



Ingeniero Técnico de Telecomunicaciones (Telemática)

Asignatura: Estadística

Curso 2005/2006

Hoja 1. Estadística Descriptiva

- 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:

Mediciones de la tensión	Nº de rejillas
$250 \leq X < 270$	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:

- Establecer la tabla de las frecuencias relativas, frecuencias absolutas acumuladas y frecuencias relativas acumuladas.
 - ¿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 $290 \leq X < 390$?
 - 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?
 - 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.
- (Propuesto)** Se ha aplicado un test sobre capacitación laboral a 90 empleados de una industria, obteniéndose los siguientes resultados:

PUNTUACIONES	Nº de EMPLEADOS
[38, 44)	7
[44, 50)	8
[50, 56)	15
[56, 62)	27
[62, 68)	18
[68, 74)	9
[74, 80]	6

Se pide:

- 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? Razona tu respuesta.
 - Calcular la puntuación media obtenida en el test. ¿En qué intervalo se sitúa la mediana? ¿Cuál es el intervalo modal?
 - ¿Qué proporción de empleados tiene una puntuación mayor o igual a 62? ¿y entre 44 y 68?
- (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

- Construye el histograma correspondiente a los datos seleccionando 5 clases.
 - Obtener la tabla de frecuencias absolutas, relativas y acumuladas usando las mismas clases del apartado anterior.
 - Calcula la media, mediana y moda a partir del histograma.
 - En base al histograma estudia la simetría o asimetría de la distribución.
 - Decide qué medida de posición puede ser representativa y calcula una medida de dispersión asociada a la medida de posición anterior.
- Con el fin de analizar el tiempo de respuesta (en milisegundos) de una base de datos de consulta se tomaron 15 datos correspondientes a 3 semanas consecutivas obteniéndose los siguientes resultados:

	Lun	Mart	Miérc	Juev	Viern
Semana 1	4.32	7.14	9.21	9.71	15.39
Semana 2	5.2	8.37	9.34	10.46	18.9
Semana 3	6.39	8.97	9.51	10.53	21.25

- Calcular la media, mediana, cuartiles y desviación típica correspondiente a estos datos.
- ¿Entre qué valores podemos decir que se encuentran los datos no atípicos? ¿Existen datos que puedan considerarse atípicos?
- ¿Qué medida de dispersión utilizarías? Razona tu respuesta.
- Realiza el diagrama de caja y bigotes para los datos del tiempo de respuesta de la base de datos. Comenta sus características más relevantes.

- (e) Con el fin de determinar la relación entre el **tiempo de respuesta de la base de datos** y el **número de usuarios** se midieron simultáneamente a la obtención del tiempo de respuesta, el número de usuarios activos en ese instante, obteniéndose un conjunto de 15 datos cuyo valor medio era 30.73 y cuya varianza era 79.40. El estudio concluyó con la siguiente relación lineal entre ambas características:

$$Tiempo = 0.51 \cdot Usuarios - 5.42$$

- i. Determinar el coeficiente de determinación asociado al ajuste.
 - ii. Determinar el número de usuarios activos, si el tiempo de respuesta de servidor es de 32.
5. **(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/cm²)</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.
6. Con el fin de determinar la relación existente entre la resistencia de una determinada pieza de plástico y uno de sus componentes (componente A) se fabrican 10 piezas de prueba, cada una con una concentración distinta y se obtienen los siguientes resultados:

<i>Pieza</i>	1	2	3	4	5	6	7	8	9	10
<i>X (% A)</i>	1.5	1.2	1.1	1.0	4.5	5.2	8.7	9.0	9.2	9.5
<i>Y (Rotura)</i>	3.04	2.96	2.66	3.17	9.82	9.68	17.71	18.18	18.32	19.30

donde la variable rotura indica la fuerza empleada a tracción en el instante de su rotura.

A partir de los datos anteriores se obtienen las siguientes cantidades:

$$\sum x_i = 50.9; \sum x_i^2 = 384.77; \sum y_i = 104.84; \sum y_i^2 = 1577.53; \sum x_i y_i = 778.33$$

- (a) Usando las cantidades anteriores, realizar un ajuste lineal de la rotura en función del contenido en el componente A. Calcular el valor de r^2 y comentar la bondad del ajuste.
 - (b) Según nuestro modelo, ¿cuál sería la resistencia para un 3% de contenido de componente A? ¿y para un contenido del 28% de A?
7. Los siguientes datos se refieren al crecimiento de una colonia de bacterias en un medio de cultivo:

<i>x</i>	3	6	9	12	15	18
<i>y</i>	115000	147000	239000	356000	579000	864000

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

- (a) Representar y en función de x para verificar que es razonable ajustar a una curva exponencial.
 - (b) Ajustar a una curva exponencial estos datos.
 - (c) Estimar el número de bacterias al término de 20 días, así como el número de bacterias al comienzo del experimento.
 - (d) ¿Cuál sería el número de bacterias después de 50 días desde la inoculación?
8. **(Propuesto)** 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?