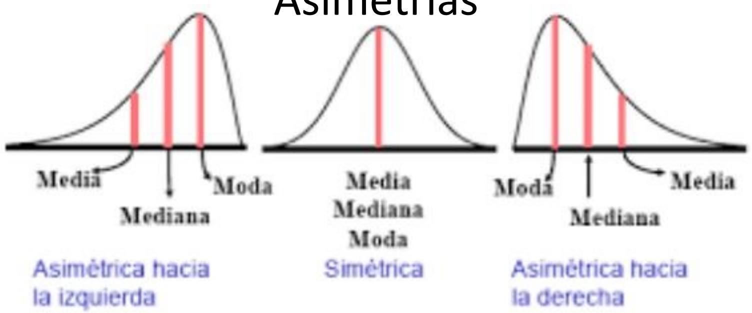
Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$	Frec. acumulada porcentual $Fr_{\%} = Fr_i.100\%$
[31,42)	36.5	3	0.075	3	0.075	7.5
[42,53)	47.5	3	0.075	6	0.150	15
[53,64)	58.5	5	0.125	11	0.275	27.5
[64,75)	69.5	13	0.325	24	0.600	60
[75,86)	80.5	8	0.200	32	0.800	80
[86,97]	91.5	8	0.200	40	1.000	100
Total		40	1.000			

Ejercicio 1

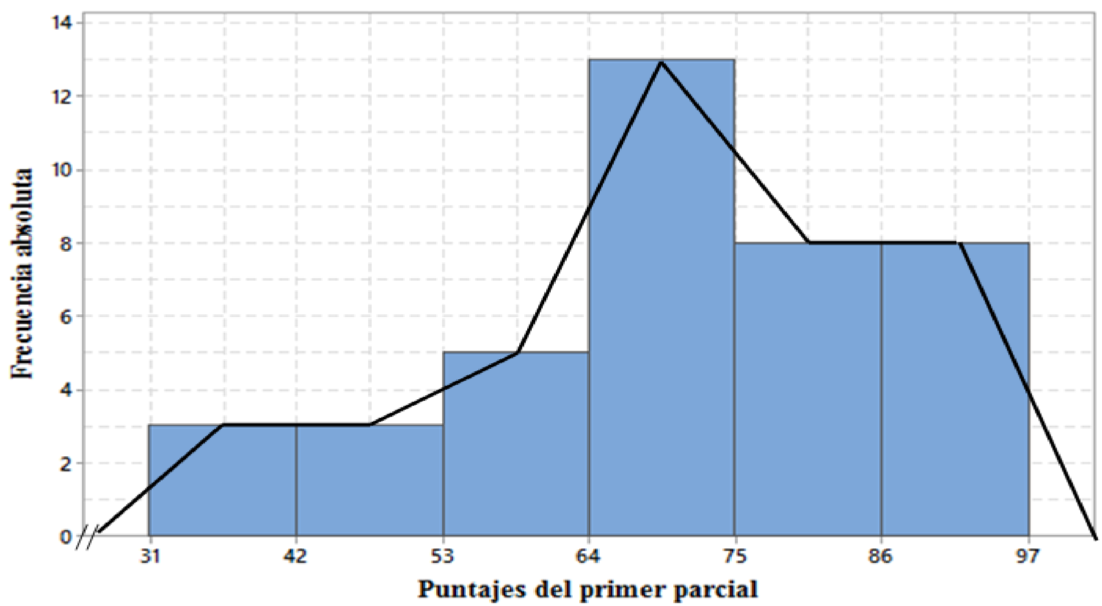
c) Realice el histograma de frecuencias absolutas y el polígono de frecuencias correspondiente. Comente acerca de la simetría del gráfico construido.

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

Asimetrías



Histograma de frecuencias absolutas y polígono de frecuencias para los puntajes del primer parcial

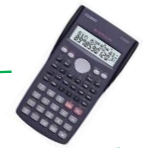


El gráfico presenta una cola hacia los puntajes más bajos. **Es un caso de asimetría negativa** o asimetría hacia la izquierda.

Ejercicio 1

d) Calcule e interprete las siguientes medidas: media, mediana, cuartiles, moda, desvío estándar y rango.

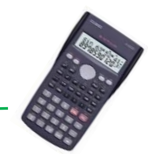
- **Media:** $\bar{x} = \sum_{i=1}^{k} \frac{x_{PM_i} f a_i}{n} = 70{,}6$ ←



Interpretación: **El puntaje promedio de la comisión es de 70,6.**

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

- **Desviación estándar:** $s = \sqrt{\sum_{i=1}^{n} \frac{(x_{PM_i} - \bar{x})^2 f a_i}{n-1}} = 16{,}105$ ←



Interpretación: **La desviación típica de los puntajes respecto de la media es 16,105.**

- **Rango:** $\mathrm{R} = x_{máx} - x_{mín} = 97 - 31 = 66$

Interpretación: **El rango de los puntajes de los alumnos es 66.**

3	178
4	79
5	28
6	0333555689
7	02223445667
8	1357789
9	0147

Ejercicio 1

- **Cuartiles:** $Q_k = Li + \frac{\frac{n}{4}.k - Fa_{(k-1)}}{fa_k} .A_{IC}$

Cuartil 1:

Buscamos la primera clase cuya frecuencia relativa acumulada es mayor o igual a 0,25.

Si $Q_1 \in [53,64)$, entonces:

$$Q_1 = 53 + \frac{\frac{40}{4}.1 - 6}{5}.11 = 61,8$$

Interpretación: **El 25% de los estudiantes obtuvo puntajes menores a 61,8.**

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

Cuartil 2 o Mediana:

Buscamos la primera clase cuya frecuencia relativa acumulada es mayor o igual a 0,50.

Si $Q_2 \in [64,75)$, entonces:

$$\tilde{x} = Q_2 = 64 + \frac{\frac{40}{4}.2 - 11}{13}.11 = 71,6$$

Interpretación: **La mitad de los estudiantes de la comisión obtuvo puntajes menores a 71.62**

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

Ejercicio 1

Cuartil 3:

Buscamos la primera clase cuya frecuencia relativa acumulada es mayor o igual a 0,75.

Si $Q_3 \in [75{,}86)$, entonces:

$$Q_3 = 75 + \frac{\frac{40}{4}.3 - 24}{8}.11 = 83{,}25$$

Interpretación: **El 75% de los estudiantes obtuvo puntajes menores a 83,25.**

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

- **Moda:** $\hat{x} = Li + \frac{d_1}{d_1+d_2}.A_{IC}$

con $d_1 = fa_k - fa_{k-1}$ y $d_2 = fa_k - fa_{k+1}$

Buscamos la clase con mayor frecuencia absoluta o relativa.

Si $\hat{x} \in [64{,}75)$, entonces:

$$\hat{x} = 64 + \frac{(13-5)}{(13-5)+(13-8)}.11 = 70{,}76$$

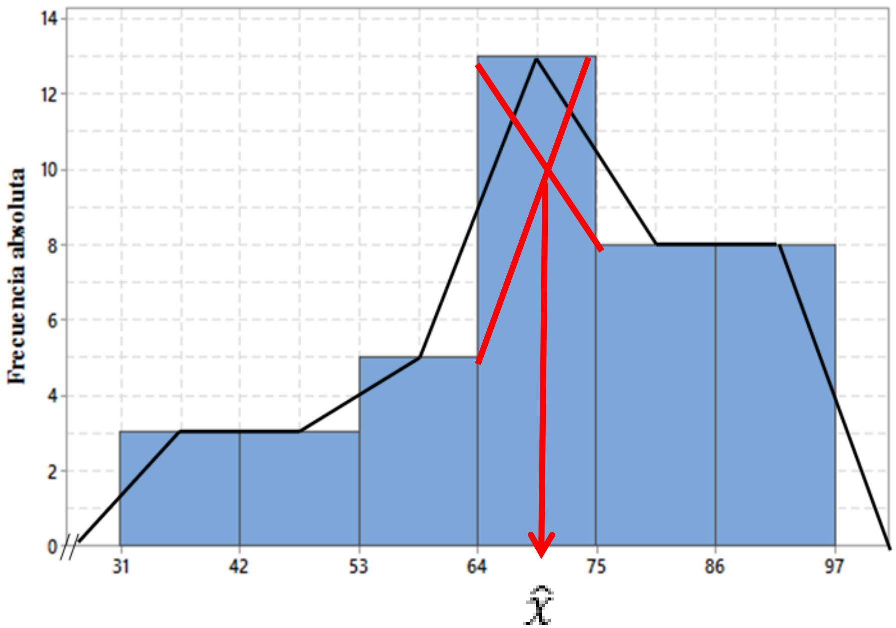
Interpretación: **Aproximadamente, el puntaje más usual es 71 puntos.**

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$
[31,42)	36.5	3	0.075	3	0.075
[42,53)	47.5	3	0.075	6	0.150
[53,64)	58.5	5	0.125	11	0.275
[64,75)	69.5	13	0.325	24	0.600
[75,86)	80.5	8	0.200	32	0.800
[86,97]	91.5	8	0.200	40	1.000
Total		40	1.000		

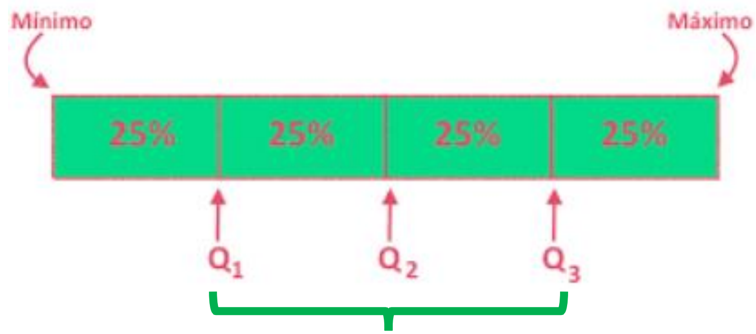
Importante!

También **podemos determinar la moda gráficamente** en el histograma de frecuencias absolutas o relativas:

$$\tilde{x} \cong 71$$



e) ¿Entre qué valores se encuentra el 50% central de la distribución de los puntajes?



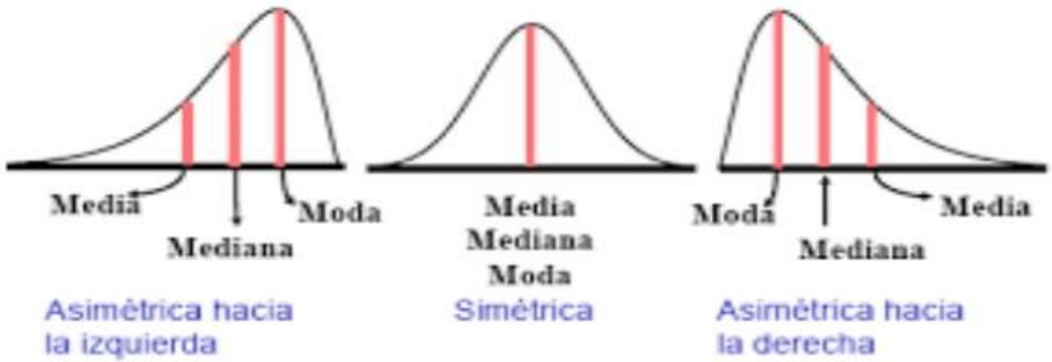
El 50% central de la distribución se encuentra entre los puntajes: 61.8 y 83.25 (cuartiles primero y tercero), aproximadamente.

Ejercicio 1

f) Calcule el coeficiente de asimetría de Pearson e interprete. ¿Guarda relación con lo concluido en el inciso c?

Coeficiente de asimetría de Pearson: $A_P = \dfrac{3(\bar{x}-\tilde{x})}{s}$

que resulta positivo si es la distribución es asimétrica a derecha, negativo si es asimétrica a izquierda, o es nulo en caso de ser una distribución simétrica.



$A_P < 0$ $\quad$ $A_P = 0$ $\quad$ $A_P > 0$

En estos datos: $A_P = \dfrac{3(70{,}6-71{,}62)}{16{,}105} \cong -0{,}19 < 0$

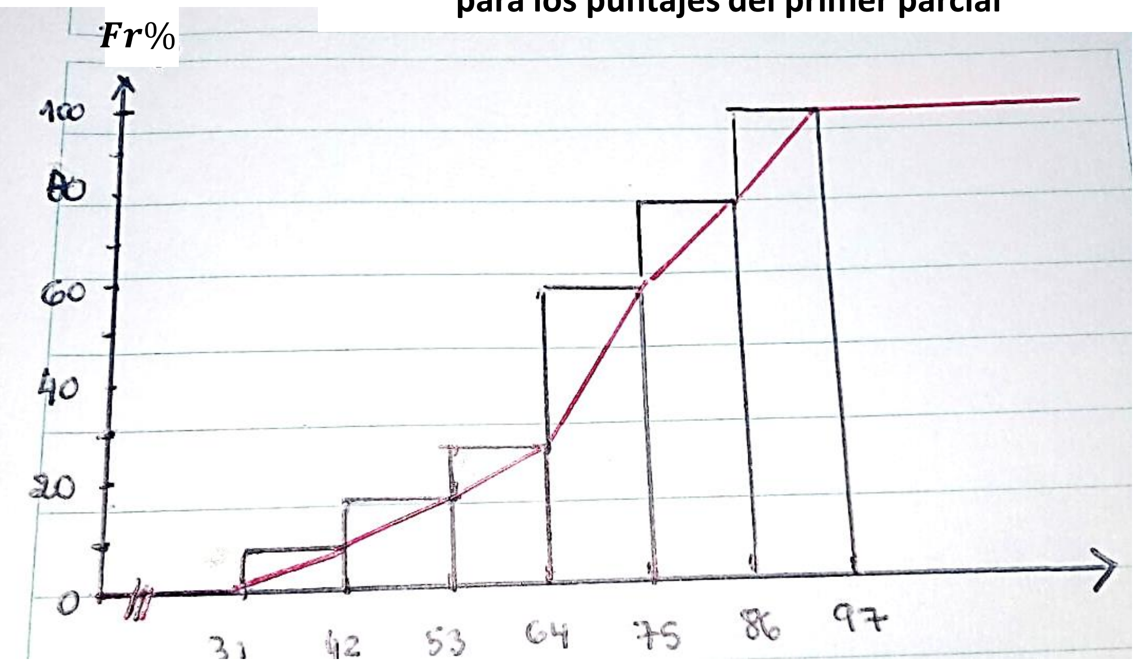
Concluimos que la distribución presenta leve asimetría hacia la izquierda, como lo observamos efectivamente en el histograma del inciso c.

Ejercicio 1

g) Construya el histograma de frecuencias acumuladas porcentuales y la ojiva correspondiente.

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$	Frec. acumulada porcentual $Fr_{\%} = Fr_i.100\%$
[31,42)	36.5	3	0.075	3	0.075	7,5
[42,53)	47.5	3	0.075	6	0.150	15
[53,64)	58.5	5	0.125	11	0.275	27,5
[64,75)	69.5	13	0.325	24	0.600	60
[75,86)	80.5	8	0.200	32	0.800	80
[86,97]	91.5	8	0.200	40	1.000	100
Total		40	1.000			

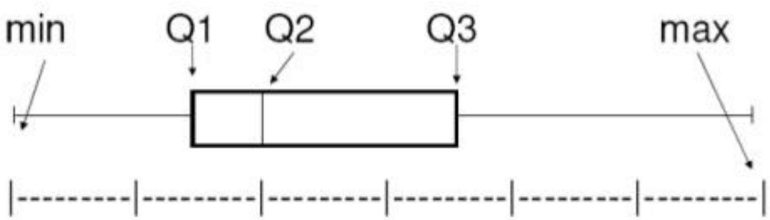
Histograma de acumuladas porcentuales y ojiva para los puntajes del primer parcial



Ejercicio 1

h) Realice el diagrama de caja y extensión de los datos. ¿Qué puede decir acerca de la asimetría del conjunto de datos a partir del gráfico?

Un gráfico asociado a los cuartiles es el **box-plot**: en un eje se ubican los siguientes 5 números extraídos de una muestra: mínimo, cuartil 1, cuartil 2, cuartil 3 y máximo.

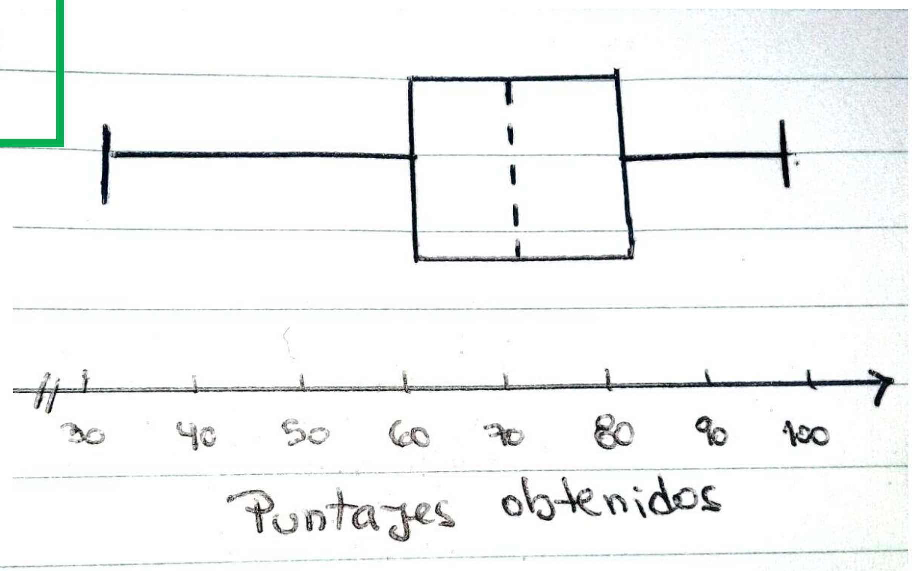


- Mínimo: $x_{mín} = 31$
- Primer cuartil: $Q_1 = 61{,}8$
- Mediana: $\tilde{x} = 71{,}6$
- Tercer cuartil: $Q_3 = 83{,}25$
- Máximo: $x_{máx} = 97$

Una regla para determinar si un dato es **anómalo** (outlier) es:

- Si un dato es < Q1 – 1.5(Q3-Q1)
- Si un dato es > Q3 + 1.5(Q3-Q1)

Diagrama de caja y extensión para los puntajes del primer parcial



¿Presentan outliers nuestros datos?

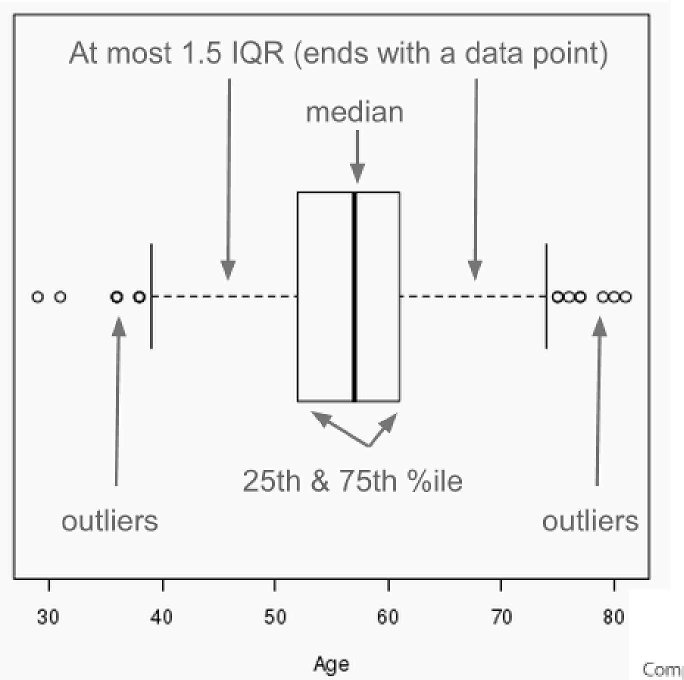
Calculamos:

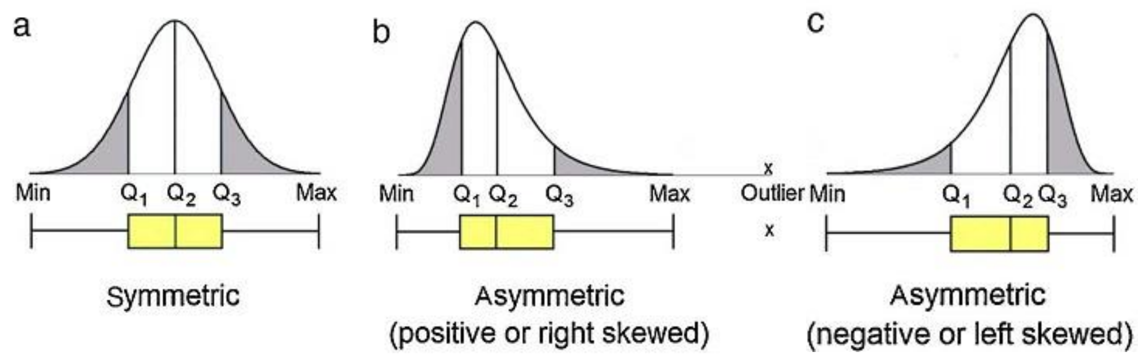
$$< Q_1 - 1{,}5(Q_3 - Q_1) = 29{,}625$$
$$> Q_3 + 1{,}5(Q_3 - Q_1) = 115{,}425$$

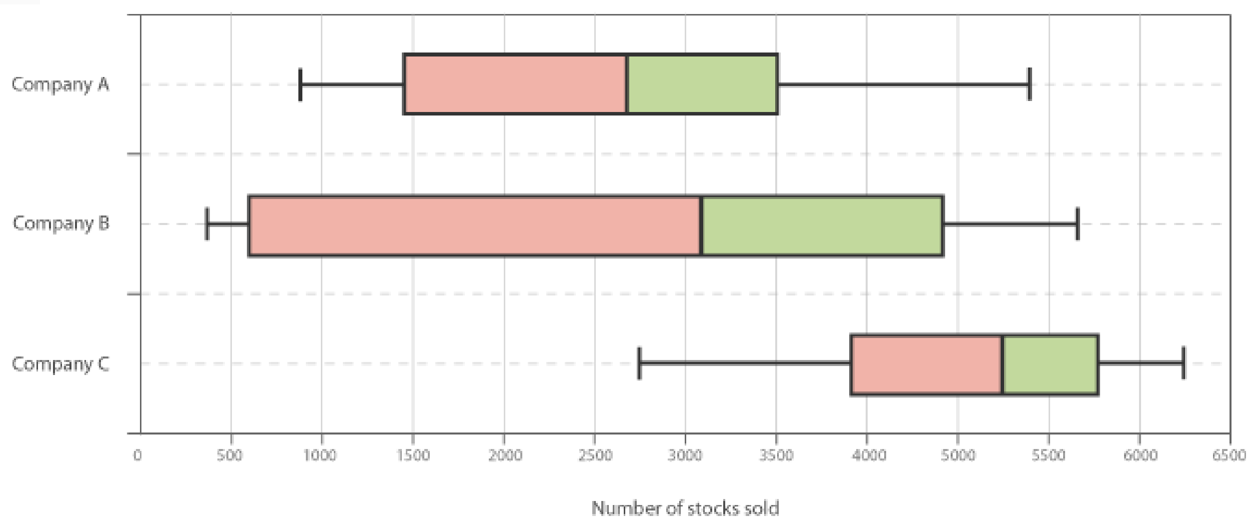
Vemos que en nuestros datos no existen observaciones menores a 29 ni mayores a 115, por lo tanto NO EXISTEN OUTLIERS en los puntajes.

Ejercicio 1

Más sobre los diagramas de caja y extensión…







Ejercicio 2

Un corrector de textos contabiliza el número de erratas que encuentra en cada página. Después de pasar este corrector por un texto de 50 páginas, se obtiene el siguiente número de erratas por página:

2 3 5 0 1 4 0 6 2 1 1 0 2 4 5 3 1 2 3 2 3 1 2 4 4
2 5 4 1 3 2 6 8 2 0 1 0 2 3 1 5 10 2 1 3 6 2 0 1 3

a) A partir del enunciado del problema, identifique la variable estadística y clasifique.

Variable: Cantidad de erratas en cada página.

Clasificación: Cuantitativa discreta.

b) Construya la tabla de frecuencias correspondiente.

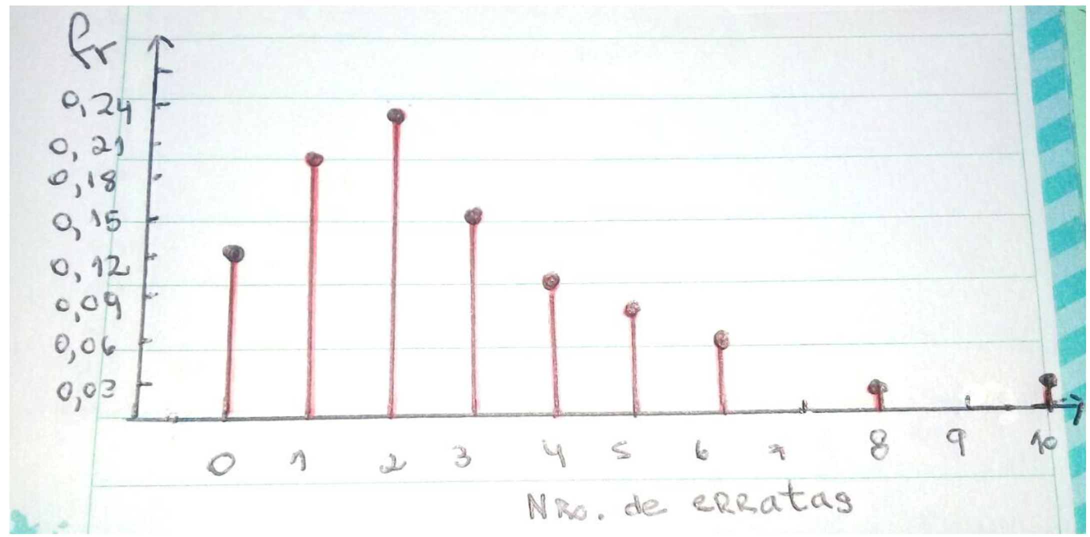
Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00
Total	50	1.00		

Ejercicio 2

c) Construya el diagrama de bastones para los datos. Solamente mirando este gráfico, ¿puede decir cuánto vale la moda?
En este caso lo construimos con la frecuencia relativa pero también puede ser utilizada la frecuencia.
Notemos que la barra más alta nos muestra el valor modal de la distribución: $\hat{x} = 2$

Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00
Total	50	1.00		

Diagrama de bastones para las cantidades de errores por página



d) ¿Qué porcentaje de las páginas tienen 2 erratas?
24% de las páginas tienen dos erratas. (Frec. rel. de 2)

¿Qué porcentaje de las páginas tiene menos de 6 erratas?
90% de las páginas tienen menos de 6 erratas. (Frec. rel. acum. de 5)

¿Qué porcentaje de las páginas tiene como mínimo 5 erratas?
18% de las páginas tiene como mínimo 5 erratas.
(sumamos las frec. rel. de 5 y maores a 5)

Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00
Total	50	1.00		

Ejercicio 2

e) Calcule los siguientes estadísticos descriptivos: media aritmética, mediana, moda, varianza y cuartiles.

- **Media :** $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 2{,}08$ ←

Interpretación: **En promedio la cantidad de errores por página es 2,08**

- **Mediana:**

Determinamos la **posición** de la mediana: $\frac{\boldsymbol{n+1}}{\boldsymbol{2}}$

Aquí $n = 50 \rightarrow \frac{n+1}{2} = \frac{50+1}{2} = 25{,}5$ →la mediana será un promedio de los 25° y 26° datos.

Buscamos el/los dato/s en la lista ordenada de observaciones.

$$\tilde{x} = \frac{x_{(25)} + x_{(26)}}{2} = \frac{2+2}{2} = 2$$

Interpretación: **La mitad de las páginas posee dos o menos errores por página.**

Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00

- **Moda:** $\hat{x} = 2$

Interpretación: **El número de erratas más frecuentes es 2.**

Ejercicio 2

e) Calcule los siguientes estadísticos descriptivos: media aritmética, mediana, moda, varianza y cuartiles.

- **Varianza:** $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = 2{,}1327^2 = 4{,}5484$ ←

Interpretación: **La variabilidad de los errores por página es aproximadamente 4.5484 errores² .**

- **Cuartiles:**

Cuartil 1:

Posición: $(50+1).\frac{1}{4} = 12{,}75 \rightarrow$Dato 13. $\therefore Q_1 = x_{(13)} = 1$

Interpretación:

El 25% de las páginas posee menos de un error.

Cuartil 3:

Posición: $(50+1).\frac{3}{4} = 38{,}25 \rightarrow$Dato 38. $\therefore Q_3 = x_{(38)} = 4$

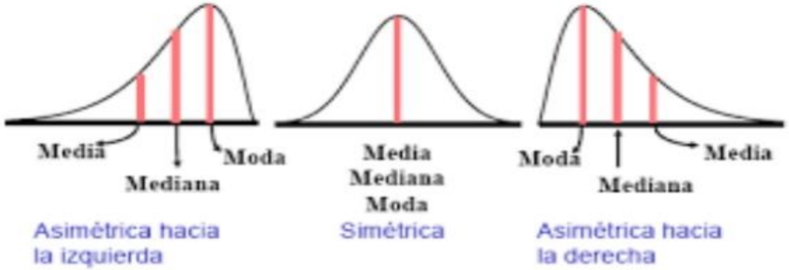
Interpretación: **El 75% de las páginas posee menos de 4 errores.**

Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00

Ejercicio 2

f) Si se comparan los estadísticos de tendencia central, ¿qué se puede decir de la simetría de la distribución?

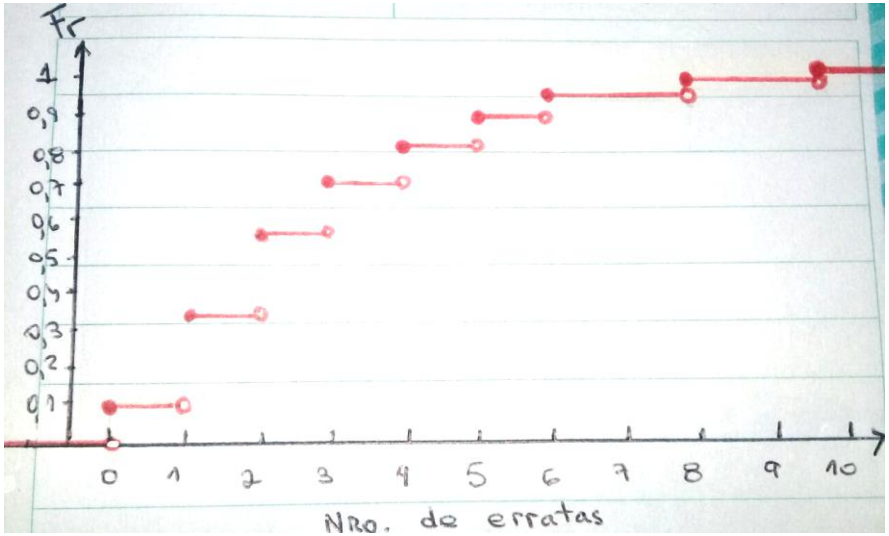
- Moda: $\hat{x} = 2$
- Mediana: $\tilde{x} = 2$
- Media: $\bar{x} = 2{,}08$



Notemos que al comparar los estadísticos de tendencia central, **la media es mayor que la mediana, indicándonos que la distribución presenta una asimetría hacia la derecha o asimetría positiva.**
La media se ve muy afectada por los valores extremos, por eso tiende a "desplazarse" a la cola de la distribución en casos de asimetría.

g) Construya el gráfico escalonado para las frecuencias acumuladas relativas.

Nro. de erratas	Frecuencia	Frec. relativa	Frec. acum.	Frec. rel. acum.
0	6	0.12	6	0.12
1	10	0.20	16	0.32
2	12	0.24	28	0.56
3	8	0.16	36	0.72
4	5	0.10	41	0.82
5	4	0.08	45	0.90
6	3	0.06	48	0.96
7	0	0.00	48	0.96
8	1	0.02	49	0.98
9	0	0.00	49	0.98
10	1	0.02	50	1.00
Total	50	1.00		



Síntesis

Medidas para datos AGRUPADOS

Nro. de intervalos de clase: $\boldsymbol{k = \sqrt{n}}$ **Amplitud de intervalos de clase:** $\boldsymbol{A_{IC} = \frac{R}{k}}$

Medidas de **tendencia central**

Media aritmética ($\bar{x}$)

$$\bar{x} = \sum_{i=1}^{k} \frac{x_{PM_i} fa_i}{n}$$

Mediana ($\tilde{x}$ o Me)

$$\tilde{x} = Li + \frac{\frac{n}{2} - Fa_{(k-1)}}{fa_k} . A_{IC}$$

Moda ($\hat{x}$ o Mo)

$$\hat{x} = Li + \frac{d_1}{d_1 + d_2} . A_{IC}$$

con $d_1 = fa_k - fa_{k-1}$ y $d_2 = fa_k - fa_{k+1}$

Medida de **forma**

Coeficiente de asimetría de Pearson(A_P)

$$A_P = \frac{3(\bar{x} - \tilde{x})}{s}$$

Medidas de **dispersión**

Rango (R)

$$R = x_{máx} - x_{mín}$$

Varianza (s^2)

$$s^2 = \sum_{i=1}^{n} \frac{(x_{PM_i} - \bar{x})^2 fa_i}{n - 1}$$

Desvío estándar (s)

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_{PM_i} - \bar{x})^2 fa_i}{n - 1}}$$

Rango intercuartil (RI)

$$RI = Q_3 - Q_1$$

Coeficiente de variación (CV)

$$CV = \frac{s}{\bar{x}}$$

Medidas de **posición**

Cuartiles (Q_k)

Números que dividen la distribución de lo datos en cuatro partes iguales.

$$Q_k = Li + \frac{\frac{n}{4}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1{,}2{,}3$

Deciles (D_k)

Números que dividen la distribución de lo datos en diez partes iguales.

$$D_k = Li + \frac{\frac{n}{10}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1, \dots, 9$

Percentiles (P_k)

Números que dividen la distribución de lo datos en cien partes iguales.

$$p_k = Li + \frac{\frac{n}{100}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1, \dots, 9$

Ejercicio 3

Los siguientes son los pesos, en kilogramos, y las estaturas, en centímetros, de una muestra de 14 individuos que participan de un ensayo clínico.

Pesos: 84, 99, 71, 64, 65, 80, 70, 70, 56, 66, 89, 60, 65, 79

Estaturas: 185, 180, 173, 168, 175, 183, 184, 174, 164, 169, 205, 161, 177, 174.

a. Para cada variable construya un diagrama de tallo y hojas.

Peso (kg)

Para construir el diagrama de tallo y hoja descomponemos las observaciones en decena y unidad, las vamos ordenando en diferentes filas y luego las reorganizamos dentro de cada fila.

5	6
6	4 5 6 0 5
7	1 0 0 9
8	4 0 9
9	9

5	6
6	0 4 5 5 6
7	0 0 1 9
8	0 4 9
9	9

Estatura (cm)

Para esta variable descomponemos cada medición en centena por un lado, decena y unidad por otro.

16	8 4 9 1
17	3 5 4 7 4
18	5 0 3 4
19	
20	5

16	1 4 8 9
17	3 4 4 5 7
18	0 3 4 5
19	
20	5

Ejercicio 3

Pesos: 84, 99, 71, 64, 65, 80, 70, 70, 56, 66, 89, 60, 65, 79

Estaturas: 185, 180, 173, 168, 175, 183, 184, 174, 164, 169, 205, 161, 177, 174.

b. Para cada variable calcule la media, la mediana, la desviación estándar y los cuartiles.

Peso (kg)

- **Media:** $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 72{,}714$

Interpretación: **El peso promedio de los pacientes que participaron del ensayo es de 72, 714 kg.**

- **Mediana:**

1°) Comenzamos **ordenando** los datos de menor a mayor.

56 – 60 – 64 – 65 – 65 – 66 – **70 – 70** – 71 – 79 – 80 – 84 – 89 – 99

2°) Determinamos la **posición** de la mediana: $\boldsymbol{\frac{n+1}{2}}$

Aquí $n = 14 \rightarrow \frac{n+1}{2} = \frac{14+1}{2} = 7{,}5$ por lo tanto la mediana será un promedio de los 7° y 8° datos.

3°) Buscamos el/los dato/s en la lista ordenada de observaciones.

$$\tilde{x} = \frac{x_7 + x_8}{2} = \frac{70 + 70}{2} = 70$$

Interpretación: **El 50% de los pacientes tiene pesos menores o iguales a 70 kilogramos.**

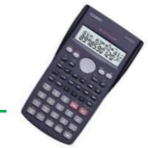
Ejercicio 3

Pesos: 84, 99, 71, 64, 65, 80, 70, 70, 56, 66, 89, 60, 65, 79
Estaturas: 185, 180, 173, 168, 175, 183, 184, 174, 164, 169, 205, 161, 177, 174.

b. Para cada variable calcule la media, la mediana, la desviación estándar y los cuartiles.

Peso (kg)

- **Desviación estándar:** $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 12{,}009 \longleftarrow$



Interpretación: **La desviación típica de los datos respecto de la media es de 12 kg, aproximadamente.**

- **Cuartiles:**

1°) Comenzamos **ordenando** los datos de menor a mayor.

56 – 60 – 64 – 65 – 65 – 66 – 70 – 70 – 71 – 79 – 80 – 84 – 89 – 99

2°) Determinamos la **posición** del cuartil k: $\boldsymbol{(n+1).\frac{k}{4}}$ siendo $k = 1{,}2{,}3$

3°) Buscamos el/los dato/s en la lista ordenada de observaciones.

Cuartil 1:

Posición: $(14+1).\frac{1}{4} = \frac{15}{4} = 3{,}75$ →Dato 4. $\therefore Q_1 = x_4 = 65$

Interpretación: **El 25% de los pacientes tiene pesos menores o iguales a 65 kg.**

Cuartil 2: Es la mediana que ya fue calculada e interpretada. $Q_2 = 70$

Cuartil 3:

Posición: $(14+1).\frac{3}{4} = \frac{45}{4} = 11{,}25$ →Dato 11. $\therefore Q_3 = x_{11} = 80$

Interpretación: **El 75% de los pacientes tiene pesos menores o iguales a 80 kg.**

Ejercicio 3

c. ¿Qué conjunto de mediciones tiene mayor variabilidad?

Para comparar variabilidad entre pesos y alturas en principio tengamos en cuenta que se trata de magnitudes diferentes, y por ende con unidades de medida diferentes, por lo tanto no sería adecuado comparar, por ejemplo, medidas de variabilidad como el rango, la desviación estándar o varianza, o el rango intercuartílico.

De todas formas, así se tratara de las mismas magnitudes, para comparar desvíos típicos entre dos conjuntos de datos, las observaciones deben tener la misma media. ¿Por qué? Porque el desvío estándar es una medida de variabilidad de los datos respecto de su media, notar que en la fórmula aparece $\bar{x}$, y si no tienen el mismo promedio la comparación de esta medida de dispersión carece de sentido.

Ante situaciones como las antes expuestas, corresponde trabajar con el **COEFICIENTE DE VARIACIÓN**.

Peso (kg)	**Estaturas (cm)**
Media: $\bar{x} = 72{,}714$	Media: $\bar{x} = 176{,}57$
Desviación estándar: $s = 12{,}009$	Desviación estándar: $s = 10{,}91$
Coeficiente de variación: $CV_{Peso} = \frac{s}{\bar{x}} = \frac{12{,}009}{72{,}714} = 0{,}165$	Coeficiente de variación: $CV_{Estatura} = \frac{s}{\bar{x}} = \frac{10{,}91}{176{,}57} = 0{,}0618$
Interpretación: **La variabilidad de los pesos de estos pacientes es 0,165 veces su media.**	Interpretación: **La variabilidad de las estaturas de estos pacientes es 0,0618 veces su media.**

CV es adimensional

Comparamos los coeficientes de variación y decidimos en cuál conjunto es mayor la variabilidad:

$$CV_{Peso} > CV_{Estatura}$$

El conjunto de mediciones del peso es el de mayor variabilidad, pues tiene un mayor coeficiente de variación.

Síntesis

Medidas para datos SIN agrupar

Medidas de **tendencia central**

Media aritmética ($\overline{x}$)

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Mediana ($\widetilde{x}$ o Me)

$$x_{\left(\frac{n+1}{2}\right)}$$

Ordenando los datos de menor a mayor.

Moda ($\widehat{x}$ o Mo)

Observación más frecuente.

Medidas de **dispersión**

Rango (R)

$$R = x_{máx} - x_{mín}$$

Varianza (S^2)

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

Desvío estándar (s)

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

Rango intercuartil (RI)

$$RI = Q_3 - Q_1$$

Coeficiente de variación (CV)

$$CV = \frac{s}{\bar{x}}$$

Medidas de **posición**

Ordenando los datos de menor a mayor.

Cuartiles $(\boldsymbol{Q_k})$

Números que dividen la distribución de lo datos en cuatro partes iguales.

Posición: $(n+1).\frac{k}{4}$

con $k = 1,2,3$

Deciles $(\boldsymbol{D_k})$

Números que dividen la distribución de lo datos en diez partes iguales.

Posición: $(n+1).\frac{k}{10}$

con $k = 1, \dots, 9$

Percentiles $(\boldsymbol{P_k})$

Números que dividen la distribución de lo datos en cien partes iguales.

Posición: $(n+1).\frac{k}{100}$

con $k = 1, \dots, 99$

Todas las medidas se expresan en la misma unidad que la variable estudiada, excepto la varianza (unidades al cuadrado) y el coeficiente de variación (carece de unidades).

Síntesis

Comparación de variabilidad en dos o más conjuntos de datos

¿Se trata de la misma magnitud?

Sí

No

¿Tienen las mismas unidades?

Coeficiente de variación

Sí

No

¿Tienen la misma media?

Coeficiente de variación

Sí

No

Varianza, desvío estándar

Coeficiente de variación, rango, rango intercuartil

Ejercicio 4

Variable: Consumo mensual de datos móviles usando red Wi-Fi en Argentina (Gb) en 2017

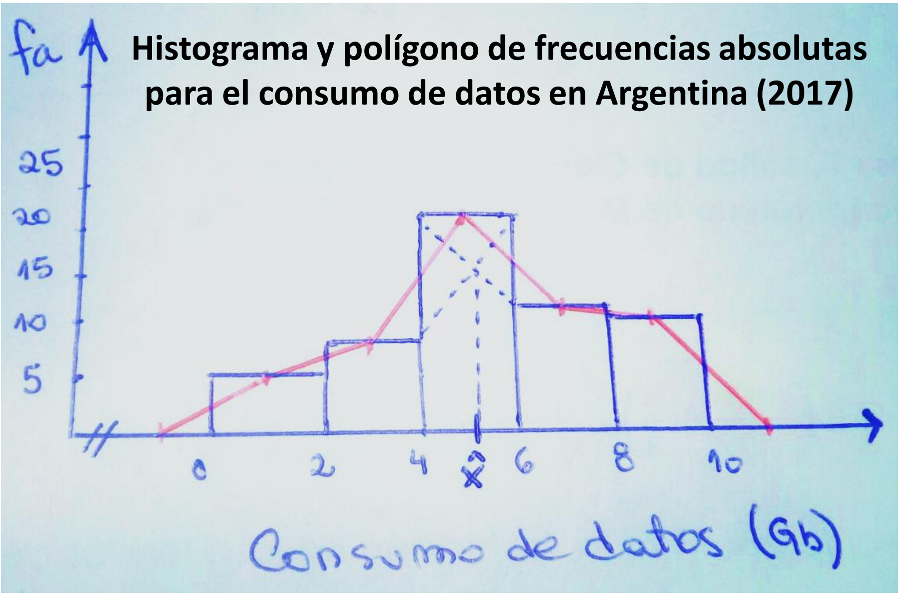
Es cuantitativa CONTINUA.

a) Completamos la tabla de distribución de frecuencias con los datos faltantes.

La frecuencia absoluta de la segunda clase la obtenemos haciendo la diferencia entre el total de datos y las otras frecuencias dadas.

Inter. de clase	X_{PM}	fa	Fa	fr	Fr
[0, 2)	**1**	6	**6**	**0,10**	**0,10**
[2, 4)	**3**	**9**	**15**	**0,15**	**0,25**
[4, 6)	**5**	23	**38**	**0,38**	**0,63**
[6 ,8)	**7**	12	**50**	**0,20**	**0,83**
[8 , 10]	**9**	10	**60**	**0,17**	**1**

b)



Moda estimada gráficamente: $\hat{x} \cong \mathbf{5}$

c) El consumo promedio mensual de datos móviles en Argentina es 5,37 Gb.

$$\bar{x} = \sum_{i=1}^{k} \frac{x_{PM_i} f a_i}{n} = 5{,}37$$

Desvío estándar: $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 2{,}37$

Ejercicio 4

d) **Mediana**:

Buscamos la primera clase cuya frecuencia relativa acumulada es mayor o igual a 0,50.

Si $\tilde{x} \in [4{,}6)$, entonces:

$$\tilde{x} = 4 + \frac{\frac{60}{2} - 15}{23} .2 = 5{,}3$$

Interpretación: **La mitad de los argentinos consumió menos de 5,3 Gb de datos móviles en 2017.**

e) Como ambos conjuntos de datos tienen diferente media comparamos los coeficientes de variación.

Consumo de datos en Uruguay (Gb)

Media: $\overline{x_U} = 8{,}4$

Desvío estándar: $s_U = 3$

Coeficiente de variación:

$$CV_U = \frac{s_U}{\overline{x_U}} = \frac{3}{8{,}4} = 0{,}357$$

Consumo de datos en Argentina (Gb)

Media: $\overline{x_A} = 5{,}37$

Desvío estándar: $s_A = 2{,}37$

Coeficiente de variación:

$$CV_A = \frac{s_A}{\overline{x_A}} = \frac{2{,}37}{5{,}37} = 0{,}44$$

Como $CV_A > CV_U$=> **Los datos de Argentina presentan mayor variabilidad que los de Uruguay.**

Ejercicio 5

Variable: Nro. de automóviles que llegan a un puesto de control en intervalos de 10 minutos.
Es cuantitativa DISCRETA.

a) Con la siguiente tabla se pueden construir los gráficos pedidos.

Nro. de autos por intervalo de tiempo	Frecuencia	Frecuencia relativa	Frecuencia acumulada
0	20	0.1888	20
1	40	0.3773	60
2	30	0.2830	90
3	10	0.0943	100
4	5	0.0472	105
5	1	0.0094	106
Total	106	1.00	

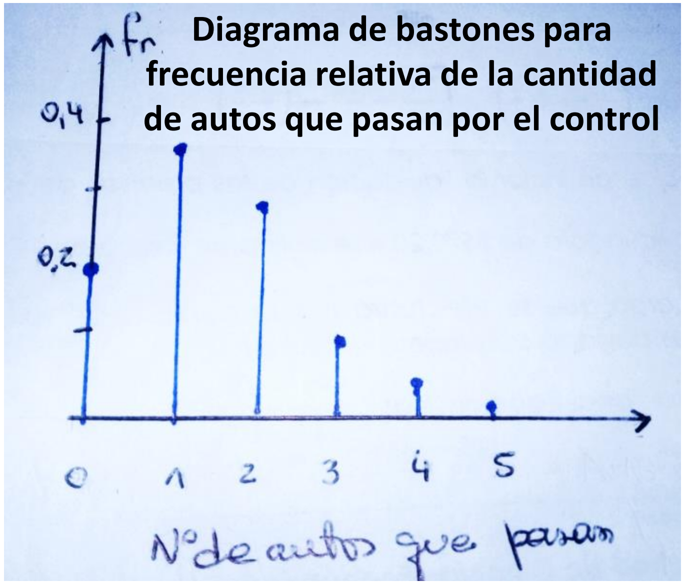


Diagrama de bastones para frecuencia relativa de la cantidad de autos que pasan por el control

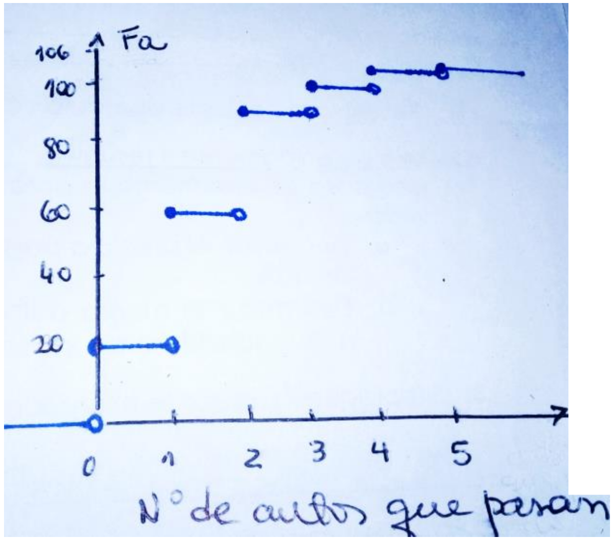


Gráfico escalonado para frecuencia absoluta de la cantidad de autos que pasan por el control

Ejercicio 5

b) **Media:** $\bar{x} = \sum_{i=1}^{k} \frac{x_{PM_i} fa_i}{n} = 1{,}46$ 

Interpretación: **En promedio, el número de autos que pasan por un puesto de control en los distintos intervalos de tiempo es de 1,46 autos.**

Mediana: Determinamos la **posición** de la mediana: $\frac{n+1}{2}$

Aquí $n = 106 \rightarrow \frac{n+1}{2} = \frac{106+1}{2} = 53{,}5$ →la mediana será un promedio de los 53° y 54° datos.

Buscamos el/los dato/s en la lista ordenada de observaciones.

Nro. de autos por intervalo de tiempo	Frecuencia	Frecuencia relativa	Frecuencia acumulada
0	20	0.1888	20
1	40	0.3773	60
2	30	0.2830	90
3	10	0.0943	100
4	5	0.0472	105
5	1	0.0094	106
Total	106	1.00	

$$\tilde{x} = \frac{x_{(53)} + x_{(54)}}{2} = \frac{1+1}{2} = 1$$

Interpretación: **En la mitad de los intervalos se contabilizaron 1 auto o menos que pasó por el control.**

Moda: Se obtiene mirando el gráfico de bastones o la clase con mayor frecuencia absoluta.

$$\hat{x} = 1$$

Interpretación: **El número de autos más frecuentes contabilizados en los intervalos de tiempo es 1.**

Ejercicio 5

c) **Varianza:** $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = 1{,}1056^2 = 1{,}2223$ 

Interpretación: **Recordar que se mide en unidades al cuadrado, conviene interpretar su raíz cuadrada que es el desvío estándar**

Desvío estándar: $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 1{,}1056$ 

Interpretación: **Un desvío típico del número promedio de autos contabilizados en los distintos intervalos de tiempo es igual a 1,106 autos (desvío estándar).**

Coeficiente de variación:

$$CV = \frac{s}{\bar{x}} = \frac{1{,}106}{1{,}4622} = 0{,}7561$$

Interpretación: **La dispersión de los datos es 0,75 veces la media.**

d) **Coeficiente asimetría de Pearson:**

$$A_p = \frac{3(1{,}4622 - 1)}{1{,}1056} = 1{,}2541 > 0$$

Significa que la distribución de los datos presenta asimetría hacia la derecha o positiva.

Ejercicio 6

a) Variable: Nro. de virus detectados con el software en PC de domicilios particulares.

- **Media:** $\overline{x_I} = \sum_{i=1}^{n} \frac{x_i}{n} = 46{,}33$
- **Mediana:**

Datos ordenados: 29 – 35 – 37 – 46 – **47** – 53 – 54 – 55 – 61

Aquí $n = 9 \rightarrow \frac{n+1}{2} = \frac{10}{2} = 5 \rightarrow \tilde{x} = x_{(5)} = 47$

- **Rango:** $R = x_{máx} - x_{mín} = 61 - 27 = 32$
- **Desvío estándar:** $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 10{,}665$

b) Por debajo del **tercer cuartil** $\boldsymbol{Q_3}$ encontramos el 75% de los datos.
Por lo tanto, calculamos la posición del mismo y luego lo determinamos en los datos ordenados.

Aquí $n = 9 \rightarrow (n+1).\frac{3}{4} = 7{,}5 \rightarrow$Hacemos un promedio de los datos 7° y 8°.

$$\rightarrow Q_3 = \frac{x_{(7)} + x_{(8)}}{2} = \frac{54+55}{2} = 54{,}5$$

En 75% de los ordenadores analizados con este software se detectaron menos de 54,5 virus.

Ejercicio 7

Para determinar cuál tienda es más consistente en sus ventas, debemos comparar la variabilidad de los conjuntos de datos.
Como tiene la misma media, podemos comparar los desvíos estándares:

Variable: Monto diario de venta en la tienda A.

Desvío estándar: $s_A = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 790{,}57$

Variable: Monto diario de venta en la tienda B.

Desvío estándar: $s_B = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 7280{,}11$

Vemos que $s_A < s_B$, por lo **cual los montos diarios de la tienda A presentan menor dispersión y por lo tanto este local es más consistente en sus ventas**.

Ejercicio 8

a) **Calificación de estudiantes en prueba inicial**

- **Media:** $\overline{x_I} = \sum_{i=1}^{n} \frac{x_i}{n} = 2{,}7$
- **Mediana:**

Datos ordenados: 1 – 1 – 1 – 2 – **2 – 3** – 3 – 4 – 5 - 5

Aquí $n = 10 \rightarrow \frac{n+1}{2} = \frac{11}{2} = 5{,}5 \rightarrow \widetilde{x_I} = \frac{x_{(5)}+x_{(6)}}{2} = \frac{2+3}{2} = 2{,}5$

- **Moda:**

Nota más frecuente. $\widehat{x_I} = 1$

- **Desvío estándar:** $\mathrm{s_I} = \sqrt{\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}} = 1{,}567$

- **Varianza:** $\mathrm{s_I^2} = 1{,}567^2 = 2{,}455$

Calificación de estudiantes en prueba final

- **Media:** $\overline{x_F} = \sum_{i=1}^{n} \frac{x_i}{n} = 6{,}3$
- **Mediana:**

Datos ordenados: 3 – 4 – 5 – 6 – **6 – 6** – 7 – 8 – 9 – 9

Aquí $n = 10 \rightarrow \frac{n+1}{2} = \frac{11}{2} = 5{,}5 \rightarrow \widetilde{x_F} = \frac{x_{(5)}+x_{(6)}}{2} = \frac{6+6}{2} = 6$

- **Moda:**

Nota más frecuente. $\widehat{x_F} = 6$

- **Desvío estándar:** $\mathrm{s_F} = \sqrt{\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}} = 2{,}003$
- **Varianza:** $\mathrm{s_F^2} = 2{,}003^2 = 4{,}012$

Ejercicio 8

b) Como los datos tienen diferente media, para averiguar cuál conjunto de datos es más disperso calculamos los coeficientes de variación:

Para las notas de la prueba inicial: $CV_I = \frac{s_I}{\overline{x_I}} = \frac{1,567}{2,7} = 0,581$

Para las notas de la prueba final: $CV_F = \frac{s_F}{\overline{x_F}} = \frac{2,003}{6,3} = 0,318$

Como $CV_I > CV_F$ concluimos que **el conjunto de calificaciones de la prueba inicial presenta mayor dispersión**.

c) Definimos: **Incremento de la calificación obtenida = $\boldsymbol{x_F - x_I}$**

Resultando: **2 – 3 – 4 - 4 – 1 – 3 – 5 – 5 – 3 – 6**

- **Media:** $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 3,6$
- **Mediana:**

Datos ordenados: 1 – 2 – 3 – 3 **– 3 – 4** – 4 – 5 – 5 – 6

Aquí $n = 10 \rightarrow \frac{n+1}{2} = \frac{11}{2} = 5,5 \rightarrow \tilde{x} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{3+4}{2} = 3,5$

- **Desvío estándar:** $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = 1,5055$