



Multiple Lineal Regression. Exercises

Machine Learning I

Ejercicio MLR.1

Se pretenden estimar los gastos en alimentación de una familia en base a la información que proporcionan las variables regresoras 'ingresos mensuales y 'número de miembros de la familia'. Para ello se recoge una muestra aleatoria simple de 15 familias, cuyos resultados se facilitan en la tabla adjunta. (El gasto e ingreso se expresan en cien mil euros).

| Gasto Alimentación | Ingresos | Tamaño |
|--------------------|----------|--------|
| 0,43 | 2,1 | 3 |
| 0,31 | 1,1 | 4 |
| 0,32 | 0,9 | 5 |
| 0,46 | 1,6 | 4 |
| 1,25 | 6,2 | 4 |
| 0,44 | 2,3 | 3 |
| 0,52 | 1,8 | 6 |
| 0,29 | 1 | 5 |
| 1,29 | 8,9 | 3 |
| 0,35 | 2,4 | 2 |
| 0,35 | 1,2 | 4 |
| 0,78 | 4,7 | 3 |
| 0,43 | 3,5 | 2 |
| 0,47 | 2,9 | 3 |
| 0,38 | 1,4 | 4 |

Realizar un análisis de regresión lineal múltiple y comentar los resultados:

Ejercicio MLR.2

El gerente de una empresa estudia las posibles relaciones entre beneficios anuales, gastos en publicidad anuales y horas extraordinarias anuales de los empleados. Para ello utiliza los datos de la tabla que sigue:

| Beneficios (millones) | Gastos Publicidad) (millones) | Horas extras (100 horas) |
|--------------------------|----------------------------------|-----------------------------|
| 1,3 | 0,3 | 4 |
| 3,5 | 1,5 | 9 |
| 2,8 | 0,7 | 6 |
| 3 | 1,1 | 7,5 |
| 3,3 | 1,2 | 8 |
| 4 | 2 | 7 |
| 3,7 | 2 | 8 |

Se pide:

- Obtener un modelo de regresión para conseguir el objetivo que busca el gerente de esta empresa. Razonar la respuesta con todos los cálculos necesarios.
- Calcular el coeficiente de determinación y el mismo corregido del modelo. Comentar el resultado obtenido.
- Los intervalos de confianza de los coeficientes obtenidos
- Demostrar si el modelo es adecuado o no.
- ¿Se puede prescindir de la variable “horas extras”?

Ejercicio MLR.3

En un estudio de los determinantes de la inversión se usaron 20 datos anuales correspondientes a las siguientes variables: inversión anual en millones de euros (Y), tipo de interés en porcentaje (X_1) y variación anual de PIB en millones de euros (X_2). Se dispone de la siguiente información:

$$\begin{array}{lll} \sum X_{1t} = 100 & \sum X_{2t} = 24 & \sum Y_t = 5 \\ \sum X_{1t}Y_t = -255 & \sum X_{2t}Y_t = 146 & \sum X_{1t}X_{2t} = 100 \\ \sum X_{1t}^2 = 680 & \sum X_{2t}^2 = 48'8 & \sum (Y_t - \bar{Y})^2 = 1200 \end{array}$$

Se pide:

- Obtener un modelo de regresión para conseguir el objetivo que busca el gerente de esta empresa. Razonar la respuesta con todos los cálculos necesarios.
- Calcular el coeficiente de determinación y el mismo corregido del modelo. Comentar el resultado obtenido.
- Los intervalos de confianza de los coeficientes obtenidos
- Demostrar si el modelo es adecuado o no.
- ¿Se puede prescindir de la variable “PIB”?

Ejercicio MLR.4

Para estimar el modelo $Y_t = \beta_0 + \beta_1 \times X_{1t} + \beta_2 \times X_{2t}$ se conoce que

$$X^tX = \begin{pmatrix} 14 & 7 & 14 \\ 7 & 4'5 & 7 \\ 14 & 7 & 15 \end{pmatrix}, \quad X^tY = \begin{pmatrix} 10 \\ 6 \\ 12 \end{pmatrix}, \quad Y^tY = 14.$$

Se pide:

- Estimar los coeficientes del modelo de regresión por MCO
- Demostrar si el modelo es adecuado o no.
- Estudiar la significación del modelo
- Los intervalos de confianza de los coeficientes obtenidos
- ¿Se puede prescindir de la variable “ X_2 ”?

Ejercicio MLR.5

Una determinada cooperativa agrícola dedicada a la producción y comercialización de fresas se plantea hacer un estudio para explicar el volumen de sus ventas (X_1), expresadas en miles de euros. Para ello, a partir de los datos semestrales obtenidos desde la creación de la cooperativa, se plantea un modelo lineal usando como variables explicativas el gasto en publicidad (X_2), expresado en miles de euros, y el número de supermercados que comercializan sus productos (X_3). Seguidamente se presenta la información de la que dispone la empresa:

| Semestre | X2 | X3 | X1 |
|------------|-----|----|----|
| 2º de 2004 | 1,5 | 15 | 13 |
| 1º de 2005 | 1,7 | 17 | 41 |
| 2º de 2005 | 2 | 19 | 16 |
| 1º de 2006 | 2,3 | 20 | 47 |
| 2º de 2006 | 2,5 | 23 | 19 |
| 1º de 2007 | 3 | 25 | 55 |
| 2º de 2007 | 3,5 | 26 | 22 |
| 1º de 2008 | 4 | 29 | 63 |
| 2º de 2008 | 4,3 | 31 | 25 |
| 1º de 2009 | 4,5 | 34 | 65 |

A partir de esos datos:

- Realizar un modelo de regresión lineal múltiple para el problema. Interpretar sus coeficientes, indicando a su vez cuál es la capacidad explicativa del mismo.
- A partir de los resultados obtenidos en el apartado anterior, ¿se podría sugerir que pueden existir otras variables no tenidas en cuenta aquí y de importancia relevante para la explicación de las ventas?

Observando los datos, nos damos cuenta que las ventas correspondientes al primer semestre son siempre mucho mayores que las del segundo semestre del mismo año. Para recoger este aspecto, se decidió introducir una nueva variable explicativa X_4 , cuyo valor sería uno si el dato corresponde al primer semestre, y cero en caso contrario. Utilizando esta variable adicional:

- Determinar el nuevo modelo de regresión planteado e interpretar sus coeficientes, indicando la capacidad explicativa del mismo.
- Bajo este modelo, ¿en cuánto se incrementarían las ventas en el segundo semestre de 2009 (con respecto al segundo semestre de 2008) si los gastos en publicidad se incrementasen en 500000 pesetas y se espera que dos nuevos supermercados comercialicen los productos de la cooperativa? ¿Y si los incrementos fuesen del 2% en ambas variables explicativas?