Estadística II. Laboratorio 7

Mayo 2025

1. Análisis de Regresión

1. Sean X e Y variables aleatorias con una distribución conjunta multinomial dada por:

$$f_{XY}(x,y) = \binom{n}{x, y, n-x-y} \theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y} \tag{1}$$

Donde $x = 0, 1, 2, \ldots, n$, $y = 0, 1, 2, \ldots, n$, y $x + y \leq n$. Determine la expresión para la regresión de Y en X, i.e., determine $\mu_{Y|X}$.

2. Sean dos variables aleatorias X e Y con densidad conjunta de probabilidad dada por:

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otro caso} \end{cases} \tag{2}$$

Determine la ecuación de regresión de X en Y, i.e., $\mu_{X|Y}$.

3. Sea la siguiente función de densidad conjunta para las variables aleatorias X e Y:

$$f(x,y) = \begin{cases} \frac{2x}{(1+x+xy)^3} & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otro caso} \end{cases} \tag{3}$$

Determine la ecuación de regresión de Y en X y la varianza de Y condicional en X, i.e., $\mu_{Y|X}$ y $Var(Y|X)$.

ID de Persona	Peso	BP
1	165	130
2	167	133
3	180	150
4	155	128
5	212	151
6	175	146
7	190	150
8	210	140
9	200	148
11	149	125
11	158	133
12	169	135
13	170	150
14	172	153
15	159	128

Cuadro 1: Muestra de Peso y Presión Sanguínea para hombres de 25 a 30 años

4. En el Cuadro 1 se muestran el peso y la presión sistólica sanguínea de 26 hombres seleccionados al azar, en el grupo de edades de 25 a 30 años. Supongamos que el peso y la presión sanguínea (BP) tienen una distribución normal conjunta. Al respecto:

   a) Determine los coeficientes de regresión lineal considerando un término constante, i.e., suponiendo una ecuación del tipo: $\mu_{y|x} = \beta_1 + \beta_2 x$, donde y será la presión sanguínea y x será el peso

   b) Determine el valor de la varianza estimada de los errores: $\hat{\sigma}^2$

   c) Pruebe la hipótesis nula de que $\beta_2 > 0$ con un $\alpha = 0.05$

   d) Pruebe la hipótesis nula de que $\beta_2 = 0.6$ con un $\alpha = 0.05$

   e) Escriba una conclusión de las dos pruebas de hipótesis planteadas

5. Un embotellador de bebidas gaseosas analiza las rutas de servicio de las máquinas expendedoras en su sistema de distribución. Le interesa

predecir el tiempo necesario para que el representante de ruta atienda las máquinas expendedoras en una tienda. Esta actividad de servicio consiste en abastecer la máquina con productos embotellados, y algo de mantenimiento o limpieza. El ingeniero industrial responsable del estudio ha sugerido que las dos variables más importantes que afectan el tiempo de entrega, y, son la cantidad de cajas de producto abastecido, x_1, y la distancia caminada por el representante, x_2. El ingeniero ha reunido una muestra aleatoria de algunas observaciones de tiempo de entrega.

Suponga que desea estimar la ecuación de regresión asumiendo una media condicional dada por:

$$\mu_{y_i|x_{i1},x_{i2},x_{i3}} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \tag{4}$$

Derivado de la operación matricial de la muestra suponga que obtiene los siguientes resultados:

$$\begin{bmatrix} \sum y_i^2 & \sum y_i x_{i1} & \sum y_i x_{i2} & \sum y_i x_{i3} \\ \sum x_{i1} y_i & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \sum x_{i1} x_{i3} \\ \sum x_{i2} y_i & \sum x_{i2} x_{i1} & \sum x_{i2}^2 & \sum x_{i2} x_{i3} \\ \sum x_{i3} y_i & \sum x_{i3} x_{i1} & \sum x_{i3} x_{i2} & \sum x_{i3}^2 \end{bmatrix}$$
$$= \begin{bmatrix} 18310.63 & 559.60 & 7375.44 & 324972.70 \\ 559.60 & 25.00 & 219.00 & 9932.00 \\ 7375.44 & 219.00 & 3055.00 & 129099.00 \\ 324972.69 & 9932.00 & 129099.00 & 6402888.00 \end{bmatrix} \tag{5}$$

Al respecto:

a) Determine los coeficientes $\boldsymbol{\beta}$ de regresión lineal

b) Determine el valor de la varianza estimada de los errores: $\hat{\sigma}^2$

c) Pruebe la hipótesis nula de que $\beta_2 = 0$ con un $\alpha = 0.05$

d) Pruebe la hipótesis nula de que $\beta_3 = 0$ con un $\alpha = 0.05$

e) Escriba una conclusión de las dos pruebas de hipótesis planteadas