Estadística Descriptiva Bivariada y Regresión

Ejercicio 1)

En una fábrica de producción de envases de productos lácteos se ha recogido la siguiente información sobre los productos defectuosos y el tipo de máquina que los ha elaborado.

Para ello se ha recogido una muestra de 500 envases producidos por tres máquinas diferentes: A, B y C y se ha observado lo siguiente: la máquina A ha producido 195 productos no defectuosos y 4 defectuosos, la máquina B ha producido 125 productos correctos y 75 defectuosos, y la máquina C ha producido 95 productos sin defecto y 6 con defecto.

a) Representa los datos de la muestra en una tabla de doble entrada. Obtén las distribuciones de frecuencias absolutas de las variables marginales: calidad de los productos y tipo de máquina.
b) Obtén la distribución conjunta de frecuencias relativas.

Ejercicio 2)

Para estudiar la relación existente entre el precio y el número de habitaciones de una vivienda en Madrid disponemos de los datos referidos a 300 operaciones de venta, de los que se deduce que el precio medio es 250.000€, el coeficiente de variación del precio es 0,25, el número medio de habitaciones es de 3, el coeficiente de variación para esta última variable es 0,10 y finalmente el coeficiente de correlación entre ambas variables se sitúa en 0,8.

a) Calcula las desviaciones típicas de ambas variables.
b) Calcula la pendiente y el intercepto de la recta de regresión que explica el precio en función del número de habitaciones. Escribe la ecuación de la recta de regresión.
c) Determina cuánto debería pedirse, de acuerdo con el mercado, por una vivienda de dos habitaciones.
d) Efectúa una valoración del grado de ajuste del modelo a los datos disponibles usando el coeficiente de determinación R^2.

Ejercicio 3)

La siguiente tabla de correlación recoge los datos correspondientes a las variables en una muestra formada por 100 habitantes de una ciudad:

X: Renta anual (en miles de euros).
Y: Gasto anual en vacaciones (en cientos de euros).

	Y		
X	**[0,20)**	**[20,40)**	$n_{i.}$
[10,30)	15	5	20
[30,60)	15	35	50
[60,90)	5	25	30
$n_{.j}$	35	65	N=100

a) Halla las tablas de frecuencia marginales de cada variable.
b) Calcula las medias y desviaciones típicas de cada variable.
c) Calcula el coeficiente de correlación entre ambas variables, sabiendo que la covarianza muestral es $s_{XY} = 73$.
d) Predice (usando un modelo de regresión lineal) el gasto vacacional de un habitante de esta ciudad cuya renta es de 35.000 euros.

Ejercicio 4)

Para estudiar el nivel de vida de los hogares de una gran ciudad, se disponen los datos de 200 hogares relativos a las siguientes variables:

X: Nº de coches disponibles en el hogar.
Y: Nº de días al año, que pasan de vacaciones en el extranjero.

$$s_X = 5, \qquad s_Y = 3, \qquad s_{XY} = 15$$

a) Calcula el coeficiente de correlación.
b) Obtén la ecuación de la recta de regresión considerando la variable Y como variable dependiente de X.
c) Según este modelo, ¿cuántos días al año pasará una familia que dispone de tres coches, en el extranjero?

Ejercicio 5)

Con objeto de analizar si existe relación lineal entre el consumo de energía eléctrica (kW por hora), variable X, y el volumen de producción en millones de pesetas, variable Y, de una empresa se ha obtenido la siguiente información:

$$\bar{x} = 0.151, \qquad \bar{y} = 94.6, \qquad s_X = 0.055, \qquad s_Y = 56.248, \qquad s_{XY} = -2.87$$

a) Ajuste la línea de regresión lineal que explica el consumo de electricidad en función del volumen de producción.
b) Razone la validez de la recta ajustada.

Ejercicio 6)

De una determinada empresa se conocen los siguientes datos, referidos al volumen de ventas (en millones de pesetas) y al gasto en publicidad (en miles de pesetas) de los últimos 6 años:

Volumen de ventas(mill. Ptas)	**Gastos Publicidad(miles ptas.)**
10	**16**
15	**32**
20	**48**
22	**56**
30	**64**
32	**80**

a) Calcula las medias, desviaciones típicas, y la covarianza entre ambas variables.
b) Con lo calculado en el inciso anterior, ¿existe relación lineal entre las ventas de la empresa y sus gastos en publicidad? Razona la respuesta.
c) Halla la recta de regresión que pone al volumen de ventas en función de los gastos en publicidad. Para un gasto en publicidad de 60.000 pesetas, ¿qué volumen de ventas obtenemos?

Ejercicio 7)

Sabiendo que:

$$\bar{x} = 3, \qquad s_X^2 = 6, \qquad s_Y^2 = 8$$

Y que la recta de regresión de Y sobre X es:

$$y = 4 - 0.667x$$

a) Obtener la recta de regresión de X sobre Y.

TAREA

Se está estudiando la relación entre el número de años que una persona está afiliada al sindicato y el nivel de satisfacción con la actuación de dicho sindicato. Para ello se parte de los datos de 7 individuos tomados aleatoriamente de personas adscritas a partidos políticos, obteniéndose:

Años	**8**	**7**	**10**	**3**	**6**	**13**	**4**
Satisfacción	**7**	**5**	**8**	**5**	**9**	**9**	**3**

a) Calcular medias, varianzas y desviaciones típicas.
b) Calcular covarianza.
c) Calcular el coeficiente de correlación lineal. Comentar el resultado obtenido.
d) Hallar la ecuación de la recta de regresión de la satisfacción en función de los años de afiliación.
e) Predecir el índice de satisfacción de una persona que lleva 11 años militando al sindicato.

FÓRMULAS ÚTILES

- Notación: Y es la variable dependiente, X es la variable independiente.
- Covarianza muestral:
  a. Datos No Agrupados:
  $$s_{XY} = \frac{1}{n}\left[\sum_{i=1}^{n} x_i y_i\right] - \bar{x}\bar{y}$$
  b. Datos Agrupados por frecuencias:
  $$s_{XY} = \frac{1}{n}\left[\sum_{i=1}^{k} n_i x_i y_i\right] - \bar{x}\bar{y}$$
- Coeficiente de correlación de Pearson:
  $$r_{XY} = \frac{s_{XY}}{s_X\, s_Y}$$
- Interpretación de r_{XY}:
  a. Siempre está entre -1 y 1.
  b. Si toma valores cercanos a -1, la correlación es fuerte e inversa.
  c. Si toma valores cercanos a 1, la correlación es fuerte y directa.
  d. Si toma valores cercanos a 0, la correlación es débil.
  e. Siempre tiene el mismo signo que la covarianza.
- Recta de regresión de Y en función de X:
  $$Y = \beta_0 + \beta_1 X$$
  $$\beta_1 = r_{XY}\frac{s_Y}{s_X}$$
  $$\beta_0 = \bar{y} - \beta_1 \bar{x}$$
- Coeficiente de determinación en regresión lineal simple: $R^2 = r_{XY}^2$
- Interpretación de R^2: el modelo explica ese % de la variabilidad de la variable Y en función de X.