

# ESTADÍSTICA Y ANÁLISIS DE DATOS

## Práctica del Tema 1. Variables estadísticas unidimensionales

### Problemas

1. Se ha contabilizado el número de días que durante un año han faltado al trabajo, por baja laboral, los trabajadores de una empresa, obteniéndose los siguientes resultados:

Número de días	1	2	3	4	5	6	7	8
Número de obreros	2	7	10	4	10	3	2	2

Se pide:

- Representar mediante el gráfico apropiado la distribución de los trabajadores según los días de baja.
  - Calcular la media, moda y cuartiles de esa distribución.
  - ¿Es el número medio de días de baja representativo de los valores de la distribución? Responde a esta pregunta utilizando el estadístico adecuado.
  - A la hora de recoger la información para elaborar la tabla del enunciado ha habido un error: 6 trabajadores no habían sido incluidos. Todos ellos estuvieron de baja un número de días igual a la media que has obtenido en el apartado b). ¿Cómo afectará la inclusión de estos 6 trabajadores al número medio de días de baja y al grado de representatividad de esta media, en comparación con los resultados obtenidos antes de incluirlos? Razona tu respuesta.
2. La siguiente tabla muestra la distribución de frecuencias de los salarios mensuales de los 80 trabajadores de una empresa:<sup>1</sup>

<i>Salarios (euro)</i>	<i>Trabajadores (%)</i>
[1500 – 1560)	15
[1560 – 1600)	15
[1600 – 1680)	25
[1680 – 1740)	20
[1740 – 1800)	15
[1800 – 1860)	10

- Elabora una tabla que muestre la distribución de los salarios en términos de frecuencias absolutas y relativas.
- Realiza un histograma de la distribución de los salarios.
- Calcula la media y mediana de la distribución. ¿Cuál es el intervalo de salarios más frecuente?

---

<sup>1</sup>Ejercicio adaptado de Agirre (2010) *Estatistikaren Oinarriak. Ariketak*. 2. argitalpena, Udako Euskal Unibertsitatea (UEU).

- d) A partir del histograma y utilizando los valores típicos pertinentes, comenta la forma que tiene la distribución.
3. Una variable estadística discreta toma los siguientes valores:  $x_1 = 1$ ,  $x_2 = 4$ ,  $x_3 = 7$ ,  $x_4 = 10$ . Sabemos que  $f_1 = f_4$  y que  $g_1(X) = 0$ .
- a) Obtener el valor de la media ( $\bar{x}$ ). Razonar la respuesta.
- b) Conociendo además que  $a_2(X) = 39,7$ , calcular los valores  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$  y realizar la representación gráfica correspondiente.
- c) Considerando la transformación  $Y = \frac{X-4}{3}$ , obtener los siguientes estadísticos:  $\bar{y}$ ,  $S_y^2$  y  $g_0(Y)$ . Comentar los resultados obtenidos.
4. Supongamos que la siguiente tabla muestra la distribución de la renta mensual familiar en miles de euros de un determinado colectivo:

$X$	[0, 1)	[1, 3)	[3, 4)	[4, 6)	[6, 7)
$f_i$	0,1	0,3	$f_3$	$f_4$	$f_5$

Se pide:

- a) Calcular  $f_3$ ,  $f_4$  y  $f_5$  sabiendo que  $\bar{x} = 3,65$  y  $M_e = 3,50$ .
- b) Calcular  $f_3$ ,  $f_4$  y  $f_5$  sabiendo que el coeficiente de asimetría es  $g_1 = 0$  ¿Cuál es el valor en este caso de  $\bar{x}$  y  $M_e$ ?
- c) Representar gráficamente las distribuciones de frecuencias de los dos casos anteriores, comentando ambas gráficas.
5. Una variable discreta toma valores  $x_1 < x_2 < x_3 < x_4 < x_5$  con frecuencias  $n_1 = n_4 = n_5 = 1$ ,  $n_2 = 2$ ,  $n_3 = 3$ . Sabemos además que  $x_i - x_{i-1} = 2$  para todo  $i$ .
- a) Representar gráficamente esta distribución.
- b) Expresar  $\bar{x}$ ,  $M_e$  y  $M_o$  en función de  $x_1$ .
- c) Hallar la desviación típica ( $S_x$ ) y el recorrido ( $R$ ).
- d) Explicar por qué la media, la mediana y la moda dependen de  $x_1$  y, sin embargo, la desviación típica y el recorrido no.
6. Del Boletín Estadístico del Banco de España hemos extraído datos mensuales del rendimiento anual de las Letras del Tesoro,  $X$ , y del índice IBEX35,  $Y$ , desde enero de 1988 hasta enero de 2008. Con los datos de estas dos variables hemos obtenido los siguientes valores típicos:

Variable	Valores típicos							
	Media	Mediana	Moda	$S^2$	$g_2$	$g_1$	Min	Max
Letras del Tesoro ( $X$ )	6,84	4,93	14,49	17,68	-1,25	0,51	1,84	14,52
IBEX35 ( $Y$ )	11,48	12,26	—	552,54	-0,02	0,35	-38,21	88,21

- a) Completa la frase:  
 “El valor tal que deja a la izquierda y a la derecha el mismo número de valores de  $X$  es .....
- b) ¿Cuál es el coeficiente de variación de  $Y$ ?
- c) ¿Cuál de las dos medias es más representativa?
- d) En un determinado mes el rendimiento de las Letras fue 11 y el del IBEX35 19, ¿cuál de estos dos valores es más grande en términos relativos, es decir, destaca más en relación al resto de observaciones?
7. En la Figura 1 puedes ver los diagramas de caja que muestran la distribución de la altura ( $X$ ), y la altura del padre ( $Y$ ) de 37 individuos. Responde a las siguientes preguntas:

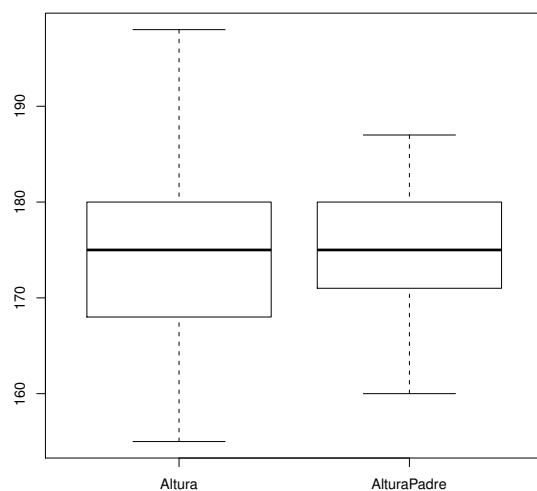


Figura 1: Diagramas de caja de las variables altura y altura del padre.

- a) Teniendo en cuenta a padres e hijos, ¿cuál es, aproximadamente, la altura de la persona más baja? ¿y la de la más alta?
- b) ¿Cuál es la altura tal que el 25% de los hijos es más bajo o igual que ella y el 75% restante es más alto o igual?
- c) ¿Cuál es la altura tal que el 75% de los hijos es más bajo o igual que ella y el 25% restante es más alto o igual?
- d) Utiliza el diagrama de caja para comparar las distribuciones de la altura y la altura del padre de los 37 individuos.
- e) ¿Hay algún individuo atípicamente alto o atípicamente bajo? Justifica tu respuesta.

8. Un hotel tiene 5 tipos de habitaciones. La tabla adjunta recoge para cada uno de ellos el precio por habitación y los ingresos totales obtenidos en un día en que se encuentra el hotel completamente ocupado.

Tipo	Precio por habitación	Ingresos totales
Primero	30	2400
Segundo	50	2000
Tercero	75	3000
Cuarto	100	2000
Quinto	$x$	2600

Se pide:

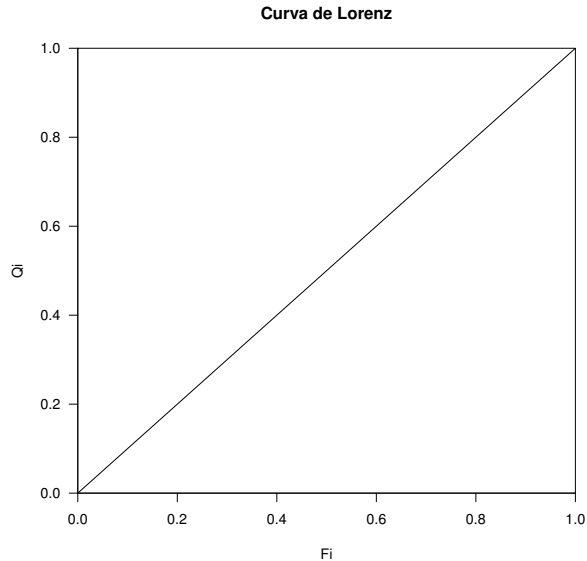
- a) Determinar cuál ha de ser el precio de la habitación del tipo superior,  $x$ , para que el precio medio por habitación en este hotel sea 60 €.
- b) ¿Cuál será entonces el valor del índice de Gini?
9. Del conjunto de perceptores de subvenciones de un Departamento Público de Promoción de Actividades Industriales, el 40 % ha percibido una subvención entre medio millón y millón y medio; otro 40 % de los perceptores la ha percibido comprendida entre un millón y medio y tres millones y medio; y el 20 % restante la ha percibido comprendida entre tres millones y medio y siete millones y medio.

Se pide:

- a) Dibujar el histograma que describa la distribución de los perceptores según la subvención recibida y estimar la moda que la variable “subvención recibida” presenta en ese colectivo.
- b) Dibujar la curva de Lorenz que describa la concentración del total de las subvenciones en su distribución en el conjunto de empresas perceptoras y calcular el índice de Gini que la mida, tomando como marcas de clase los puntos medios de los intervalos de las subvenciones.
10. En la siguiente tabla aparecen los porcentajes acumulados de la renta ganada por el 20 % más pobre, el 40 % más pobre, etc. de la población en tres países diferentes, para los años 1997-2000:

País	Porcentaje acumulado de población				
	20 %	40 %	60 %	80 %	100 %
Dinamarca	8,3	23,0	41,2	64,1	100,0
España	7,0	19,1	35,5	58,0	100,0
Sudáfrica	3,5	9,8	19,8	37,8	100,0

- a) Dibuja en el siguiente gráfico las correspondientes curvas de Lorenz.



- b) Ordena de menor a mayor los índices de concentración de las distintas curvas.  
¿Qué indica dicha clasificación en cuanto a la equidad de la distribución de la renta?
- c) Comenta la frase:  
“La renta total de Dinamarca es mayor que la renta total de España, la cual es además mayor que la renta total de Sudáfrica”.

11. En la empresa Ehunaga, S.A.L., de Yurre, existe la siguiente distribución de ingresos mensuales en €:

Nivel	Ingreso	nº de trabajadores
I	1000	60
II	2000	30
III	4000	10

Se ha tomado el acuerdo en esta empresa de que, en adelante, toda subida de remuneración será lineal, es decir, todos los niveles recibirán el mismo incremento del salario en €.

- a) ¿Cuánto habrá de aumentar cada ingreso mensual para que el coeficiente de variación  $g_0$  de la distribución de los trabajadores según sus salarios se reduzca a la mitad?
- b) Calcula el índice de Gini de la distribución de salarios antes de la subida. Razona cómo variará el valor de este índice tras el aumento de sueldo calculado en el apartado a).
- c) Razona qué ocurre con el índice de Gini si se multiplica cada valor de la variable por una constante  $k$ .
12. Un abuelo decide repartir los 25 millones a los que asciende su fortuna entre sus 10 nietos de la siguiente forma: un 30 % de su fortuna entre el 20 % de sus nietos, un 15 % entre el 30 %, un 25 % entre el 10 % y el 30 % que le queda entre el resto de los nietos. (Dentro de cada clase del reparto todos los nietos reciben lo mismo).

- a) Calcular la cantidad media percibida así como la cantidad que ha recibido un mayor número de nietos.
- b) Estudiar la dispersión de esta distribución.
- c) Estudiar el grado de concentración, tanto de forma gráfica como numérica.
- d) Comentar los resultados obtenidos.

## Cuestiones

1. Sea la siguiente tabla estadística:

$x_i$	-5	-3	0	3	5
$n_i$	1	2	4	5	4

El valor de la media es:

- (A) 5    (B) 10    (C) 0    (D) Un valor entre -5 y 5    (E) Un valor mayor que 5
2. Si la mediana de la altura de los alumnos de una clase es 1,65m., podemos estimar que:
- (A) Al menos la mitad de los alumnos de la clase miden exactamente 1,65m.
  - (B) Al menos la mitad de los alumnos de la clase miden más de 1,50m.
  - (C) La mitad de los alumnos de la clase mide más de 1,80m.
  - (D) 1,65m. es la medida que más se repite en esta clase
  - (E) Todos los alumnos miden más de 1,65m.
3. Para una variable continua agrupada en intervalos y unimodal, su moda estará siempre situada:
- (A) Dentro del intervalo de mayor frecuencia absoluta
  - (B) Dentro del intervalo de mayor amplitud
  - (C) Dentro del intervalo de mayor frecuencia relativa
  - (D) Dentro del intervalo central
  - (E) Todo falso
4. Supongamos que en una empresa a todos y cada uno de los trabajadores se les sube el sueldo en un 9%. Según la varianza, la dispersión respecto a la media:
- (A) Se mantendrá igual    (B) Disminuirá ligeramente    (C) Todo falso
  - (D) Se duplicará    (E) Aumentará ligeramente
5. Si para una variable  $X$  ocurre que  $m_4(X) = 5S_x^4$ , podemos afirmar que la distribución de esta variable es:
- (A) Más apuntada que una distribución normal
  - (B) Dos veces más apuntada que una distribución normal

- (C) Dos veces menos apuntada que una distribución normal
- (D) Menos apuntada que una distribución normal
- (E) Todo falso
6. Si te dijeran que para dos variables estadísticas tipificadas  $T_x$  y  $T_y$  la relación entre sus coeficientes de variación es  $g_0(T_x) = 3g_0(T_y)$ , ¿cuál de las siguientes afirmaciones es correcta?:
- (A) Es absurdo considerar cualquier relación entre los coeficientes de variación de dos variables tipificadas
- (B) El recorrido intercuartílico de  $T_x$  es tres veces el de  $T_y$
- (C) La dispersión respecto de la media de  $T_x$  es tres veces menor que la de  $T_y$
- (D) La dispersión respecto de la media de  $T_x$  es tres veces mayor que la de  $T_y$
- (E) Todo falso
7. Si  $S_x^2 = 25$ ,  $m_3(X) = -36$  y realizamos la transformación lineal  $Y = -3X + 2$ , ¿cuál de las siguientes afirmaciones es correcta?:
- (A) El coeficiente de asimetría de  $Y$  es 0,5    (B)  $S_y = -5$     (C)  $m_3(Y) = -36$
- (D)  $m_3(Y) = 36$     (E) Todo falso
8. Si para una variable continua agrupada en intervalos ocurre que un intervalo  $[L_{i-1}, L_i]$  tiene una frecuencia absoluta acumulada  $N_i = \frac{N}{2}$ , entonces:
- (A)  $M_e = \bar{x}$     (B)  $M_e = L_i$     (C) Todo falso    (D)  $M_e = L_{i-1}$     (E)  $M_e = \frac{L_{i-1} + L_i}{2}$
9. Si para una variable estadística cuantitativa centrada tenemos que  $a_2(X) = -2$ , podemos afirmar que:
- (A) La media aritmética de esta variable también es negativa
- (B) Se trata de una distribución con asimetría negativa
- (C) Se trata de una distribución con poca dispersión
- (D) Se trata de una variable que toma valores negativos
- (E) Todo falso
10. Si a partir de una variable estadística  $X$  definimos una nueva variable  $Y$  como  $Y = \frac{X-b}{a}$ , ¿cuál de las siguientes afirmaciones es correcta?:
- (A) Todos los estadísticos de posición y de dispersión se ven afectados por los dos parámetros  $a$  y  $b$
- (B) Los estadísticos de posición se ven afectados solamente por el parámetro  $a$
- (C) Los estadísticos de dispersión, salvo  $g_0$ , se ven afectados solamente por el parámetro  $a$
- (D) Los estadísticos de posición se ven afectados solamente por el parámetro  $b$
- (E) Todo falso

**Las dos próximas cuestiones se refieren al siguiente enunciado.**

Se ha hecho un estudio sobre el precio, en miles de euros, de las viviendas en dos ciudades diferentes  $A$  y  $B$ , obteniéndose las siguientes tablas de resultados,  $\{x_i\}$  e  $\{y_i\}$  para  $A$  y  $B$ , respectivamente:

Precio ( $x_i$ )	Nº viv.	Precio ( $y_i$ )	Nº viv.
300	34	250	50
250	38	150	78
170	56	120	82
140	72	70	130
100	100	40	160
	300		500

Con los datos anteriores se han calculado los siguientes estadísticos:

$$\bar{x} = 164,3 \quad \bar{y} = 99,1 \quad S_x^2 = 4554,17 \quad S_y^2 = 4086,79$$

11. Ante estos resultados, ¿cuál de las siguientes afirmaciones consideras más adecuada?:
- (A) Existe comparativamente mayor dispersión en la distribución de los precios de las viviendas en la ciudad  $A$
  - (B) Son dos colectivos de diferente tamaño, por lo que no son comparables
  - (C) Existe comparativamente mayor dispersión en la distribución de los precios de las viviendas en la ciudad  $B$
  - (D) Las dos distribuciones tienen igual dispersión
  - (E) Todo falso
12. ¿Cuál de las siguientes afirmaciones es correcta?:
- (A) Las viviendas de 250 mil euros son relativamente más caras en la ciudad  $B$  que en la ciudad  $A$
  - (B) Las viviendas de 250 mil euros son relativamente más caras en la ciudad  $A$  que en la ciudad  $B$
  - (C) —
  - (D) No se pueden hacer este tipo de afirmaciones al ser distribuciones no comparables
  - (E) Todo falso

**Las dos próximas cuestiones se refieren al siguiente enunciado.**

En la siguiente tabla aparece la distribución del gasto en tabaco al mes, correspondiente a 130 personas con edad superior a 18 años.

gasto (decenas de euros)	nº de personas
[0, 2)	23
[2, 3)	47
[3, 6)	60

13. El gasto mensual en tabaco más frecuente en este colectivo estará entre:  
 (A) 30 y 60 euros (B) 0 y 20 euros (C) Todo falso (D) 20 y 30 euros (E)  
 —
14. El Ministerio de Sanidad va a comenzar una campaña para disminuir el consumo de tabaco. En un ensayo experimental comenzará el tratamiento con una muestra correspondiente al 50 % de las personas del colectivo anterior cuyo gasto mensual en tabaco sea más elevado. Aproximadamente entrarán a formar parte de este grupo de ensayo aquellas personas cuyo gasto mensual en tabaco sea superior a:  
 (A) 27 euros (B) 20 euros (C) 30 euros (D) 28,93 euros (E) Todo falso
15. Sea  $Y = aX + b$  ( $a, b \neq 0$ ) una transformación lineal de una variable tipificada  $X$ . Entonces:  
 (A)  $g_0(Y) = |a|$  (B)  $S_y = 1$  (C)  $g_0(Y) = \frac{|a|}{|b|}$  (D)  $S_y^2 = a^2 + b$  (E) Todo falso
16. Una compañía inmobiliaria tiene 200 apartamentos para alquilar cuya superficie media es de 67 m<sup>2</sup>. El alquiler mensual de un apartamento se determina en función de su superficie (el m<sup>2</sup> se alquila a razón de 4 €/mes) más una cantidad fija mensual en concepto de comisión. El valor de esta comisión, si el alquiler medio de los apartamentos es de 300 €, es de:  
 (A) 384 € (B) 2,67 € (C) Todo falso (D) 32 € (E) 28 €
17. Se desea obtener el grado de concentración existente en la distribución de los salarios entre los 50 empleados de un supermercado. Se sabe que todos los empleados ganan 800 € al mes. ¿Cuál será el valor del índice de Gini?:  
 (A) 1 (B) 0,5 (C) Todo falso (D) 0  
 (E) Necesitamos más datos para obtener el valor del índice de Gini
18. En una empresa  $A$  se han repartido 50 millones de beneficios entre 20 socios. En otra empresa  $B$  se han repartido 100 millones entre 40 socios. ¿Cuál será la relación entre los índices de concentración obtenidos en cada reparto?:  
 (A)  $I_G(B) = 1/2I_G(A)$   
 (B)  $I_G(B) = 2I_G(A)$   
 (C)  $I_G(B) \neq I_G(A)$   
 (D)  $I_G(B) = I_G(A)$   
 (E) El  $I_G$  depende de cómo se realice el reparto en cada empresa
19. De una tabla que recoge, por un lado, el porcentaje acumulado de la masa salarial a repartir y, por otro, el porcentaje acumulado de individuos entre los que se distribuye, se desprende el siguiente resultado: “El 10 % de los individuos que menos ganan reciben el 40 % de la masa salarial total” ¿Cómo interpretarías este resultado?:  
 (A) Los datos aportados no son suficientes  
 (B) Se trata de un reparto concentrado  
 (C) Los datos aportados no son compatibles

(D) Se trata de un reparto muy equitativo

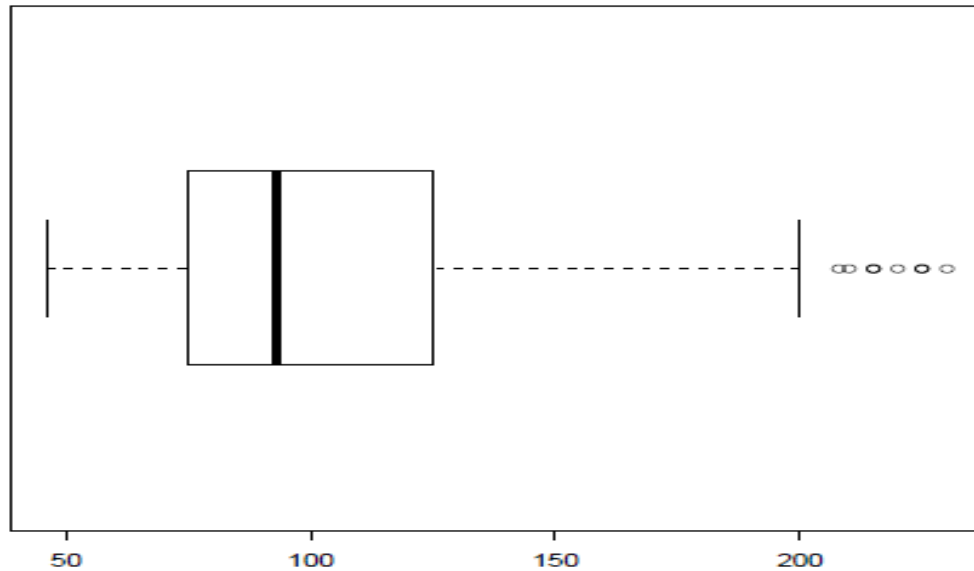
(E) Se trata de una distribución dispersa

20. Para una distribución con un índice de concentración igual a 0, ¿cuál será el valor de  $Q_i$  siendo  $P_i = 30\%$ ?:

(A) 0% (B) 10% (C) No hay datos suficientes (D) 100% (E) Todo falso

**Las dos próximas cuestiones se refieren al siguiente enunciado.**

El siguiente diagrama de caja recoge la distribución de la velocidad (Km/h) a la que circulaba un colectivo de 391 vehículos en una carretera comarcal vizcaína:



21. Observando el diagrama podemos afirmar que:

(A) El 50% central de los valores de la variable velocidad se encuentra, aproximadamente, entre 125 y 200 Km/h.

(B) El 25% de los vehículos que circulaban a menor velocidad está más disperso que el 25% de los vehículos que circulaban a mayor velocidad.

(C) El primer cuartil es aproximadamente 200 Km/h.

(D) La distribución de la variable velocidad es asimétrica por la derecha.

(E) Todo falso.

22. Supongamos que la velocidad máxima permitida en este tramo era de 100 Km/h, entonces es cierto que:

(A) Un vehículo circulaba a menos de 25 Km/h.

(B) La mitad de los vehículos circulaban a más de 120 Km/h.

(C) Ningún vehículo circulaba a más de 200 Km/h.

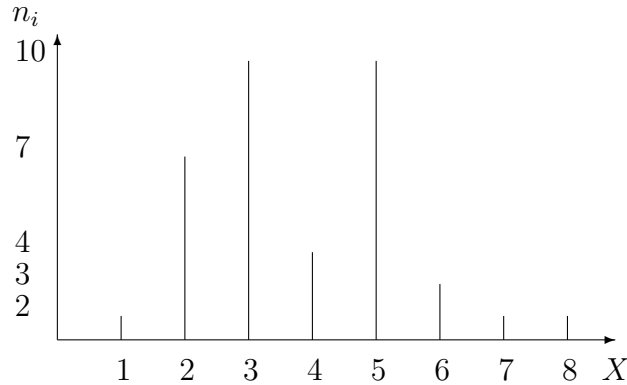
(D) Más de la mitad de los conductores respetaba el límite de velocidad.

(E) Todo falso.

# Solución a los problemas y cuestiones del Tema 1

## Problemas

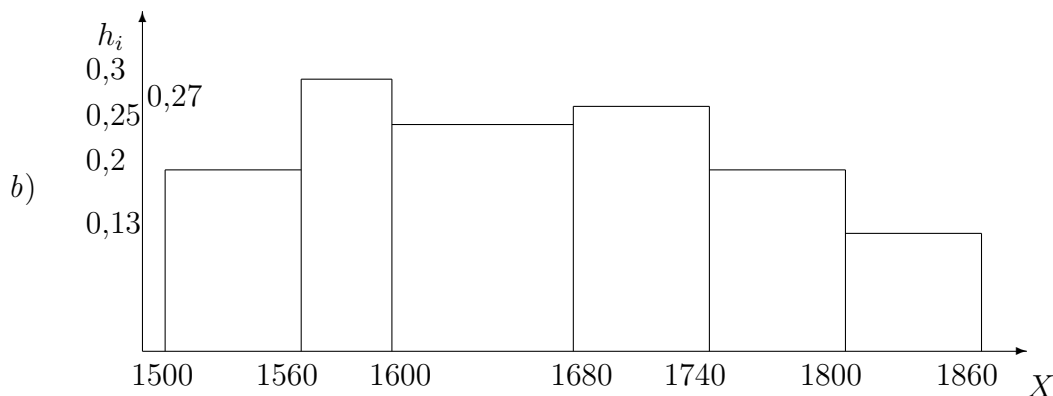
1. a) Diagrama de barras.



- b)  $\bar{x} = 4$  días;  $M_o = 3$  y  $5$  (bimodal);  $q_1 = 3$  días,  $q_2 = M_e = 4$  días y  $q_3 = 5$  días.  
 c)  $g_0(X) = 0,44$   
 d) La media no varía ya que los individuos no incluidos inicialmente tienen el mismo valor de  $X$  que la media,  $\bar{x}$ . Sin embargo, la dispersión de los datos será menor, por lo que el coeficiente de variación disminuye y la representatividad de la media aumenta.
2. a)  $N = 80$  es el tamaño del colectivo objeto de estudio.  $X$ : salario mensual en euros es la variable a estudiar. La distribución de frecuencias se representa en la siguiente tabla:

	$x_i$	$f_i$	$n_i = 80f_i$	$F_i$	$N_i$	$c_i$	$h_i$
[1500, 1560)	1530	0,15	12	0,15	12	60	0,20
[1560, 1600)	1580	0,15	12	0,30	24	40	0,30
[1600, 1680)	1640	0,25	20	0,55	44	80	0,25
[1680, 1740)	1710	0,20	16	0,75	60	60	0,27
[1740, 1800)	1770	0,15	12	0,90	72	60	0,20
[1800, 1860)	1830	0,10	8	1,00	80	60	0,13

En la tabla no sólo hemos obtenido las marcas de clase, frecuencias absolutas y relativas, sino que además hemos obtenido las acumuladas, la amplitudes de intervalos y las alturas de los rectángulos que forman el histograma. Así podemos realizar fácilmente el apartado siguiente.



c) Para calcular la media,

$$\bar{x} = \sum_{i=1}^6 x_i f_i = 1667 \text{ euros.}$$

Para calcular la mediana, podemos hacerlo en términos relativos, para ello, miramos cuál es el primer intervalo tal que  $F_{i-1} < 0,5$  y  $F_i > 0,5$ . Vemos que es el tercer intervalo,  $[1600, 1680)$ . Para estimar un valor concreto de salario mediano, recordamos que es aquel valor que deja un 50 % de área en el histograma a su izda. y un 50 % a su dcha.

$$Me(X) = L_{i-1} + \frac{0,5 - F_{i-1}}{f_i} c_i = 1600 + \frac{0,5 - 0,30}{0,25} 80 = 1664 \text{ euros.}$$

también podemos hacerlo en términos absolutos, para ello, miramos cual es el primer intervalo tal que  $N_{i-1} < 40 = \frac{N}{2}$  y  $N_i > 40$ . Comprobamos de nuevo que es el tercer intervalo,  $[1600, 1680)$ . La expresión que nos permite estimar la mitad del área en el histograma, es ahora

$$Me(X) = L_{i-1} + \frac{40 - N_{i-1}}{n_i} c_i = 1600 + \frac{40 - 24}{20} 80 = 1664 \text{ euros.}$$

La respuesta a cuál es el intervalo de salario más frecuente, es el intervalo modal, esto es aquel que presenta una mayor altura (frecuencia por unidad) en el histograma. Se trata del segundo intervalo,  $[1560, 1600)$ .

d) El hecho de que  $\bar{x} > Me(X)$  indica una asimetría por la derecha. Lo comprobamos mediante el coeficiente de asimetría,

$$g_1 = \frac{m_3}{S_x^3} = \frac{123486}{818607,399} = 0,150 > 0$$

donde  $m_3 = \sum_{i=1}^6 (x_i - 1667)^3 f_i = 123486$ , y  $S_x^2 = \sum_{i=1}^6 (x_i - 1667)^2 f_i = 8751$ . Luego,  $S_x = 93,546$ .

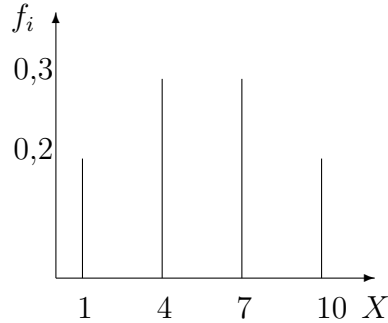
Efectivamente,  $g_1(X)$  es distinto de cero, y positivo, aunque no es un valor muy grande. Podemos confirmar la ligera asimetría positiva o por la derecha que se aprecia en el dibujo.

Si calculamos el coeficiente de curtosis,

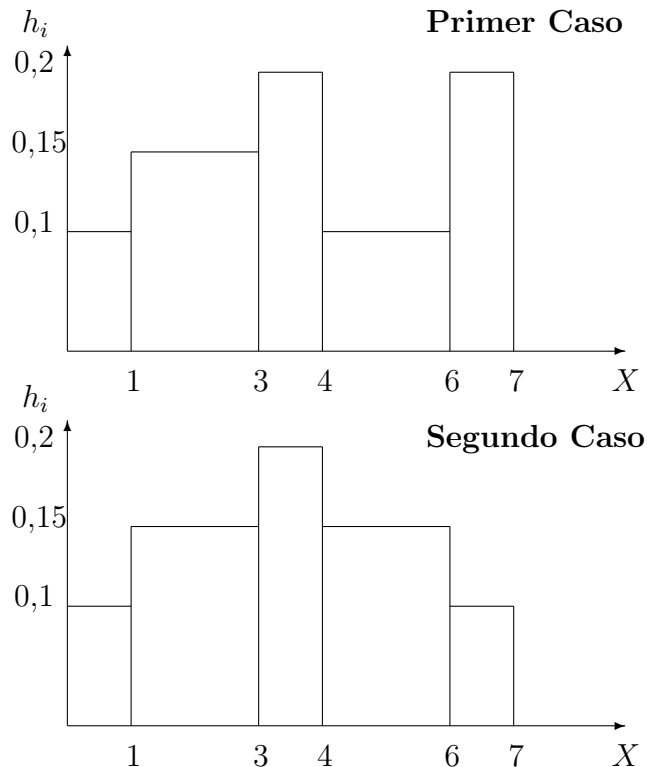
$$g_2 = \frac{m_4}{S_x^4} - 3 = \frac{149725197}{76577447,76} - 3 = -1,044 < 0$$

donde  $m_4 = \sum_{i=1}^6 (x_i - 1667)^4 f_i = 149725197$ . Se comprueba que el coeficiente de curtosis es negativo, lo que indica que se trata de una distribución con forma platicúrtica, ésto es más plana que la correspondiente normal  $N(1667, S_x = 93,543)$ .

3. a) Al ser una distribución simétrica,  $f_2 = f_3$ . Por tanto,  $\bar{x} = \sum_{i=1}^4 x_i f_i = 5,5$ .  
 b)  $f_1 = f_4 = 0,2$  y  $f_2 = f_3 = 0,3$ . La representación gráfica adecuada de esta distribución es el diagrama de barras.



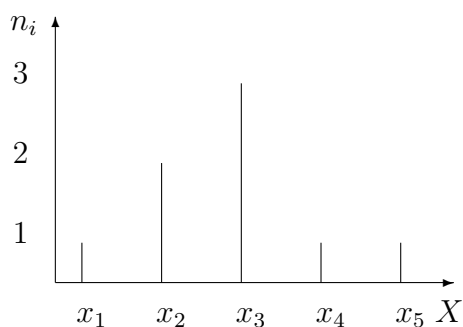
- c)  $\bar{y} = 0,5$ ;  $S_y^2 = 1,05$  y  $g_0(Y) = 2,049$ .
4. a)  $f_3 = f_4 = f_5 = 0,2$   
 b)  $f_3 = 0,2$ ;  $f_4 = 0,3$ ;  $f_5 = 0,1$ ;  $\bar{x} = M_e = 3500$  €  
 c) Representaciones gráficas: histogramas.



La diferencia entre ambos está en la asimetría. En el segundo caso, el gráfico es simétrico, mientras que en el primero no (la media aritmética es mayor que la

mediana, lo que indica cierta asimetría positiva). La distribución de las rentas hasta 4000 euros es igual en ambos casos. En el primero, las rentas más altas vienen representadas por una superficie mayor y las comprendidas entre 4000 y 6000 euros tienen menor frecuencia que en el segundo caso. Por ello, en el primer caso hay mayor dispersión respecto a la media.

5. a) Diagrama de barras.



- b)  $\bar{x} = x_1 + 3,75$ ;  $M_e = M_o = x_1 + 4$ .
- c) Realizando la transformación  $U = X - x_1$  se obtiene que  $R_u = R_x = 8$  y  $S_u = S_x = 2,33$ .
- d)  $\bar{x}$ ,  $M_e$  y  $M_o$  dependen de  $x_1$  por ser estadísticos de posición y varían con los cambios de origen. Sin embargo,  $S_x$  y  $R$  son estadísticos de dispersión y no les afectan los cambios de origen.
6. a) “El valor tal que deja a la izquierda y a la derecha el mismo número de valores de  $X$  es 4,93”.
- b)

$$g_0^Y = \frac{\sqrt{552,54}}{11,48} = 2,04757$$

- c) Calculamos el coeficiente de variación de  $X$  y lo comparamos con el coeficiente de variación de  $Y$  arriba obtenido:

$$g_0^X = \frac{\sqrt{17,68}}{6,84} = 0,61473 < g_0^Y = 2,04757$$

Dado que  $g_0^X$  es menor, la variación es menor en la distribución de la variable  $X$  y  $\bar{x}$  es más representativa que  $\bar{y}$ .

- d) Comparamos los valores tipificados de ambos rendimientos:

$$\frac{11 - \bar{x}}{S_X} = \frac{11 - 6,84}{4,2047} = 0,9893$$

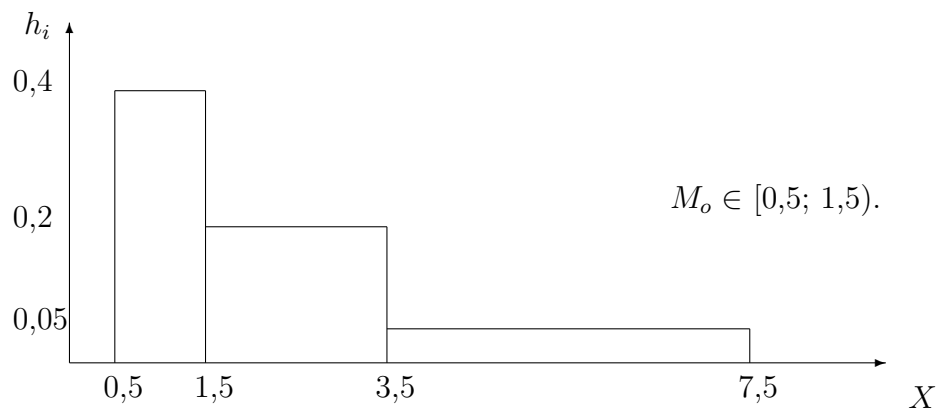
$$\frac{19 - \bar{y}}{S_Y} = \frac{19 - 11,48}{23,506} = 0,3199$$

Acabamos de comprobar que el valor tipificado en el caso de las Letras del tesoro es mayor que el valor tipificado del IBEX35, por lo tanto el rendimiento de las Letras ese mes destaca más en relación al resto de observaciones que el del IBEX35.

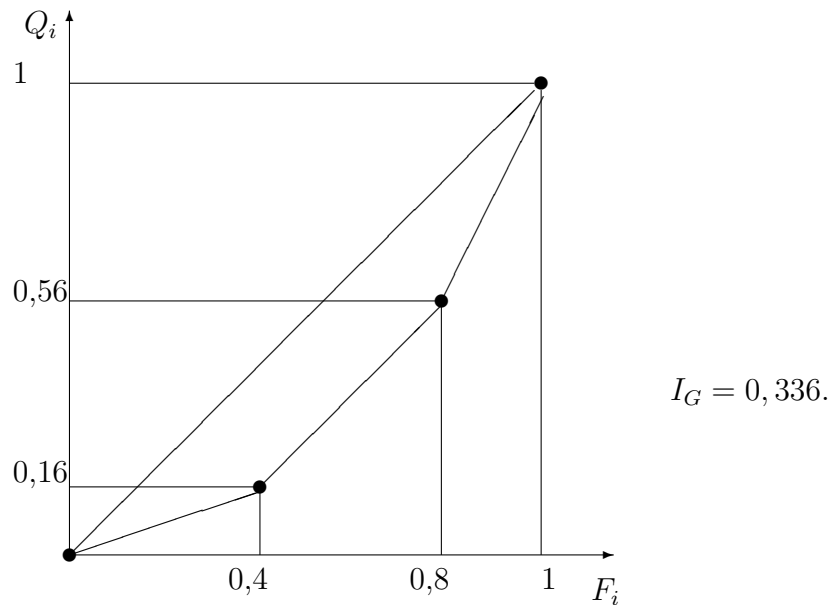
7. a) La persona más baja medirá alrededor de 155 cm. Y la más alta alrededor de 198 cm.
- b) Aproximadamente 168 cm.
- c) Aproximadamente 180 cm.
- d) Se pueden comentar muchas cosas como por ejemplo:
- En ambos casos la mediana está aproximadamente en 175 cm., de modo que en ambos casos la altura del 50 % de los individuos es menor o igual a 175 cm.
  - Sin embargo el 50 % central de los valores de  $X$  está entre 168 cm. y 180 cm. mientras que el 50 % central de los de  $Y$  está entre 173 cm. y 180 cm. Los individuos “medianos” en el caso de los padres son un poco más altos que en el caso de los hijos.
  - El 25 % de los valores más altos de  $X$  está entre 180 cm. y 198 cm., mientras que el 25 % de los valores más altos de  $Y$  están entre 180 cm. y 188 cm. De modo que los individuos más altos lo son más que los padres más altos.
  - El recorrido de la distribución altura es mayor que el de la distribución altura del padre, por lo tanto la dispersión absoluta de la primera es mayor que la de la segunda.
- e) No, porque ninguno de los diagramas de caja posee datos atípicos.
8. a) 130 €.
- b)  $I_G = 0,2916$ .
9. a) La distribución de los perceptores según la subvención recibida se representa en la siguiente tabla:

	$x_i$	$f_i$	$F_i$	$c_i$	$h_i$	$q_i$	$Q_i$	$F_i^*$
[0,5;1,5)	1	0,4	0,4	1	0,4	0,16	0,16	0,2
[1,5;3,5)	2,5	0,4	0,8	2	0,2	0,4	0,56	0,6
[3,5;7,5)	5,5	0,2	1	4	0,05	0,44	1	0,9

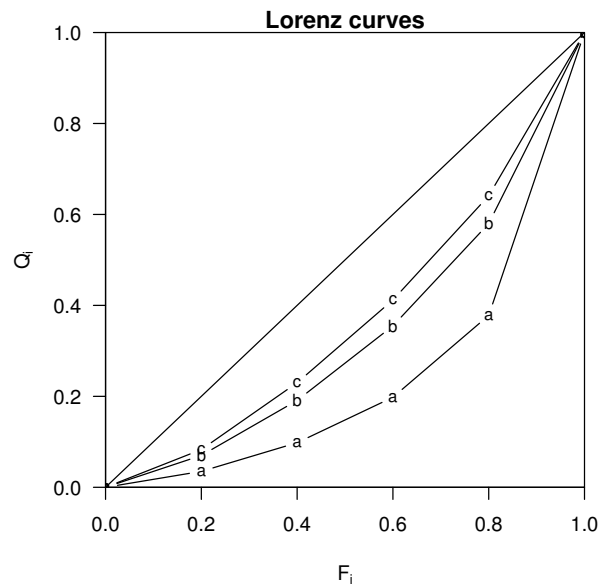
El histograma que representa la distribución anterior es:



b) La Curva de Lorenz es:



10. a) Siendo a=Sudáfrica, b=España y c=Dinamarca



b)

$$I_G(a) > I_G(b) > I_G(c)$$

La distribución de la renta en Sudáfrica es menos equitativa que en España y ésta, a su vez, menos equitativa que en Dinamarca.

c) Una curva de Lorenz de renta no ofrece información sobre cantidad de renta sino sobre concentración (grado de equidad de un reparto) de renta.

11. a) 1600 €.

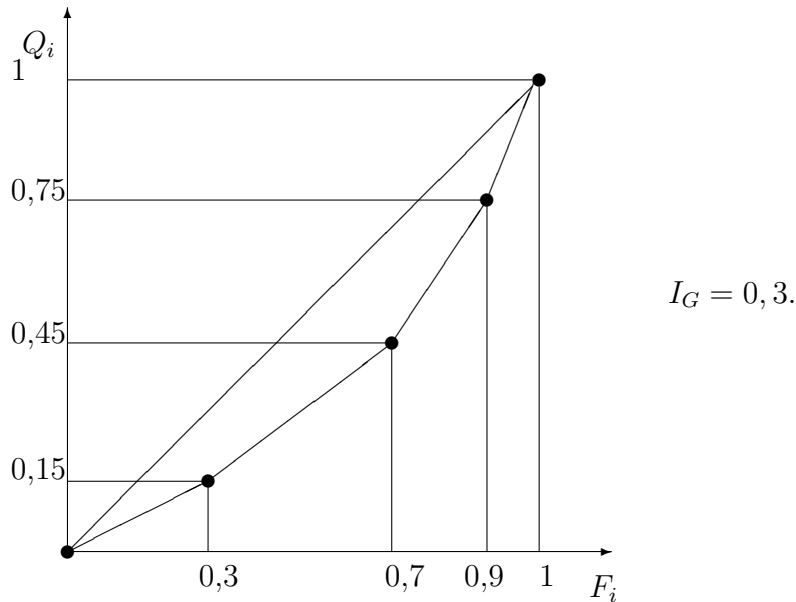
b) El índice de Gini antes de la subida es  $I_G(X) = 0,2625$ . Tras el aumento del sueldo,  $X + 1600$ , el valor de este índice disminuirá (puede comprobarse que se reduce a la mitad,  $I_G(X + 1600) = 0,13125$ ).

c) Si se multiplica cada valor de la variable por una constante  $k$ ,  $kx_i$ , el índice de Gini no varía; puesto que este cambio no afecta ni a los  $F_i^* = F_i - (f_i/2)$  ni a los  $q_i = \frac{kx_i n_i}{\sum_{j=1}^k kx_j n_j} = \frac{x_i n_i}{\sum_{j=1}^k x_j n_j}$ .

12. a)  $\bar{x} = 2,5$  millones.  $M_o = 1,875$  millones.

b)  $g_0 = \frac{S_x}{\bar{x}} = \frac{1,522}{2,5} = 0,608$ , luego la dispersión no es muy alta.

c) La Curva de Lorenz es:



d) Tanto el estudio de la dispersión como el de la concentración nos llevan a decir que estamos ante una distribución de valores no muy dispersos y donde la masa total está medianamente concentrada. La media no es un mal representante de la distribución.

## Cuestiones

- |      |      |      |      |      |
|------|------|------|------|------|
| 1.D  | 2.B  | 3.E  | 4.E  | 5.A  |
| 6.A  | 7.E  | 8.B  | 9.E  | 10.C |
| 11.C | 12.A | 13.D | 14.D | 15.C |
| 16.D | 17.D | 18.E | 19.C | 20.E |
| 21.D | 22.D |      |      |      |