



**Ingeniero Técnico de Minas**  
**Asignatura: Estadística**  
**Curso 2005/2006**  
**Hoja 1. Estadística Descriptiva**

1. Para tener una buena imagen de la pantalla del ordenador es necesario que la tensión de la rejilla metálica situada detrás de la pantalla no sea ni demasiado alta ni demasiado baja. Por este motivo, durante la producción el fabricante controla la tensión de dicha rejilla. Los siguientes resultados corresponden a estas mediciones sobre 300 rejillas:

<i>Mediciones de la tensión</i>	<i>Nº de rejillas</i>
$250 \leq X < 270$	120
$270 \leq X < 290$	70
$290 \leq X < 310$	50
$310 \leq X < 330$	30
$330 \leq X < 350$	20
$350 \leq X < 370$	10

Se pide:

- (a) Establecer la tabla de las frecuencias relativas, frecuencias absolutas acumuladas y frecuencias relativas acumuladas.
- (b) ¿Qué porcentaje de rejillas de la muestra tienen una tensión inferior a 330? ¿y una tensión superior o igual a 310?. ¿Qué porcentaje de rejillas tienen una tensión comprendida en el intervalo  $270 \leq X < 350$ ?, ¿y en el intervalo  $290 \leq X < 390$ ?
- (c) Calcular la tensión media de las rejillas analizadas, la varianza, la desviación típica y el coeficiente de variación. ¿En qué intervalo se sitúa la mediana? ¿Cuál es el intervalo modal?
- (d) Representar gráficamente la variable mediante un histograma. Comentar las características más relevantes de dicho histograma. ¿Qué medidas de centralización y dispersión son más adecuadas para resumir los datos? Razonar la respuesta.
2. **(Propuesto)** Los responsables de calidad realizan periódicamente un análisis de la medición del trabajo con el fin de determinar el tiempo requerido para generar una unidad de producción. En una planta de procesamiento se registró durante 20 días el número de horas-obrero totales requeridas para realizar cierta tarea. Los datos recogidos son

128	116	95	97	124	109	124	132	100	111
145	125	106	135	119	131	133	131	112	135

- (a) Realiza el diagrama de caja y bigotes para los datos del tiempo de respuesta de la base de datos. Comenta sus características más relevantes.

- (b) Construye el histograma correspondiente a los datos seleccionando 5 clases.
- (c) Obtener la tabla de frecuencias absolutas, relativas y acumuladas usando las mismas clases del apartado anterior.
- (d) Calcula la media, mediana y moda a partir del histograma.
- (e) En base al histograma estudia la simetría o asimetría de la distribución.
- (f) Decide qué medida de posición puede ser representativa y calcula una medida de dispersión asociada a la medida de posición anterior.

3. **(Propuesto)** Con el fin de determinar si existe relación entre la cantidad de polímeros de látex incluida durante el proceso de mezclado de cemento Portland y su resistencia adhesiva a tensión, una empresa encargada de realizar certificaciones de obras toma una muestra de tamaño 10, obteniendo los siguientes resultados:

<i>Polímeros latex (mgr/kg)</i>	13.5	11.0	13.0	11.2	12.0	13.2	12.0	13.5	11.2	13.0
<i>Resistencia (kgf/cm<sup>2</sup>)</i>	17.5	16.6	17.2	16.6	17.0	17.3	16.9	17.3	16.8	17.1

- (a) Calcular la media y varianza asociada a cada una de las variables.
- (b) Calcular la covarianza existente entre ambas variables así como el coeficiente de correlación.
- (c) Realizar un ajuste por mínimos cuadrados de la resistencia respecto a la cantidad de polímeros añadida en la mezcla.
- (d) Deducir, suponiendo que la relación proporcionada por las rectas de regresión es válida:
- i. El valor estimado para la resistencia si la cantidad de polímero agregado es de  $11.5mgr/kg$ .
  - ii. Si un determinado constructor desea que la argamasa tenga una resistencia de  $16.5kgf/cm^2$ , calcular a partir de la recta de regresión correspondiente, cuál debe de ser la cantidad de polímero de látex que se debe añadir.
4. Con el fin de determinar la cantidad de contaminación eliminada del aire, a partir de la cantidad de precipitación pluvial, se realizaron las correspondientes mediciones en 15 días de lluvia. La media y la varianza de las precipitaciones recogidas durante los 15 días de lluvia son 40 y 4, respectivamente, y la recta ajustada por mínimos cuadrados para los datos es:

$$y = 0.25 + 0.01x$$

- (a) Calcular la covarianza entre la cantidad de lluvia y la cantidad de contaminación eliminada
- (b) ¿Cuál es el signo de la correlación entre la cantidad de lluvia y la cantidad de contaminación eliminada? Justifica tu respuesta. ¿Qué interpretación tiene dicho signo?
- (c) ¿Cuánto vale la media aritmética de las cantidades de contaminación eliminadas durante los 15 días de lluvia observados?

5. Los siguientes datos se refieren al crecimiento de una colonia de bacterias en un medio de cultivo:

$x$	3	6	9	12	15	18
$y$	115000	147000	239000	356000	579000	864000

siendo " $x$ " los días desde la inoculación e " $y$ " el número de bacterias.

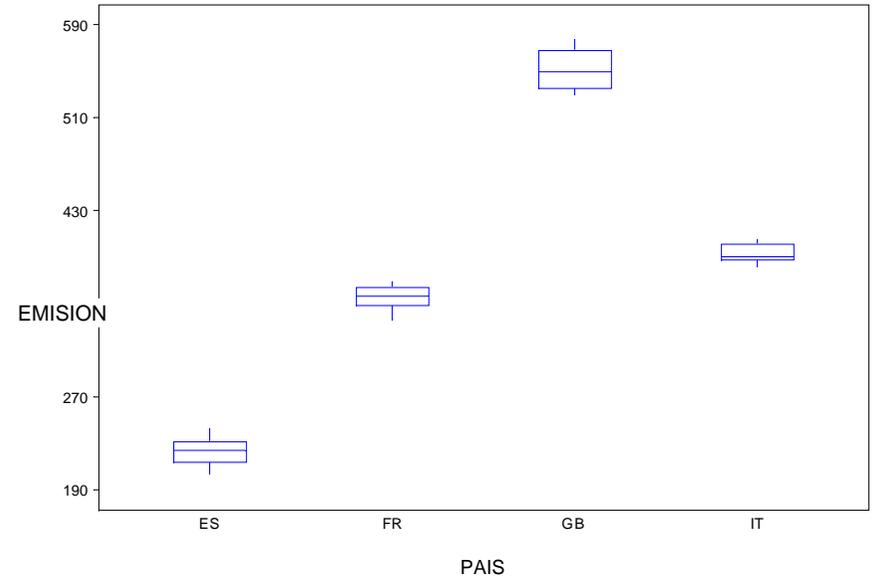
- Representar  $y$  en función de  $x$  para verificar que es razonable ajustar a una curva exponencial.
  - Ajustar a una curva exponencial estos datos.
  - Estimar el número de bacterias al término de 20 días, así como el número de bacterias al comienzo del experimento.
  - ¿Cuál sería el número de bacterias después de 50 días desde la inoculación?
6. En la dirección <http://dataservice.eea.eu.int/dataservice/>, se pueden encontrar los datos de emisión de CO2 por fuentes fósiles para los distintos países europeos entre los años 1985 y 1997. Se indican a continuación los datos de emisión total de este contaminante para España entre los años 1990 y 1997.

Año ( $X$ )	90	91	92	93	94	95	96	97
Emisión total ( $Y$ ) ( Megatoneladas)	203.8	215.5	225.1	211.5	222.8	237.5	226.6	242.8

Se pueden utilizar las cantidades numéricas siguientes:

$$\sum x_i = 748; \sum x_i^2 = 69980; \sum y_i = 1785.6; \sum y_i^2 = 399722; \sum x_i y_i = 167142$$

- Realizar el ajuste lineal de la emisión total en función del año. Calcular el valor de  $R^2$  y comentar la bondad del ajuste.
  - Según nuestro modelo, ¿se está produciendo un aumento o un descenso de la emisión de CO2?
  - Según nuestro modelo, ¿cuál sería la cantidad de CO2 emitida en 1998 por España?
7. (**Propuesto**) A continuación se presentan los diagramas de caja-bigotes para la emisión total de cuatro países de la Unión europea entre 1990 y 1997. ( ES="España", FR="Francia", GB="Gran Bretaña", y IT="Italia").



- ¿Cómo se construye un diagrama de caja-bigotes?
  - ¿Cómo clasificaría estos cuatro países en cuanto a contaminación por CO2?
  - ¿Cuál es el país que presenta mayor dispersión entre sus datos de contaminación? ¿Qué quiere decir?
  - Si se realiza un ajuste lineal de la emisión de CO2 en función del año para cada uno de los países, obtenemos las rectas siguientes:
 

España	Emisión= $-196.4 + 4.48$ año
Francia	Emisión= $403.2 - 0.51$ año
Gran Bretaña	Emisión= $1126.2 - 6.15$ año
Italia	Emisión= $192.9 + 2.143$ año
- ¿Qué países han ido reduciendo sus emisiones entre 1990 y 1997? ¿Qué países las han incrementado? ¿Cuál es el país que más ha reducido sus emisiones de CO2? ¿Cuál es el país que más las han incrementado?