

TEMA 3: Correlación y Regresión

Examen Septiembre 2003.- Como parte de la evaluación de un sistema de computadoras, un gerente de sistemas quiere predecir el tiempo de respuesta de las terminales de computadora. El “tiempo de respuesta de una terminal” se define como el tiempo (T), en segundos, que la computadora tarda en responder a un comando enviado desde una terminal, Aunque son muchas las variables que influyen en el tiempo de respuesta de las terminales, el gerente de sistemas va a modelar dicho tiempo como una función del número de usuarios simultáneos (N) en el momento en que envía el comando, El gerente ha recabado los datos de estas dos variables en cinco instantes de tiempo, de forma que:

$$\sum_{i=1}^5 N_i = 15, \quad \sum_{i=1}^5 T_i = 4,6 \quad \sum_{i=1}^5 N_i T_i = 16,97,$$

$$\sum_{i=1}^5 N_i^2 = 55, \quad \sum_{i=1}^5 T_i^2 = 5,2826$$

- ¿Existe una relación lineal entre las variables? Usar algún coeficiente que indique cómo es esta relación.
- Calcular la recta de regresión por el método de mínimos cuadrados, que relaciona el tiempo de respuesta de una terminal con el número de usuarios simultáneos en es momento.
- Interpretar los parámetros de ajuste de regresión.
- Razonar si es posible que la varianza residual tome el valor de 0,3.

Examen Septiembre 2002.- Tenemos cinco alumnos de un programa de apoyo escolar en la asignatura de matemáticas. Se pretende establecer una relación entre los resultados académicos (obtenidos por el orientador escolar a través de un test psicopedagógico) que obtuvieron antes de entrar al programa (X) y los obtenidos ahora (Y). Los resultados son los siguientes:

X	a	-4	0	3	5
Y	3	1	2	6	5

En el caso de que “a” tome el valor de -1:

- ¿Cuál de las dos medidas está más dispersa?
- Determina y dibuja la recta de regresión de Y sobre X, así como el coeficiente de correlación. ¿Cómo se interpreta el resultado del coeficiente de correlación?
- Determina los resultados esperados para un alumno que se incorporó más tarde al programa y que obtuvo como valoración inicial 2.
- ¿Qué relación existe entre el coeficiente de regresión de Y sobre X y el coeficiente de correlación?
- Calcula los residuos obtenidos y la desviación típica residual.

- f) El psicopedagogo ha traspapelado el valor de “a”, pero recuerda que la desviación típica de la variable X coincidía con su coeficiente de variación. ¿Cuál es el valor de “a”?

Examen Junio 2001.- Un gestor financiero sostiene que existe una relación lineal directa entre las deudas totales de la empresa y sus recursos propios. Para comprobar la validez de esta creencia consideró una muestra piloto de 151 empresas, midiendo las deudas (en millones de pesetas) y el capital(en millones de pesetas). Obtuvo los siguientes resultados usando el programa Statgraphics:

Regresión Análisis = Linear model: $Y = a + b * X$

Dependent variable: deudas

Independent variable: recursos

Parameter	Estimate	Standard Error	T Statistic	p-value
Intercept	1.27581	5.33732	0.239035	0.8114
Slope	0.0327699	0.00194528	16.8458	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	p-value
Model	58673.5	1	58673.5	283.78	0.0000
Residual	30806.5	149	206.755		
Total (Corr.)	89480.0	150			

Correlation Coefficient = 0.809763

R-squared = 65.5716 percent

Standard Error of Est. = 14.379

- Según los resultados obtenidos, ¿parece razonable la hipótesis del gestor? ¿Cómo se interpreta el coeficiente $b = 0,0328$?
- Utilizando otro método, el gestor obtuvo la recta $Y = 1,0 + 0,02 X$, con una desviación típica residual igual a 25,5. Según el criterio de los mínimos cuadrados, ¿cuál de las dos rectas utilizaría para el estudio?
- Si se tiene la información de que una empresa dispone recursos que alcanzan los 5000 millones de pesetas, ¿cuál será aproximadamente la deuda de la empresa?

Examen Junio 1999.- Se dispone de las edades (años) y de los niveles de calcio (mg / kg) de 10 enfermos de osteoporosis. Los resultados son:

	1	2	3	4	5	6	7	8	9	10
Edad (x)	68	55	61	66	71	71	81	86	53	60
Calcio (y)	390	410	350	310	270	220	110	90	400	330

$$\sum_{i=1}^{10} x_i = 672 \quad \sum_{i=1}^{10} y_i = 2880 \quad \sum_{i=1}^{10} x_i y_i = 183320$$

$$\sum_{i=1}^{10} x_i^2 = 46174 \quad \sum_{i=1}^{10} y_i^2 = 949200$$

- Calcular el coeficiente de correlación entre edad y nivel de calcio.
- Hallar la recta de regresión.
- Predecir el nivel de calcio en los huesos para una persona de 65 años con osteoporosis .
- Predecir el nivel de calcio de una persona centenaria con osteoporosis utilizando la recta de regresión. ¿ La predicción le parece razonable? Comentar.
- Obtener la desviación típica residual.
- Van a seleccionar al azar a uno de los enfermos de la muestra para aplicarle un tratamiento nuevo, ¿Cuál es la probabilidad de que tenga mas de 65 años?

Examen Junio 1998.- Una clínica de adelgazamiento está realizando un estudio con el propósito de analizar la relación existente entre el número de horas dedicadas al deporte y el número de kilos perdidos. Sean las variables:

X: Número de horas mensuales dedicadas al deporte
Y: Número de kilos perdidos en un mes.

Los resultados obtenidos tras examinar a 80 pacientes son los siguientes:

$$\sum_{i=1}^{80} x_i = 359 \quad \sum_{i=1}^{80} y_i = 94,5 \quad \sum_{i=1}^{80} x_i y_i = 548,7$$

$$\sum_{i=1}^{80} x_i^2 = 2100 \quad \sum_{i=1}^{80} y_i^2 = 205,16$$

- Calcular el coeficiente de correlación lineal entre las variables x e y. Si aumenta el número de horas mensuales dedicadas al deporte, ¿ aumenta también el número de kilos perdidos? Razonar la respuesta.
- Obtener la recta de regresión e interpretar sus coeficientes.
- Supongamos que un nuevo cliente decide dedicar 4,5 horas mensuales al deporte. ¿ Cuántos kilos se espera que pierda en un mes?
- Hallar e interpretar la varianza residual.
- Definir la covarianza entre dos conjuntos de datos y enumerar sus propiedades
- ¿Qué relación existe entre covarianza y correlación? ¿ Y entre el coeficiente de correlación y la pendiente de la recta de regresión?

Examen Junio 1996.- Una empresa desea realizar un estudio estadístico sobre la relación existente entre la renta y el consumo correspondiente a seis familias españolas en el año 1995. Los datos obtenidos son los siguientes:

Renta (x_i)	3	3,4	4	4,9	6	4,1
Consumo (y_i)	2,6	2,2	3,1	4,3	5,2	3,5

- Calcular la recta de regresión.
- Hallar los residuos.
- Calcular la desviación típica residual.
- Hallar la desviación típica de y .
- ¿Cuándo coincide la desviación típica residual con la desviación típica de y ? Razona la respuesta.

Examen Junio 1995.- Los beneficios en millones de pesetas obtenidos este año por las sucursales en cinco ciudades de dos empresas que compiten por ganar cuota de mercado son los siguientes:

Empresa A	5	20	15	10	25
Empresa B	5	30	20	20	25

Se pide:

- Calcular la cantidad media en que los beneficios de las sucursales de empresa B superan a lo de las sucursales de la empresa A.
- Hallar el coeficiente de correlación entre los beneficios obtenidos por las sucursales de ambas empresas.
- Un ejecutivo de empresa B decide modelizar los beneficios de las sucursales de su empresa en función de los de las sucursales de la empresa A. Propone el siguiente modelo lineal:

$$y=a+bx$$

donde y = beneficios sucursales de B, y x = benéficos sucursales de A. Utilizando los datos disponibles, calcular a y b ajustando por mínimos cuadrados.

- Una sucursal de empresa A situada en una sexta ciudad ha conseguido 30 millones de pesetas de benéficos. Utilizando el modelo del apartado (c), predecir los beneficios obtenidos por una sucursal de la empresa B situada en la misma ciudad.