

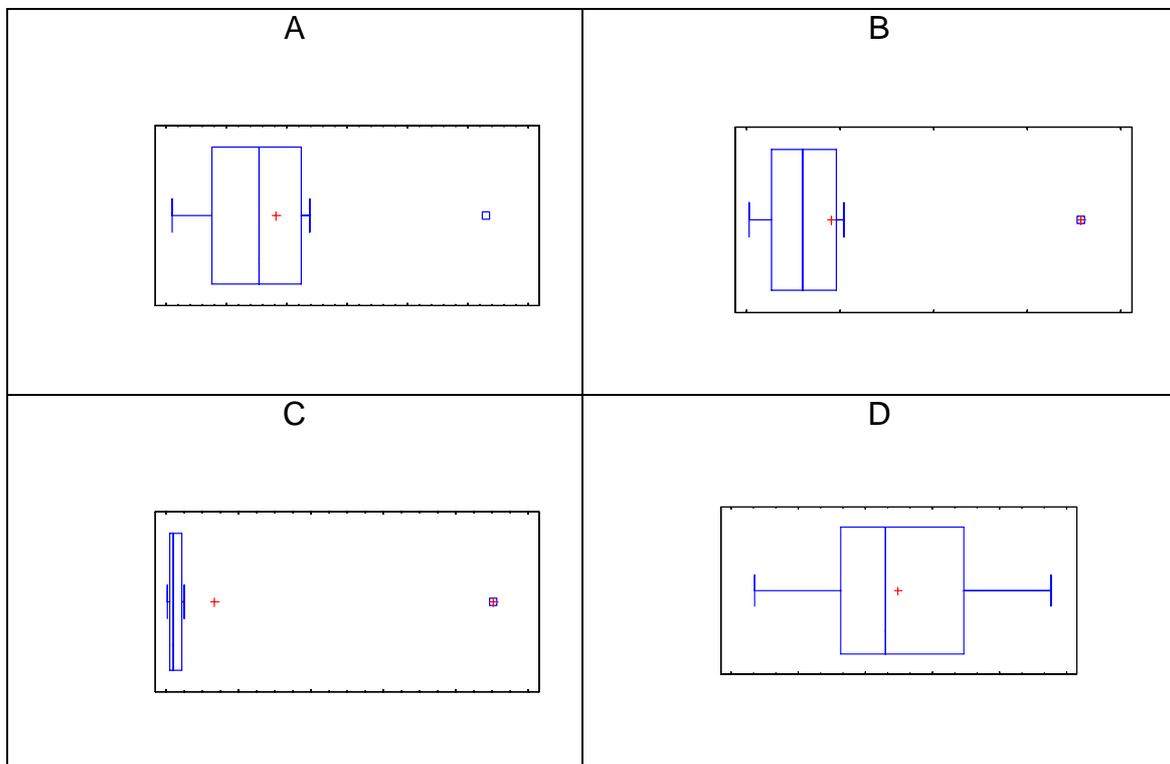
Examen de Estadística Descriptiva y Análisis de Datos
Diplomatura de Estadística
7 de febrero de 2006

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

1. (3 puntos) El número de días que 9 trabajadores escogidos al azar de una empresa han estado de baja son los siguientes:

15 7 8 85 19 12 8 22 14.

- a) Calcula la media aritmética, la mediana y la desviación típica 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 medida de tendencia central sería preferible utilizar.
- c) ¿Qué se puede decir de la asimetría de estos datos? Calcula algún coeficiente de asimetría que creas oportuno.
- d) ¿Qué transformaciones conoces que creas que puedan ser útiles para simetrizar estos datos?
- e) 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 , $\log(X)$, $X^{1/2}$, $1/X$?



2. (5 puntos) En una encuesta sobre los hábitos de los españoles se han obtenido los siguientes resultados sobre el número de horas semanales que miran la televisión:

	Número de horas					
	[0,20]		(20,30]		(30,40]	
Edad	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
[15,25]	1	1	4	6	10	9
(25,45]	10	18	6	6	3	1
(45,65]	2	1	1	1	9	11

- ¿De qué tipo son las variables que intervienen en este estudio? Calcula sus distribuciones marginales.
- Calcula la mediana de edad para este grupo de personas.
- En media, ¿quién mira más la televisión, los hombres o las mujeres?
- ¿Qué porcentaje de hombres mira la televisión más de 20 horas a la semana?
- ¿Cuál es la edad más frecuente de las mujeres que miran la televisión como máximo 30 horas a la semana?
- ¿Qué variable presenta más dispersión? Justifica que medida de dispersión utilizas.

g) De las medidas siguientes:

- coeficiente de correlación de Spearman,
- V de Cramer,
- coeficiente de correlación lineal de Pearson,
- τ_B de Kendall,
- coeficiente de contingencia de Pearson,

¿cuál (o cuáles) crees que se podrían utilizar para determinar si existe algún tipo de relación entre la edad de los encuestados y el número de horas semanales que miran la televisión? Justifica por qué. Calcula la medida que creas más oportuna (solamente una) e interpreta el resultado.

h) De las medidas siguientes:

- coeficiente de correlación de Spearman,
- estadístico chi-cuadrado,
- D de Sommer,
- coeficiente de contingencia de Pearson,
- V de Cramer,

¿cuáles crees que se podrían utilizar para determinar si existe algún tipo de relación entre el sexo de los encuestados y el número de horas semanales que miran la televisión? Justifica por qué. Explica, de forma breve y sin hacer los cálculos, cuál sería el procedimiento a seguir.

3. (2 puntos) Como parte de la encuesta anterior, se han recogido datos de 100 familias sobre el número de televisores que tiene cada familia en casa (variable Y) y los ingresos mensuales (variable X, en euros). A partir de estos datos, se obtiene que estas familias tienen una media de 2 televisores (con una varianza de 0.04) y una media de ingresos mensuales de 930 euros (con una varianza de 2410). Teniendo en cuenta que el coeficiente de correlación lineal de Pearson vale 0.9,

- Calcula los coeficientes de la recta de regresión entre ambas variables.
- Valora la bondad del ajuste mediante algún coeficiente que te parezca adecuado para ello. ¿Qué porcentaje de variabilidad queda sin explicar?
- Supón que tienes unos vecinos con unos ingresos mensuales de 900 euros. ¿Cuántos televisores crees más probable que tengan en casa: uno o dos? Justifícalo.
- ¿Cuál de los siguientes gráficos (A, B, C o D) crees que corresponde a la nube de puntos asociada a estos datos? Justifica tu respuesta.

