

**Examen de Estadística Descriptiva y Análisis de Datos**  
**Diplomatura de Estadística- Grupo 25**  
**1 de septiembre de 2006**

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

1. (3 puntos) La siguiente tabla contiene el número de defunciones de mujeres por agresión (homicidios) en las provincias españolas durante el año 2004 (según datos del INE).

defunciones	provincias	$N_i$	$f_i$
0	15		0,2885
1	10	25	
2		33	
3	6	39	
4			0,0769
5			
6	1	48	0,0192
7	1	49	0,0192
11	1 (Barcelona)	50	0,0192
16	1 (Valencia)	51	0,0192
91	1 (Madrid)	52	0,0192

- Completa los valores que faltan en la tabla anterior. Redondea a valores enteros los resultados que obtengas para la columna "provincias".
- Dibuja el diagrama de caja (box-plot).
- Interpreta el diagrama de caja: ¿qué representa la caja? ¿qué tipo de asimetría (si es que la hay) presenta esta variable? ¿hay valores atípicos (di cuáles son, si es que los hay)?
- Entre las siguientes medidas, escoge una de centralización y una de dispersión que sean útiles para resumir estos datos. Justifica por qué.
  - media aritmética,
  - desviación típica,
  - varianza,
  - mediana,
  - media geométrica,
  - rango intercuartílico,
  - coeficiente de variación de Pearson,
  - moda.

2. (4 puntos) Una empresa fabricante de teléfonos móviles ha realizado un amplio estudio de mercado en la ciudad de Madrid con el fin de conocer sus expectativas de crecimiento. Uno de los frutos de este estudio es la siguiente tabla de contingencia donde se relaciona la variable sexo con el gasto mensual en llamadas telefónicas (en euros), obtenida a partir de 500 encuestas:

Gasto	Hombres	Mujeres
[0,3)	26	25
[3,6)	33	51
[6,12)	42	62
[12,30)	54	51
[30,60)	67	39
[60,120)	28	22

- ¿De qué tipo son las variables que intervienen en este estudio? Calcula sus distribuciones marginales.
- Calcula la mediana de gasto para este grupo de personas.
- Compara el gasto medio en llamadas de hombres y mujeres.
- ¿Cuál es el gasto mensual más frecuente entre los hombres?
- ¿Qué porcentaje de las personas que hacen un gasto mensual inferior a 12 euros son hombres?
- ¿Qué porcentaje de mujeres hacen un gasto de entre 20 y 40 euros?
- De las medidas siguientes:
  - coeficiente de correlación de Spearman,
  - V de Cramer,
  - coeficiente de correlación lineal de Pearson,
  - $\tau_B$  de Kendall,
  - coeficiente de contingencia de Pearson,
  - estadístico chi-cuadrado,
  - D de Sommer.

¿Cuáles crees que se podrían utilizar para determinar si existe algún tipo de relación entre el sexo de los encuestados y el gasto mensual en llamadas? Calcula las que creas oportunas e interpreta los resultados.

**3. (3 puntos)** Como complemento de los resultados anteriores se desea ajustar una recta de regresión entre el gasto mensual en llamadas (G) y la edad (E) teniendo en cuenta que la media del gasto mensual es 25.73 y su varianza es 690.31, mientras que la variable edad tiene una media de 47.23 y una varianza de 136.78. Además también se sabe que  $\sum_{i=1}^{500} E_i G_i = 688758$ .

- a) Calcula e interpreta el coeficiente de correlación lineal de Pearson.
- b) Encuentra la recta de regresión que expresa el gasto mensual en llamadas en función de la edad de los encuestados.
- c) Utilizando los resultados del apartado anterior, calcula una predicción para el gasto mensual en llamadas de una persona de 30 años y haz una valoración de la fiabilidad de esta predicción.
- d) Si quisiéramos utilizar las pesetas como unidad de referencia, ¿cuáles serían los coeficientes de la recta de regresión? (1 euro=166 pts). Justifica tu respuesta.