

# Prueba de Estadística descriptiva

Apellidos:

Nome:

DNI:

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas<sup>1</sup>.

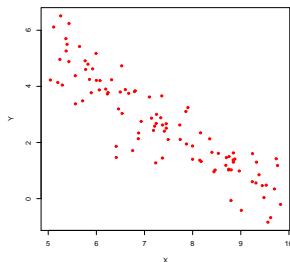
Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

1. 2ptos Si el rango intercuartílico vale 5, entonces:

- a) Por lo menos el 50 % de los valores está a una distancia igual a 5 de la mediana.
- b) Por lo menos el 50 % de los valores está a una distancia menor de 5 de la media.
- c) Por lo menos el 50 % de los valores está a una distancia menor de 5 entre ellos.
- d) Ninguna de las anteriores.

sol c)

2. 2ptos El coeficiente de correlación lineal correspondiente a las variables X e Y, con el gráfico de dispersión siguiente, podría ser:



a) 0.62

b) 1.00

c) -0.01

d) -0.886

**solución correcta: d)**

3. 2ptos Indique cuál de las siguientes afirmaciones es falsa:

- a) La mediana y la media aritmética nunca coinciden, pero aportan información complementaria de la distribución.
- b) La media aritmética es una medida de posición central y en su cálculo interviene toda la información muestral.
- c) El valor de la media aritmética siempre está comprendido entre el menor y el mayor valor observado.
- d) La media aritmética es más sensible a los valores extremos o anómalos que la mediana.

sol a)

1. (4 puntos) El archivo de texto *datos.sulfato.nitrato.csv*<sup>2</sup> contiene información sobre medidas de sulfato y nitrato diarios en una estación de monitorización, concretamente las variables son:

Date:	fecha
sulfate:	concentración diaria de sulfato
nitrate:	concentración diaria de nitrato
ID:	indicador de estación de monitorización

```
datos <- read.table('http://dl.dropbox.com/u/29008031/datos.sulfato.nitrato.csv',header=TRUE,sep='  
head(datos); dim(datos); names(datos)
```

a) ¿Qué tipo de variables estadísticas son?.

b) Agrupar la variable *sulfato* en 4 subintervalos ([0,1], (1,2], (2,3], (3,5]). Da su distribución de frecuencias completa. Dar la representación gráfica más adecuada manteniendo esos intervalos.

```
ss <- cut(datos$sulfate, breaks=c(0,1,2,3,5), include.lowest = TRUE, right = TRUE)  
# distribución de frecuencias absolutas  
table(ss)
```

<sup>1</sup> respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

<sup>2</sup>El archivo de datos *datos.sulfato.nitrato.csv* se puede descargar desde la url <http://dl.dropbox.com/u/29008031/datos.sulfato.nitrato.csv>

```
# distribución de frecuencias relativas
table(ss)/sum(table(ss))
# distribución de frecuencias acumuladas absolutas
cumsum(table(ss))
# distribución de frecuencias acumuladas relativas
cumsum(table(ss))/sum(table(ss))
# densidade de frecuencia
ampl <- c(1,1,1,2)
dens <- table(ss)/sum(table(ss))/ampl
```

```
# Representación gráfica
hist( datos$sulfate, breaks=c(0,1,2,3,5),freq=FALSE)
```

- c) Resume numéricamente la variable. Cuantos días sobrepasó el valor de 4. Responder a ambas preguntas con la variable original y agrupada.

```
# Variable orixinal
summary(datos$sulfate); sd(datos$sulfate) # interpretación da media, ....
sum(datos$sulfate>4)
# Variable agrupada
marca <- c(0.5,1.5,2.5,4)
ss.arg <- rep( marca, table(ss))

summary(ss.arg); sd(ss.arg) # interpretación da media, ....
# Como o 4 está no medio do subintervalo (3,5], entón a frec. relativa é:
table(ss)[4]/2/sum(table(ss))
```

- d) Que valor es el más plausible de concentración de sulfato agrupado.

```
# Está a pedir a moda. Como é dunha v. agrupada con subintervalos con
# lonxitude distinta teño que coller as densidades de frecuencia
# Intervalo modal,
which.max(dens) # o 1º intervalo [0,1], moda:
0 + dens[2]/dens[2] * 1
```

# Prueba de Estadística descriptiva

Apellidos:

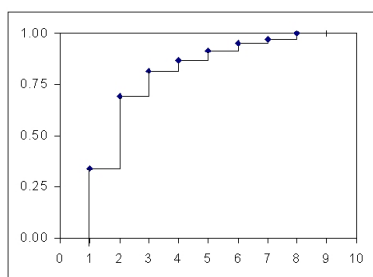
Nome:

DNI:

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas<sup>3</sup>

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

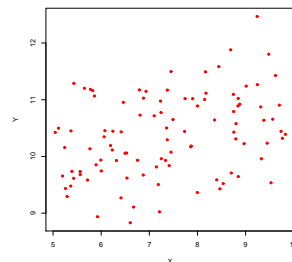
1. (2 puntos) La siguiente distribución de frecuencias relativas acumuladas corresponde a la variable  $X = \text{'número de plantas de los edificios censados en determinada localidad'}$ . Es cierto que:



- a) El 50 % de los edificios tienen más de 4 plantas
- b) Como mínimo el 50 % de los edificios tienen entre 1 y 3 plantas
- c) El 80 % de los edificios tienen exactamente 4 plantas
- d) El 75 % de los edificios tienen más de 3 plantas

**sol b)**

2. (2 punto) El coeficiente de correlación lineal correspondiente a las variables  $X$  e  $Y$ , con el gráfico de dispersión siguiente, podría ser:



- a) 0.75
- b) 0.086
- c) -0.82
- d) -0.37

**solución correcta: b)**

3. (2 punto) Si la desviación típica de una variable  $X$  es 2, ¿Cuánto vale la desviación de la variable  $Y = 1 + 2X$ ?

- a) 2
- b) 4
- c) 5
- d) 8

**solución correcta b)**

1. (4 puntos) El archivo de texto *datos.sulfato.nitrato.csv*<sup>4</sup> contiene información sobre medidas de sulfato y nitrato diarios en una estación de monitorización, concretamente las variables son:

Date:	fecha: <i>cualitativa ordinal</i>
sulfate:	concentración diaria de sulfato: <i>v. continua</i>
nitrate:	concentración diaria de nitrato: <i>v. continua</i>
ID:	indicador de estación de monitorización: <i>cualitativa ordinal</i>

```
datos <- read.table('http://dl.dropbox.com/u/29008031/datos.sulfato.nitrato.csv',header=TRUE,sep='  
head(datos); dim(datos); names(datos)
```

- a) ¿Qué tipo de variables estadísticas son?.

<sup>3</sup>1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

<sup>4</sup>El archivo de datos *datos.sulfato.nitrato.csv* se puede descargar desde la url <http://dl.dropbox.com/u/29008031/datos.sulfato.nitrato.csv>

- b) Agrupar la variable *nitrate* en 4 subintervalos  $([0, 1], (1, 2], (2, 3], (3, 5])$ . Da su distribución de frecuencias completa. Dar la representación gráfica más adecuada manteniendo esos subintervalos.

```
ss <- cut(datos$nitrate, breaks=c(0,1,2,3,5), include.lowest = TRUE, right = TRUE)
# distribución de frecuencias absolutas
table(ss)
# distribución de frecuencias relativas
table(ss)/sum(table(ss))
# distribución de frecuencias acumuladas absolutas
cumsum(table(ss))
# distribución de frecuencias acumuladas relativas
cumsum(table(ss))/sum(table(ss))
# densidade de frecuencia
ampl <- c(1,1,1,2)
dens <- table(ss)/sum(table(ss))/ampl

# Representación gráfica
hist( datos$nitrate, breaks=c(0,1,2,3,5),freq=FALSE)
```

- c) Resume numéricamente la variable. Cuantos días el valor fue inferior a 1. Responder a ambas preguntas con la variable original y agrupada.

```
# Variable orixinal
summary(datos$nitrate); sd(datos$nitrate) # interpretación da media, ....
mean(datos$nitrate<1)
# Variable agrupada
marca <- c(0.5,1.5,2.5,4)
ss.arg <- rep( marca, table(ss))

summary(ss.arg); sd(ss.arg) # interpretación da media, ....
# Como o 1 é o límite sup dun intervalo entón entón a frec. relativa é:
table(ss)[1]/sum(table(ss))
```

- d) Una zona se clasifica de riesgo si el 10 % de los días de mayor concentración de nitrato este valor supera el 1.75. Responde si esta zona es de riesgo o no.

```
quantile(datos$nitrate, prob=0.9) # polo tanto non é de risco porque non supera o 1.75
```

# Prueba de Estadística descriptiva

Apellidos:

Nome:

DNI:

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas<sup>5</sup>

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

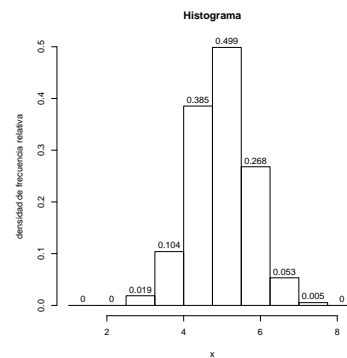
1. (2 puntos) ¿Cuál de las afirmaciones siguientes son ciertas?

- a) El coeficiente de correlación lineal y las varianzas de las variables siempre toman el mismo signo.
- b) La covarianza, al igual que el coeficiente de correlación lineal, toma valores entre -1 y 1.
- c) Al coeficiente de correlación lineal no le afectan los cambios de origen.
- d) Ninguna de las anteriores.

2. (2 puntos) Se sabe que el montaje de unas determinadas piezas requiere un promedio de 4 horas. Si la media del taller que monta el 40 % de las piezas es 4,45, ¿cuál es la media del resto de talleres?

- a) 4.375
- b) 3.325
- c) 4.25
- d) 3.7

3. (2 pts) En el histograma que aparece a continuación se muestran las densidades de frecuencia relativa de una variable continua agrupada en intervalos de longitud 0.75. Si  $N=1000$ , ¿cuántos valores hay en el intervalo  $(3,25, 4]$ ?



- a) 104
- b) 78
- c) 7.8
- d) Ninguno de los anteriores

1. (6 puntos) El archivo de texto *polen.csv*<sup>6</sup> contiene información sobre mediciones de granos de polen diarios en la ciudad de Ourense, concretamente las variables son:

día:	
polen.h:	granos de polen por $m^3$ de aire aspirado en el recuento horizontal
polen.v:	granos de polen por $m^3$ de aire aspirado en el recuento vertical

- a) Agrupar la variable *polen.h* en 6 subintervalos  $([0, 50], (50, 150], (150, 250], (250, 350], (350, 450], (450, \max])$ . Da su distribución de frecuencias completa. Dar la representación gráfica más adecuada manteniendo esos subintervalos.
- b) Resume numéricamente la variable. Si concentraciones superiores a 30 *granos*/ $m^3$  pueden provocar síntomas graves alérgicos en la población, cuantos días se sobrepasó este valor en la ciudad de Ourense, responder a ambas preguntas con la variable original y agrupada.
- c) Una zona se clasifica de riesgo alérgico si el 10 % de los días de mayor concentración de granos este valor no supera el 80 *granos*/ $m^3$ . Responde si la ciudad de Ourense es o no de riesgo alérgico.

<sup>5</sup>1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

<sup>6</sup>El archivo de datos *polen.csv* se puede descargar desde la url <http://dl.dropbox.com/u/29008031/polen.csv>

# Prueba de Estadística descriptiva

Apellidos:

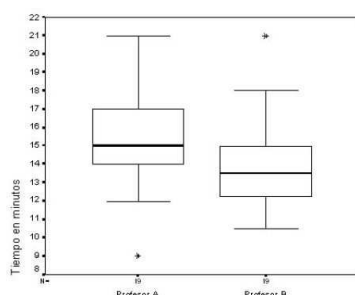
Nome:

DNI:

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas<sup>7</sup>

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

1. (2 puntos) En el grafico de cajas siguiente se muestra el salario mensual para hombres y mujeres de una determinada empresa. ¿Cuál de las afirmaciones siguientes son ciertas?



- a) Los hombres tienen un salario superior a las mujeres en la empresa
- b) Por lo menos el 75 % de los hombres tienen un salario superior al salario del 75 % de las mujeres.
- c) El 50 % de las mujeres tiene un salario inferior al 25 % de los hombres con un salario menor.
- d) Ninguna de las anteriores

2. (2 puntos) Si la varianza de una variable  $X$  es 2, ¿Cuánto vale la varianza de la variable  $Y = 1 + 2X$ ?

- a) 2
- b) 4
- c) 5
- d) 8

3. (2 pts) Del sector automovilístico tenemos la información de los Precios (agrupados en intervalos) y del Número de unidades vendidas en este último año. Si con respecto al año anterior los precios de los automóviles se han mantenido y los deciles 7, 8 y 9 se han desplazado considerablemente a la derecha, entonces podemos afirmar que en el último año:

- a) Se ha vendido un mayor porcentaje de coches de gama baja.
- b) Se ha vendido un mayor porcentaje de coches de gama media
- c) Se ha vendido un mayor porcentaje de coches de gama alta.
- d) Como los precios no han variado los deciles no pueden haber aumentado.

1. (6 puntos) Una muestra de comerciantes de material electrónico reveló las siguientes ventas y el número de empleados del año pasado:

Ventas (miles de euros) \ n° empleados	2	5	10	15
100 a 160	30	50	20	10
160 a 180	20	35	40	20
180 a 220	10	60	65	40
220 a 300	6	45	45	90

- a) La distribución de la variable ventas. Da su distribución de frecuencias completa y la venta más frecuente.
- b) La distribución de la variable número de empleados y el número de empleados medio.
- c) Da una medida de variabilidad de la variable *Ventas en miles de Euros*. Misma medida pero de la *v. Ventas en Euros*.

<sup>7</sup>1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

# Prueba de Estadística descriptiva

Apellidos:

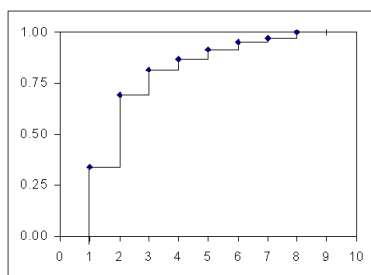
Nome:

DNI:

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas<sup>8</sup>

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

1. (2 puntos) La siguiente distribución de frecuencias relativas acumuladas corresponde a la variable  $X = \text{número de plantas de los edificios censados en determinada localidad}$ . Es falso que:



- a) El 50 % de los edificios tienen más de 4 plantas
- b) Como mínimo el 50 % de los edificios tienen entre 1 y 3 plantas<sup>o</sup>
- c) El 50 % de los edificios tienen menos de 3 plantas
- d) Ninguna de las anteriores

**sol a)**

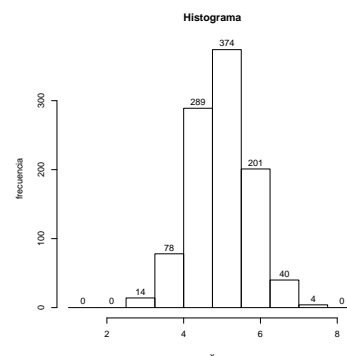
2. (2 puntos) Si el rango intercuartílico de una variable  $X$  es 2, ¿Cuánto vale el rango de la variable  $Y = 1 + 2X$ ?

- a) 2
- b) 4

- c) 5
- d) Ninguno de los anteriores

**solución correcta: RIQ4**

3. (2 pts) En el histograma que aparece a continuación se muestran las frecuencias absolutas de una variable continua agrupada en intervalos de longitud 0.75. Si  $N=1000$ , ¿cuántos valores hay mayores que 5.25?



- a) 500
- b) 619
- c) 381
- d) Ninguno de los anteriores

**sol) 1000-14-78-289- 374/3 \* 2=369.667**

1. (6 puntos) El archivo *tempos.csv* que está en dropbox<sup>9</sup> contiene información sobre el tiempo dedicado a la materia Estadística cada semana. Concretamente

curso:	
materia:	
semana:	
grupo:	
temp:	tiempo expresado en minutos

```
datos <- read.table('http://dl.dropbox.com/u/29008031/tempos.csv',header=TRUE,sep=';',dec='.')
head(datos); dim(datos); names(datos)
```

<sup>8</sup>1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

<sup>9</sup>El archivo de datos *tempos.csv* se puede descargar desde la url <http://dl.dropbox.com/u/29008031/tempos.csv>

- a) Describe numéricamente y gráficamente las v. *temp* y *semana*. Interpreta los valores dados. ¿Cuál es el tiempo mínimo dedicado a la materia del 80 % de los alumnos que más tiempo dedican?.

```
# Variable temp
summary(datos$temp); sd(datos$temp) # interpretación da media, ....
hist(datos$temp, freq=FALSE)
# Variable temp
summary(datos$semana); sd(datos$semana) # interpretación da media, ....
barplot(table(datos$semana)/sum(table(datos$semana))) #frecuencias relativas
# Cuantil 0.20
quantile(datos$temp, prob=0.2) # sol: 60 minutos
```

- b) Agrupa la variable *temp* en intervalos de longitud 1 hora. Resume numéricamente la variable agrupada. Da la frecuencia relativa por minuto de estudio y por intervalo.

```
ss <- cut(datos$temp, breaks=60*(0:ceiling(max(datos$temp)/60)), include.lowest = TRUE, right

# Variable agrupada
marca <- 60*(1:ceiling(max(datos$temp)/60)) - 30
ss.arg <- rep( marca, table(ss))
summary(ss.arg); sd(ss.arg) # interpretación da media, ....
# densidade de frecuencia
ampl <- 60
dens <- table(ss)/sum(table(ss))/ampl
```

- c) Da una medida de dispersión de la variable *tiempo agrupado para los grupos 1 y 2*.

```
datos2 <- subset(datos, subset=(datos$grupo=='EST-1')|(datos$grupo=='EST-2'))
sd(datos2$temp) # sin agrupar
ss2 <- cut(datos2$temp, breaks=60*(0:ceiling(max(datos2$temp)/60)), include.lowest = TRUE, right
marca <- 60*(1:ceiling(max(datos2$temp)/60)) - 30
ss2.arg <- rep( marca, table(ss2))
sd(ss2.arg) # agrupado
```