

Problemas

Conocer una variable

Vicente Manzano-Arrondo, 2010-2014

Problemas de cálculo

Ejercicio resuelto completo

En 50 ocasiones hemos preguntado a Asensio por su número favorito. En contra de lo que cabe esperar, Asensio ha respondido en cada ocasión con un número diferente. Sus respuestas son las que siguen:

13	17	19	19	13	19	15	13	15	13
11	19	11	11	19	15	11	15	15	19
13	15	17	11	15	11	17	19	17	13
15	19	19	17	17	17	13	15	13	15
11	13	17	15	13	15	15	11	13	17

Con estos datos, construye una tabla de frecuencia al uso y calcula los estadísticos moda, mediana y media aritmética con sus correspondientes índices de bondad de representación o de dispersión.

Solución

La tabla de frecuencia es la que sigue, con las columnas de valores, frecuencias, frecuencias acumuladas, porcentajes y porcentajes acumulados, respectivamente. He añadido un par de columnas más para los cálculos intermedios (cada valor por su frecuencia, para el cálculo de la media aritmética; y cada distancia cuadrática del valor a la media por su frecuencia, para el cálculo de la desviación tipo). Y otra pequeña tabla más a la derecha, para el cálculo de la mediana de las distancias a la mediana. Con todo ello:

Xi	fi	Fi	%i	%ai
11	8	8	16	16
13	11	19	22	38
15	13	32	26	64
17	9	41	18	82
19	9	50	18	100
Σ	50		100	

Xifi	d2fi
88	128,0
143	44,0
195	0,0
153	36,0
171	144,0
750	352,0

Xi-Mdn	fi	Fi
0	13	13
2	20	33
4	17	50
Σ	50	

Para el cálculo de los estadísticos:

- Moda y bondad de la moda (% de la moda)
Según puede observarse en la tabla, el valor con mayor frecuencia es $X_i=15$. Luego, Moda = 15 y % Moda = 26.
- Mediana y bondad de la mediana (MAD)
Como $n=50$, la mediana ocupa la posición $51/2=25,5$. Desde la posición 20 hasta la 32, ambas inclusive, todos los datos ordenados tienen el valor 15. Luego, Mdn = 15.

Para calcular MAD (mediana de distancias a la mediana), hay que calcular las distancias a la mediana primeramente. Esta información se encuentra en la pequeña tabla de la derecha. Los 13 datos con valor 15 coinciden con Mdn, por lo que su distancia es 0. Los 11 datos con valor 13 y los 9 datos con valor 17 se distancian de la mediana en 2 unidades, luego la frecuencia para la distancia 2 es de $11+9 = 20$. Por último, los 8 datos con valor 11 y los 9 datos con valor 19 se distancian de la mediana en 4 unidades, luego la frecuencia de la distancia 4 es 17. Observando las frecuencias acumuladas, la posición 25,5 se encuentra en el intervalo de las posiciones 14 a 33, ocupadas por el valor de distancia 2. Luego, la mediana de las distancias a la mediana es MAD = 2.

- Media aritmética y bondad de la media (desviación tipo)

Para interpretar más fácilmente la dispersión, acudimos también al coeficiente de variación de Pearson, según el cual, la desviación tipo es un 18% de la media, una dispersión muy aceptable y, por tanto, un apoyo a la media aritmética como medida de representación numérica del conjunto.

$$\bar{X} = \frac{\sum X_i f_i}{n} = \frac{750}{50} = 15$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{352}{50}} = 2,65$$

$$CV = \frac{S}{\bar{X}} 100 = 18\%$$

$$S^a = \frac{100 S}{\sqrt{(\bar{X} - \min)(\max - \bar{X})}} =$$

$$= \frac{100 \cdot 2,65}{\sqrt{(15 - 11)(19 - 15)}} = 66\%$$

Ejercicios propuestos

1. Hemos preguntado a 40 estudiantes por cuántos euros (en términos redondos) suelen gastar cada día en alimentación en la universidad. Las respuestas son:

4	9	2	9	4	1	6	4	6	4
2	1	2	6	1	6	2	6	6	1
4	6	9	2	6	2	9	4	9	4
6	9	1	9	9	9	4	6	4	6

Construye la tabla de frecuencias y calcula los mismos estadísticos que en problema anterior.

2. Un grupo de niños y niñas de 8 años se han enfrentado a una prueba en la clase de educación física. Tenían que dar cuatro patadas a un balón sin que tocara el suelo. Hay quien lo ha conseguido y hay quien no. Lo que sigue es el número de patadas seguidas que han conseguido dar. Resuelve del mismo modo que en los ejercicios anteriores.

3	1	3	1	1	1	2	2	1
2	1	1	1	1	2	2	4	2
1	2	2	4	3	3	3	2	1
2	2	1	4	1	2	4	1	3

Soluciones sintéticas

Ejercicio propuesto 1

X_i	f_i	F_i	$\%i$	$\%ai$
1	5	5	12,5	12,5
2	6	11	15,0	27,5
4	9	20	22,5	50,0
6	11	31	27,5	77,5
9	9	40	22,5	100,0
Σ	40		100	

$X_i f_i$	$d^2 f_i$
5	80,0
12	54,0
36	9,0
66	11,0
81	144,0
200	298,0

$ X_i - \text{Mdn} $	f_i	F_i
1	20	20
3	6	26
4	14	40
Σ	40	

n	40
Media	5,00
D.t.	2,73
Sa	68
CV	55

Mdn	5
MAD	2
Moda	6
%Moda	27,5

Ejercicio propuesto 2

X_i	f_i	F_i	$\%i$	$\%ai$
1	14	14	39	39
2	12	26	33	72
3	6	32	17	89
4	4	36	11	100
Σ	36		100	

$X_i f_i$	$d2f_i$
14	14,0
24	0,0
18	6,0
16	16,0
72	36,0

$ X_i - Mdn $	f_i	F_i
0	12	12
1	20	32
2	4	36
Σ	36	

n	36
Media	2,00
D.t.	1,00
Sa	71
CV	50

Mdn	2
MAD	1
Moda	1
%Moda	28

Ejercicios más conceptuales

1. Observa la siguiente tabla de frecuencias:

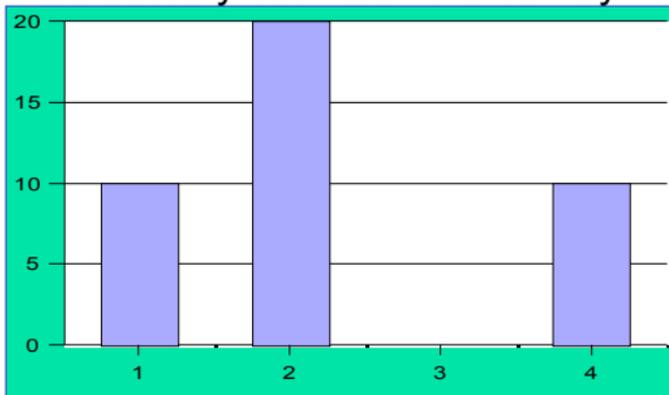
X_i	5	7	11	12	15	19	20	21	25	26	28	32	35	38	44	52	61
f_i	124	53	264	15	23	62	54	85	65	159	86	92	78	56	84	121	38

En relación a la tabla, responde a las siguientes preguntas sin realizar cálculos:

- ¿Cuál es el valor con el mayor porcentaje?
- ¿Y el que posee el menor?
- ¿Cuál es el valor con el porcentaje acumulado más alto?
- ¿Y el que posee el más bajo?

2. Observa la siguiente representación gráfica.

Sabiendo que se trata de una variable nominal y que el eje vertical se refiere a las frecuencias de cada valor, reproduce la tabla de frecuencias que le corresponde. Indica cuántos datos y cuántos valores hay.



3. Hemos preguntado a un conjunto de estudiantes de psicología en qué provincia han nacido. Han respondido 120 en Sevilla, 68 en Córdoba, 72 en Cádiz, 60 en Huelva, 36 en Badajoz y 44 en otros lugares. Construye la tabla de frecuencias más idónea y escoge la mejor representación numérica.

4. La siguiente tabla muestra información sobre la variable “visitas al supermercado durante el pasado mes”. Completa la tabla con la columna de frecuencias.

X_i	2	5	8	9	10	14	16	17	21	24	27
F_i	6	8	11	22	31	42	50	51	57	62	70

5. Identifica los errores en las siguientes tablas:

X_i	f_i	X_i	F_i	X_i	$\%a_i$	X_i	$\%i$	X_i	$\%i$
2	1	1	1	1	2	1	2	1	2
4	4	2	3	2	7	2	7	2	7
6	2	4	7	4	11	4	-8	4	11
5	6	5	9	5	23	5	13	5	115
9	5	7	13	7	25	7	42	7	47
11	3	8	12	8	30	8	-1	8	30
15	2	9	22	9	39	9	11	9	15
17	1	10	27	10	58	10	6	10	6

6. Hemos proyectado 10 películas en un ciclo de cine, ante un grupo de 40 personas. Les hemos preguntado

por cuáles les han gustado más y las películas escogidas han sido las numeradas como 7, 3, 5, 2 y 9.

Construye la tabla de frecuencias sabiendo que:

- El conjunto de datos tiene dos modas.
- El 10% ha escogido la película 3.
- La película 5 se ha llevado la quinta parte de las elecciones.
- 4 personas escogieron la película 7.

7. Tenemos el siguiente conjunto de datos: 2, 1, 9, 1, 12, 5, 1, 2, 3. Añade un dato más, de tal forma que consigas que:

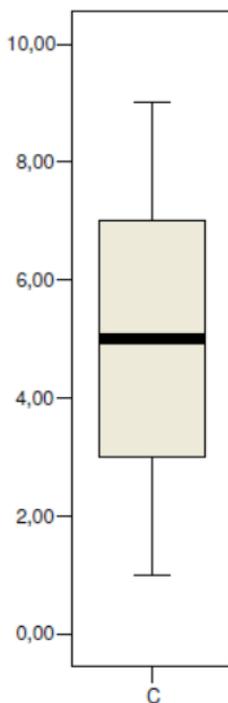
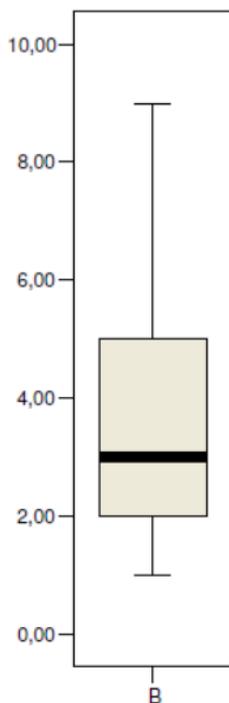
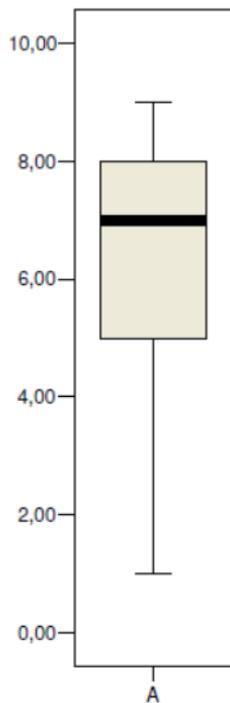
- A. la distribución pase a ser bimodal.
- B. o bien, que la media aritmética no varíe.
- C. o bien que la mediana no varíe.

8. A los siguientes datos, $\{1, 2, 3, 4, 10, 10\}$ ¿Qué valor hemos de añadir para conseguir que el valor de la media aritmética sea el doble que la mediana?

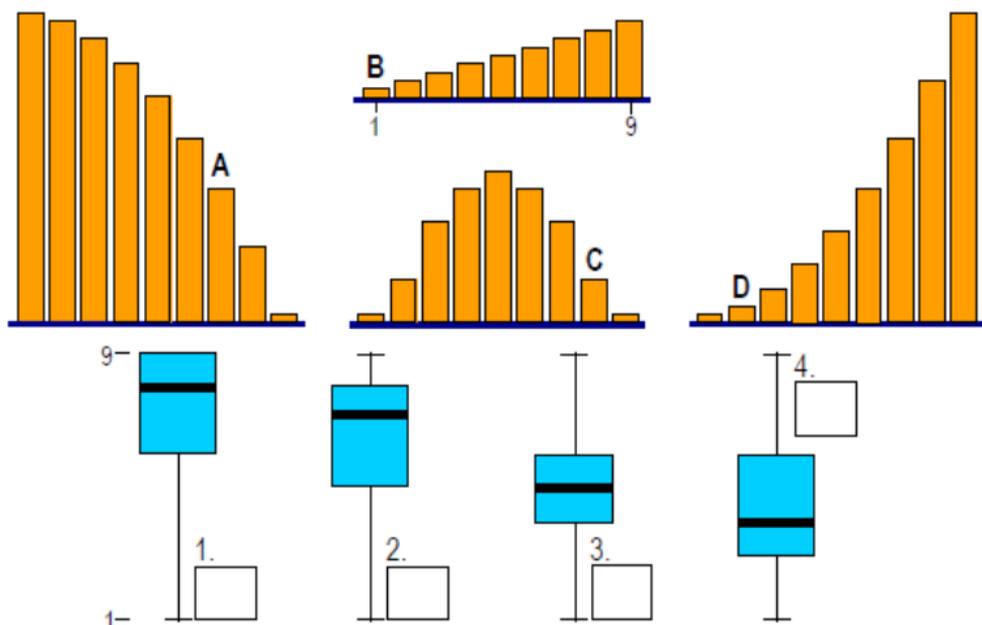
Ejercicios de exámenes

6. A continuación observarás tres representaciones gráficas. Señala cuál de ellas se corresponde con la distribución de datos siguiente. Justifica tu respuesta. (1 punto)

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



4. Para cada uno de cuatro conjuntos de datos hemos realizado un diagrama de barras y uno de caja y patillas. Reconstruye las parejas, indicando en la casilla de cada box-plot la letra del diagrama de barras con el que se corresponde. {4 puntos}



1. Rodea el número de las columnas que escogerías para construir la tabla de frecuencias de la variable "Nombre del mejor político nacional actual". {6 puntos}

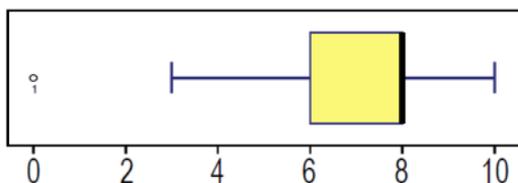
- | | | |
|-----------------|----------------------------|----------------------------|
| 1. Valores. | 2. Etiquetas de valor. | 3. Frecuencias absolutas. |
| 4. Porcentajes. | 5. Frecuencias acumuladas. | 6. Porcentajes acumulados. |

3. En un estudio se han abordado las siguientes variables: *actitud frente al matrimonio* (en una escala de 0 a 13); *adscripción política* (izquierdas; centro; derechas), *edad* (en años), *DVD en casa* (si; no) y *libro leído preferido*. Indica

<i>Variable</i>	<i>Moda</i>	<i>Mediana</i>	<i>Media</i>
Actitud matrimonio			
Adscripción política			
Edad			
DVD en casa			
Libro preferido			

(marcando con una cruz), cuál es el mejor índice de posición del conjunto de datos para cada una de las variables. {5 puntos}

5. Anota en las casillas correspondientes el valor de los tres cuartiles y de los valores mínimo y máximo provenientes de la distribución de datos que ha sido representada en el diagrama de caja y patillas. {5 puntos}

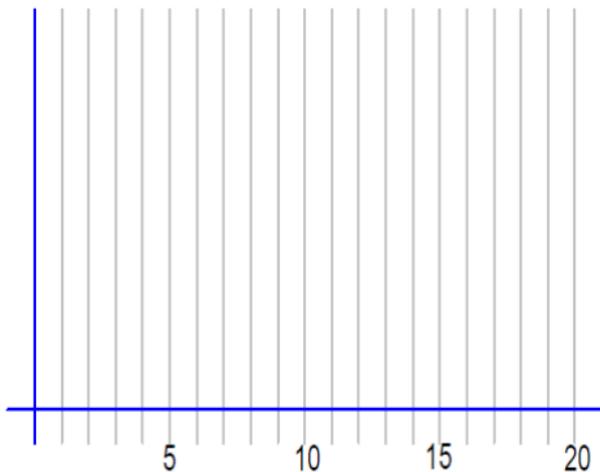


Mín.	Máx.	Q ₁	Q ₂	Q ₃

3. Hemos obtenido el valor de unos cuantos percentiles en una variable cuantitativa. Con la información que te suministran esas medidas realiza las siguientes dos acciones:

$$P_1 = 3; P_{10} = 3; P_{25} = 4; P_{50} = 6; P_{75} = 9; P_{90} = 14; P_{98} = 14; P_{99} = 19$$

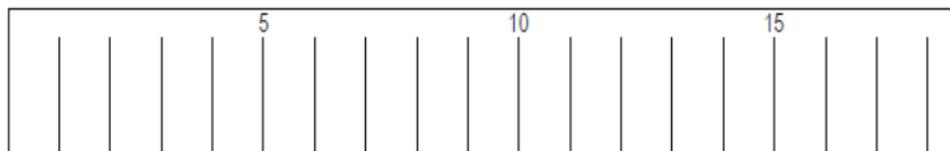
A) Construye un diagrama horizontal de caja y patillas en la siguiente gráfica vacía. (6 puntos)



B) Escribe los valores que definen el 50% central de los datos (2 puntos):

de a

4. En el siguiente recuadro dibuja un diagrama de caja y patillas sabiendo que el punto de valor 7 deja al 50% de los datos por arriba y al otro 50% por debajo; que el 50% central se encuentra entre 6 y 10; que la práctica totalidad de los datos se hallan entre 5 y 14; y que hay un dato con un valor raro, el 17. {6 puntos}

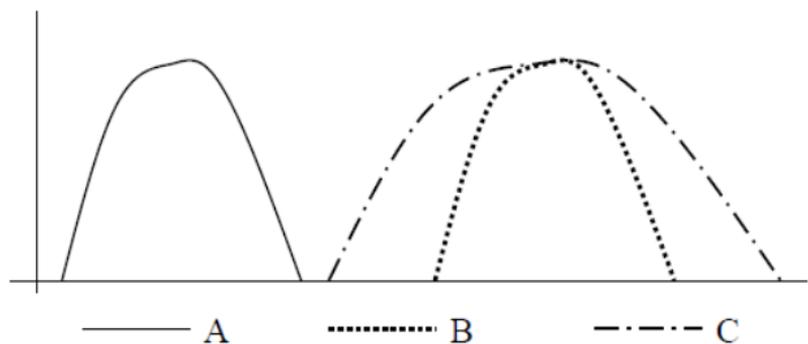


3. En un estudio se han abordado las siguientes variables: *porcentaje de grasa corporal*; *cercanía al centro de trabajo* (cerca; distancia media; lejos), *peso* (en gramos), *procedencia* (autóctono; extranjero), *título de la última película vista en el cine* y *número*

<i>Variable</i>	<i>MAD</i>	<i>D.T.</i>	<i>C.V.</i>
% grasa corporal			
Cercanía trabajo			
Peso			
Procedencia			
Última película			
Número de orden			

de orden que ocupa en el listado de hermanos. Indica (marcando con una cruz), cuál es el mejor índice de dispersión del conjunto de datos para cada una de las variables: desviación mediana de las distancias a la mediana (MAD), desviación tipo (DT) o coeficiente de variación de Pearson (CV). {6 puntos}

3. Observa la siguiente representación gráfica, donde se encuentran tres distribuciones de datos (A, B y C). A continuación se encuentra un listado con cuatro parejas (1, 2, 3 y 4) de índices descriptivos (media y desviación tipo). Asigna a cada una de las tres distribuciones su correspondiente par de índices. (1,5 puntos)



Par	1	2	3	4
Media	6	6	19	19
Desviación Tipo	2	4	2	4