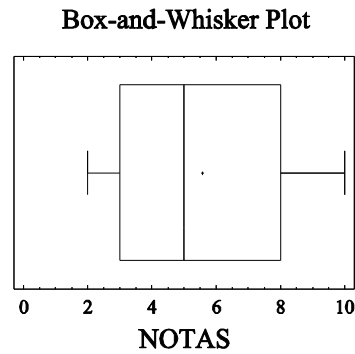


EJERCICIOS U.D. 2-ESTADÍSTICA DESCRIPTIVA

1. En el siguiente diagrama Box-Whisker están representadas las notas obtenidas por un grupo de estudiantes de 1º de Ingeniería Informática en el 1º parcial de la asignatura estadística



Contesta los siguientes apartados, justificando todas las respuestas:

- ¿Cuál es la población en estudio? ¿Cuál es la variable aleatoria implicada?
- ¿Cuál es el valor del rango intercuartílico?
- ¿Cuál sería el parámetro de posición más adecuado para describir esta muestra?
- ¿Cuál sería el parámetro de dispersión más adecuado para describir esta muestra?
- A partir del diagrama Box-Whisker, ¿se podría determinar cuantas notas se han utilizado para representarlo? En caso afirmativo, decir cuantas.

SOLUCIÓN:

a)

La población de estudio son los **estudiantes de 1.º curso de Ingeniería Informática**. La variable aleatoria o característica que se está estudiando es la **nota que los alumnos obtienen en el parcial de la asignatura**.

b)

A través del diagrama box-and-whisker podemos conocer, aproximadamente, el valor de los cuartiles 1.º y 3.º.

- Cuartil 1.º (C1) = lado izquierdo de la caja = 3.
- Cuartil 3.º: (C3) = lado derecho de la caja = 8.

El rango o recorrido intercuartílico (RI) se define como la diferencia entre los cuartiles 3.º y 1.º, es decir:

$$RI = C3 - C1 = 8 - 3 = 5.$$

c)

La muestra presenta una ligera asimetría positiva (*síntomas* de asimetría positiva: el bigote derecho es más largo que el izquierdo, la mediana no está situada en el centro de la caja, y la media está situada a la derecha de la mediana). Por tanto, **el parámetro de posición más adecuado para resumir los datos es la mediana** (dado que la media se vería afectada por la asimetría de los datos).

Podemos obtener la mediana, a través del box-and-whisker, como la línea vertical dentro de la caja:

$$\text{Mediana} = (\text{línea vertical dentro de la caja}) = 5.$$

d)

Dada la asimetría de la muestra, **el parámetro de dispersión más adecuado para resumir los datos es el rango intercuartílico** (la desviación típica se vería afectada por la asimetría de los datos).

En la pregunta 1.b) ya hemos calculado el valor del rango intercuartílico.

e)

En general, a partir del diagrama box-and-whisker **no es posible conocer el tamaño muestral**, es decir, la cantidad de datos recogidos para construir el gráfico.

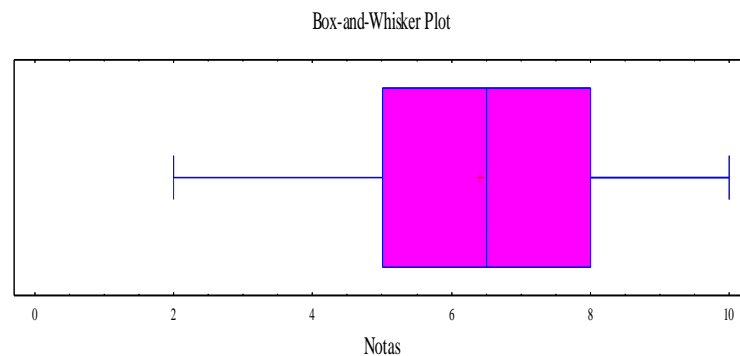
En principio, se podría obtener un gráfico como el mostrado prácticamente con cualquier cantidad de datos.

2. Las notas obtenidas por un grupo de 10 alumnos en el examen parcial de la asignatura Estadística que se imparte en la Escuela de Informática de una determinada universidad son:

2 5 8 7 4 8 9 6 5 10

- a) Indicar la población, la muestra y la variable aleatoria que intervienen en el estudio, así como el tipo del que es esta última.

El diagrama Box-Whisker de los datos es:



- b) ¿Es en este caso la media un buen parámetro de posición o sería más conveniente utilizar otro?
c) Calcular un parámetro de dispersión de los datos. Justificar por qué se ha elegido.

SOLUCIÓN:

- a) Población: Todos los estudiantes de la asignatura
Muestra: Los 10 estudiantes
Variable aleatoria: La nota obtenida, es cuantitativa continua
- b) Al ser aproximadamente simétrico el BW, la media muestral sí que es un buen parámetro de posición.
- c) Un buen parámetro de dispersión puede ser tanto la varianza como la desviación típica, ya que el BW es simétrico.

3. Una cibernauta que tiene un blog desde el 9 de marzo de 2009 desea hacer un estudio sobre el número de visitas que recibe diariamente. Mirando el contador que tiene instalado en la página web, el número de visitas que tiene registrado en la última semana es el siguiente:

Lunes 25 mayo	Martes 26 mayo	Miércoles 27 mayo	Jueves 28 mayo	Viernes 29 mayo	Sábado 30 mayo	Domingo 31 mayo
5	2	4	6	7	8	9

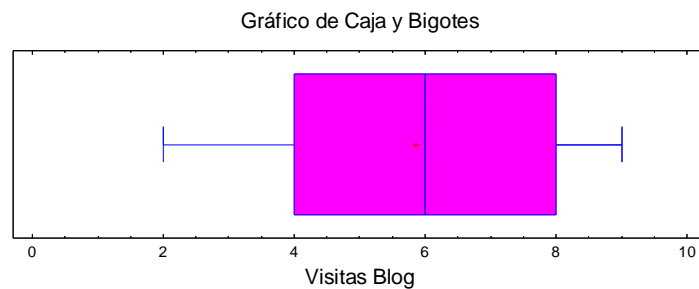
a) Indicar quién es la población, la muestra, y la variable aleatoria y el tipo de ésta

POBLACIÓN: Todos los días desde el 9 de marzo de 2009

MUESTRA: Los 7 días de la semana anterior

VARIABLE: El número de visitas diarias. Es una variable cuantitativa discreta

El diagrama Box and Whisker de los datos es:



b) A la vista de la gráfica obtenida, indicar y calcular un parámetro de posición que represente correctamente los datos. ¿Por qué lo has elegido?

Se observa en el gráfico que los datos siguen una distribución simétrica.

Parámetro de posición más adecuado cuando los datos son simétricos: la media muestral, que resulta igual a 5,86.

c) ¿Sería el intervalo intercuartílico un buen parámetro de dispersión? ¿Por qué? Si no lo es, indicar y calcular un parámetro de dispersión que represente bien los datos.

Parámetro de dispersión más adecuado cuando hay simetría sería: la varianza muestral, que resulta igual a 5,8 o la desviación típica, que vale en este caso 2,4.

4. Una cibernauta que tiene un blog desde el 9 de marzo de 2009 desea hacer un estudio sobre el número de visitas que recibe diariamente. Mirando el contador que tiene instalado en la página web, el número de visitas que tiene registrado en la última semana es el siguiente:

Lunes 24 agosto	Martes 25 agosto	Miércoles 26 agosto	Jueves 27 agosto	Viernes 28 agosto	Sábado 29 agosto	Domingo 30 agosto
3	4	5	6	5	10	19

a) Indicar cuáles son la población, la muestra, y la variable aleatoria y el tipo de ésta

POBLACIÓN: Todos los días desde el 9 de marzo de 2009

MUESTRA: Los 7 días de la semana anterior

VARIABLE: El número de visitas diarias. Es una variable cuantitativa discreta

El diagrama Box and Whisker de los datos es:



b) A la vista de la gráfica obtenida, indicar y calcular el parámetro de posición que mejor representa los datos. ¿Por qué?

El mejor parámetro de posición sería la mediana porque el diagrama Box-Whisker indica que la distribución de los datos es asimétrica positiva (o por la derecha)

c) ¿Sería el intervalo intercuartílico un buen parámetro de dispersión? ¿Por qué? Si no lo es, indicar y calcular un parámetro de dispersión que represente bien los datos.

El intervalo intercuartílico vale 6 y es un buen parámetro de dispersión dada la asimetría positiva de los datos.

5. Se ha medido la velocidad de la carga de una determinada página web en 10 momentos diferentes del día y se han obtenido los siguientes valores en décimas de segundo:

2 3 1,5 2,2 4 2,1 2 8 12 5

El gráfico Box-Whisker de estos datos es:

Datos ordenados de menor a mayor 1,5 2 2 2,1 2,2 3 4 5 8 12

Mediana= $2,6=(2,2+3)/2$

Primer cuartil= $C1=2$

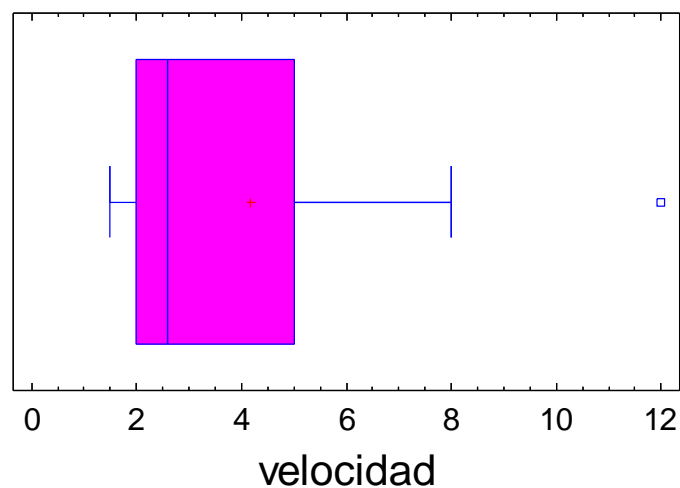
Tercer cuartil= $C3=5$

Intervalo intercuartílico=3

$1,5 \times 3=4,5$

Media muestral =4,18

Box-and-Whisker Plot



a) ¿Se trata de una variable simétrica?

No. Es asimétrica por la derecha (asimetría positiva)

b) ¿Qué parámetros de posición y dispersión serían los más adecuados para describir la muestra?

Para posición la mediana (2,6) y para dispersión el intervalo intercuartílico (3).

6. En el marco de un estudio sobre la posible incidencia que tiene la religión profesada por los distintos matrimonios en la presencia de una mayor o menor frecuencia de divorcios, se ha tomado una muestra aleatoria simple a nivel mundial de tamaño 32000
 Nota: Sólo se han considerado matrimonios donde los dos cónyuges tenían la misma religión .

Nota: Los datos del presente ejercicio no proceden de un estudio real y el único objeto del presente ejercicio es la evaluación de los aprendizajes.

Religión / Divorcio	Divorcio	No divorcio
Católicos	1435	7565
Ateos	845	2155
Musulmanes	160	7840
Protestantes	610	4390
Otros	1250	5750

- Basando tus razonamientos y afirmaciones en las frecuencias relativas que resulten más informativas para este estudio, señalar cuál es la religión donde los matrimonios presentan una mayor probabilidad de terminar en divorcio.
- Obtener las frecuencias marginales de la variable divorcio.

SOLUCIÓN:

a)

Si el objetivo es estudiar cómo se distribuyen los divorcios dentro de cada religión (para ver si hay alguna donde sean más frecuentes), **las frecuencias relativas que resultan más adecuadas son las de la variable divorcio condicionadas a cada religión (es decir, “porcentajes que suman 100% por filas”)**, ya que nos permiten ver la proporción de divorcios en cada religión, para posteriormente comparar dicha proporción entre las diferentes religiones. Esto es:

Religión	Divorcio	No divorcio	Total
Católicos	1.435 (15,94%)	7.565 (84,06%)	9.000 (100%)
Ateos	845 (28,17%)	2.155 (71,83%)	3.000 (100%)
Musulmanes	160 (2%)	7.840 (98%)	8.000 (100%)
Protestantes	610 (12,2%)	4.390 (87,8%)	5.000 (100%)
Otros	1.250 (17,86%)	5.750 (82,14%)	7.000 (100%)

La “religión” que presenta una mayor proporción de divorcios es la de los “ateos” (28,17%), seguida por la categoría de “otras religiones” (17,86%).

Nótese que, por ejemplo, la cantidad (frecuencia absoluta) de matrimonios católicos que se divorcian es muy alta, en comparación con la de ateos ($1.435 > 845$), pero representa “sólo” un 15,94% del total de matrimonios católicos. Por tanto, si hubiéramos elegido las frecuencias relativas “que suman 100% en cada columna”, habríamos obtenido resultados y conclusiones incorrectos.

b)

Las frecuencias marginales (absolutas y relativas) de la variable “divorcio” son las frecuencias totales observadas para dicha variable, independientemente de la variable religión, y se obtienen sencillamente sumando el total de cada columna. Es decir:

Religión	Divorcio	No divorcio
Católicos	1.435	7.565
Ateos	845	2.155
Musulmanes	160	7.840
Protestantes	610	4.390
Otros	1.250	5.750
Total (frecs. marginales de “divorcio”)	4.300 (13,44%)	27.700 (86,56%)

Estas frecuencias, por tanto, nos informan de qué cantidad y proporción de la muestra de 32.000 matrimonios se divorcian (independientemente de la religión que profesen).

7. A un grupo de 68 alumnos se les ha consultado sobre la creencia que tienen acerca de que la selección española de fútbol pase de cuartos en la Eurocopa 2008 de Suiza y Austria. En la siguiente tabla se indica el sexo del alumno y si espera que España pase de cuartos.

	CHICOS	CHICAS	TOTAL FILA
SÍ	18	15	33
NO	27	8	35
TOTAL COLUMNA	45	23	68

Indicar justificando si las siguientes afirmaciones son ciertas o falsas:

a) La tabla anterior se denomina tabla de frecuencias condicionales.

FALSO, es una tabla de frecuencias cruzadas

b) Aproximadamente el 26,5% de los alumnos son chicos y piensan que la selección pasa de cuartos.

VERDADERA, ya que $18/68 = 0,2647$

c) Aproximadamente, el 65% de los que piensan que la selección pasa de cuartos son chicas.

FALSO, ya que $15/33 = 0,4545$

- d) Prácticamente el 23% de los alumnos que piensan que la selección no pasa de cuartos son chicas.

VERDADERO, ya que $8/35 = 0,22857$

- e) En el grupo hay aproximadamente una chica por cada dos chicos.

VERDADERO, ya que $23/68 = 0,3382$

8. En la siguiente tabla, se muestran los resultados de una encuesta de opinión, realizada a cuatro grupos de estudiantes de diferentes ingenierías, (agrónomos (AGR), informáticos (INF), industriales(IND), ing. de caminos (CAM)). En dicha encuesta se les preguntaba sobre su preferencia energética a la hora de conseguir reducciones en las emisiones de CO₂ (nuclear (N), fotovoltaica (F), eólica (E), hidráulica (H)).

	N	F	E	H
AGR	6	35	23	25
INF	3	45	20	25
IND	10	57	34	27
CAM	3	10	15	49

Contesta a las siguientes preguntas justificando los cálculos realizados, representando en los cálculos hasta la segunda cifra decimal aplicando redondeo.

- a) Calcula el porcentaje de ingenieros informáticos que prefieren la energía fotovoltaica. Indica el tipo de frecuencia que se pide.

$$Porc(F | INF) = \frac{45}{93} \cdot 100 = 48,39\%$$

Es la frecuencia relativa de energía fotovoltaica condicionada a ingenieros informáticos.

- b) ¿Cuál es el porcentaje de individuos que estudian agrónomos y prefieren la energía eólica?. Indica el tipo de frecuencia que se pide.

$$Porc(AGR \cap E) = \frac{23}{387} \cdot 100 = 5,94\%$$

Es la frecuencia relativa conjunta de agrónomos y energía eólica.

- c) ¿Qué porcentaje de los estudiantes que eligieron la energía hidráulica estudian industriales?. Indica el tipo de frecuencia que se pide.

$$Porc(IND | H) = \frac{27}{126} \cdot 100 = 21,43\%$$

Es la frecuencia relativa de industriales condicionada a energía hidráulica.

- d) ¿Cuál es el porcentaje de estudiantes en la encuesta que estudia caminos?. Indica el tipo de frecuencia que se pide.

$$Porc(CAM) = \frac{77}{387} \cdot 100 = 19,90\%$$

Es la frecuencia relativa marginal de estudiantes de caminos.

- e) ¿Cuál es el porcentaje de individuos que han elegido la energía nuclear?. Indica el tipo de frecuencia que se pide.

$$Porc(N) = \frac{22}{387} \cdot 100 = 5,68\%$$

Es la frecuencia relativa marginal de estudiantes que prefieren la energía nuclear.

9. En un estudio sobre las reclamaciones que reciben las compañías aéreas de bajo coste, se han tomado al azar 250 reclamaciones entre todas las que se han recibido en los últimos seis meses en las compañías O'Clock Air, Isillet Air, Brian Air y Iberfly y se han clasificado de acuerdo con su naturaleza en los siguientes tipos: reclamaciones sobre equipajes, reclamaciones en el servicio recibido durante el vuelo, reclamaciones relacionadas con las facturas o pagos y reclamaciones por cancelaciones de vuelo. Con los datos recogidos se ha confeccionado la siguiente tabla de frecuencias:

Frequency Table for Reclamación by Compañía

	Brian Air	Iberfly	Isillet	O'Clock Air	Row Total
Cancelaciones	30	2	15	50	97
Equipajes	15	12	5	57	89
Facturas-pago	5	8	5	20	38
Servicio durante vuelo	6	3	2	15	26
	56	25	27	142	250

- a) ¿Cuál es la probabilidad de que si se escoge una reclamación al azar, ésta resulte ser una reclamación de equipaje y de la compañía Iberfly?

$$P(\text{Equipaje} \cap \text{Iberfly}) = 12/250 = 0.048$$

- b) ¿Cuál es la probabilidad de que si se escoge una reclamación de O'Clock Air al azar, ésta se refiera a una cancelación?

$$P(\text{Cancelación/O'Clock}) = 50/142 = 0.352$$

- c) Obtener las frecuencias marginales relativas correspondientes a la variable compañía aérea.

$$\begin{aligned}\text{Frec. Marginal Rel. Brian Air} &= 56/250 \times 100 = 22.4 \% \\ \text{Frec. Marginal Rel. Iberfly} &= 25/250 \times 100 = 10 \% \\ \text{Frec. Marginal Rel. Isillet} &= 27/250 \times 100 = 10.8 \% \\ \text{Frec. Marginal Rel. O'Clock Air} &= 142/250 \times 100 = 56.8 \%\end{aligned}$$

- d) Obtener las frecuencias relativas condicionales “compañía/reclamación” para las reclamaciones de equipajes.

Frec. Relativas Condicionales:

$$\begin{aligned}\text{Brian Air/Equipajes} &= 15/89 \times 100 = 16.85 \% \\ \text{Iberfly/Equipajes} &= 12/89 \times 100 = 13.48 \% \\ \text{Isillet/Equipajes} &= 5/89 \times 100 = 5.62 \% \\ \text{O'Clock Air/Equipajes} &= 57/89 \times 100 = 64.06\%\end{aligned}$$

10. En una encuesta realizada a 310 valencianos se ha recogido información sobre su deporte favorito y se ha clasificado teniendo en cuenta el sexo de los individuos en la siguiente tabla de frecuencias cruzadas:

	Tenis	Fútbol	Fórmula 1	Vela
Hombres	27	86	35	22
Mujeres	27	69	26	18

- a) ¿Cuál es la frecuencia relativa marginal del suceso “preferir el fútbol”?

$$(86+69)/310 \times 100 = 50\%$$

- b) ¿Cuál es la frecuencia relativa de preferir la vela y ser mujer?

$$P(\text{ser mujer} \cdot \text{preferir la vela}) = 18/310 = 0.058$$

- c) ¿Cuál es la probabilidad de que un individuo sacado al azar que resultó ser chico tenga preferencia por el fútbol?

$$86/(27+86+35+22) = 86/170 = 0.506$$