

**Examen de Estadística Descriptiva y Análisis de Datos**  
**Diplomatura de Estadística**  
**16 de febrero de 2007**

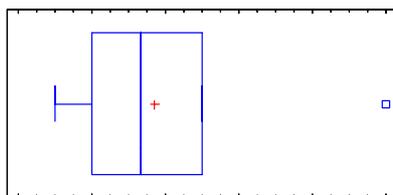
**Realiza los cálculos intermedios con 4 decimales y redondea el resultado final a 2 decimales.**

1. Una determinada empresa sospecha que sus empleados se conectan demasiadas veces a internet con fines particulares. Por esto toma una muestra de 10 empleados y controla cuántas veces al día miran su correo electrónico. Los datos son:

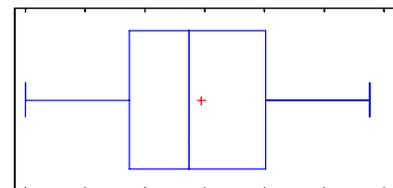
2      5      6      3      10      4      3      2      3      1.

- a) Calcula la media aritmética, la mediana, la desviación típica y la MEDA de este conjunto de datos.
- b) Representa este conjunto de datos mediante un diagrama de caja (box-plot). ¿Se detecta algún valor atípico? A la vista de estos resultados, justifica qué medidas de tendencia central y de dispersión sería preferible utilizar.
- c) ¿Qué se puede decir de la simetría o asimetría de estos datos? Calcula algún coeficiente de asimetría que consideres oportuno.
- d) Los siguientes diagramas de caja muestran los datos transformados. Si llamamos  $X$  a la variable original, ¿qué diagrama (A, B, C o D) crees que corresponde a cada una de las transformaciones siguientes:  $X^2$ ,  $\log(X)$ ,  $X^{1/2}$ ,  $1/X$ ?

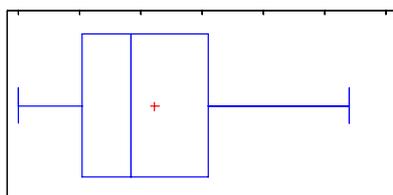
A



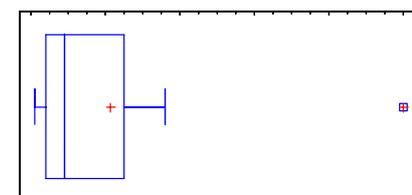
B



C



D



2. Como parte de un estudio sobre la evolución de los precios de determinados alimentos se ha realizado una encuesta a 1500 establecimientos de Madrid para obtener datos sobre distintos productos. La tabla siguiente contiene los resultados referidos al precio de un litro de leche y de una docena de huevos, ambos de una determinada marca:

Precio del litro de leche (€)	Precio de la docena de huevos (€)				Total
	[0.75,0.85)	[0.85,0.89)	[0.89,0.93)	[0.93,0.99)	
[0.54,0.58)	97	15	2	0	114
[0.58,0.59)	184	134	29	3	350
[0.59,0.60)	148	195	87	19	449
[0.60,0.61)	85	118	146	38	387
[0.61,0.65)	12	43	87	58	200
Total	526	505	351	118	1500

- a) ¿Qué producto es en media más caro?
- b) ¿Qué producto presenta una mayor variación en el precio? Justifica qué medida utilizas.
- c) ¿Cuál es el precio más frecuente de un litro de leche?
- d) De los establecimientos que venden la docena de huevos por debajo de 0.90 €, ¿qué porcentaje venden el litro a más de 0.60 €?
- e) ¿Cuál es el precio máximo de la docena de huevos para el 70% de los establecimientos? ¿Y el precio mínimo?
- f) De las medidas siguientes:
  - g1) coeficiente de correlación de Spearman,
  - g2) V de Cramer,
  - g3) coeficiente de correlación lineal de Pearson,
  - g4)  $\tau_B$  de Kendall,
  - g5) coeficiente de contingencia de Pearson,

justifica cuáles **no** se pueden utilizar para determinar si existe algún tipo de relación entre el precio del litro de leche y el precio de la docena de huevos.

3. En un estudio sobre olmos se pretende predecir el diámetro de su tronco en función de la longitud del mismo. Para ello, a partir de 30 ejemplares se calculan las siguientes medidas:

	diámetro	longitud
media	0.4265	0.5417
Coef. varianción (%)	13.6964%	12.3614%
mínimo	0.335	0.44
máximo	0.56	0.705

Si el coeficiente de determinación (en %) vale 92.8523%,

- ¿Cuánto vale el coeficiente de correlación lineal de Pearson? Justifica la respuesta.
- Calcula los coeficientes de la recta de regresión entre ambas variables e interprétalos.
- Valora la bondad del ajuste de la recta de regresión. ¿Qué porcentaje de variabilidad queda sin explicar?
- ¿Cuál es el diámetro esperado para los siguientes olmos? ¿Son fiables las predicciones?

Olmo	Longitud
1	0.30
2	0.595
3	0.705

- Calcula cuál es el error de predicción para un olmo de longitud 0.595 y diámetro 0.495.