Regression Analysis

May 15, 2023

1 Introduction

2 Simple Linear Regression

Consider the hypothetical question we are analyzing if there is a relation between hours studied and marks scored in an exam. Here hours studied is independent variable and marks scored is dependent of hours studied. Regression analysis will help us quantify this effect. The fundamental assumption that we make in linear regression is *there is a linear relation between* dependent and independent variables.

$$\hat{y} = \beta_0 + \beta_1 x_i \tag{1}$$

We will define a quanity Sum of squared residuals (SSR) as $(y_i - \hat{y}_i)^2$

$$\begin{aligned} SSR &= (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \\ &= \sum_{i} \left[y_i^2 + (\beta_0 + \beta_1 x_i)^2 - 2y_i(\beta_0 + \beta_1 x_i)\right] \\ &= \sum_{i} \left[y_i^2 + \beta_0^2 + \beta_1^2 x_i^2 + 2\beta_0\beta_1 x_i - 2\beta_0 y_i - 2\beta_1 x_i y_i\right] \end{aligned} \tag{2}$$

Find β_0 and β_1 such that $y_{actual} - y_{predicted}$ is small. To do this we need to minimise the SSR.

From Equation 2 we get,

$$SSR = \sum_i \left[y_i^2 + \beta_0^2 + \beta_1^2 x_i^2 + 2\beta_0\beta_1 x_i - 2\beta_0 y_i - 2\beta_1 x_i y_i\right]$$
$$\frac{\partial SSR}{\partial \beta_0} = \sum_i \left[2\beta_0 + 2\beta_1 x_i - 2y_i\right]$$

We will set $\frac{\partial SSR}{\partial \beta_0}$ to zero that is

$$\frac{\partial SSR}{\partial \beta_0} = \sum_i \left[2\beta_0 + 2\beta_1 x_i - 2y_i\right] := 0$$

Now,

$$2\sum_{i=1}^{n} \left[\beta_0 + \beta_1 x_i - y_i\right] = 0$$

$$n\beta_0 = \sum_i^n y_i - \sum_i^n \beta_1 x_i$$

$$\beta_0 = \frac{\sum_i^n y_i}{n} - \frac{\sum_i^n \beta_1 x_i}{n}$$

That is $\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}$

Now to find β_1 we need to minimise SSR i.e. equation 2 wrt β_1

From Equation 2 we get,

$$SSR = \sum_i \left[y_i^2 + \beta_0^2 + \beta_1^2 x_i^2 + 2\beta_0\beta_1 x_i - 2\beta_0 y_i - 2\beta_1 x_i y_i\right]$$
$$\frac{\partial SSR}{\partial \beta_1} = \sum_i \left[2\beta_1 x_i^2 + 2\beta_0 x_i - 2x_i y_i\right]$$

setting this to zero, we get

$$\sum_i \left[2\beta_1 x_i^2 + 2\beta_0 x_i - 2x_i y_i\right] = 0$$

$$-2\sum_i x_i\left[y_i - (\beta_0 + \beta_1 x_i)\right] = 0$$

$$\sum_i x_i\left[y_i - (\beta_0 + \beta_1 x_i)\right] = 0 \tag{3}$$

Now we can substitute the already derived value of β_0 in equation 3.

$$\sum_i x_i\left[y_i - ({\color{red}\beta_0} + \beta_1 x_i)\right] = 0$$

$$\sum_i x_i\left[y_i - ((\bar{y} - \beta_1\bar{x}) + \beta_1 x_i)\right] = 0$$

$$\sum_i x_i\left[y_i - \bar{y} + \beta_1\bar{x} - \beta_1 x_i\right] = 0$$

$$\sum_i x_i(y_i - \bar{y}) - \sum_i \beta_1 x_i(x_i - \bar{x}) = 0$$

$$\sum_i x_i(y_i - \bar{y}) = \sum_i \beta_1 x_i(x_i - \bar{x})$$

$$\sum_i x_i(y_i - \bar{y}) = \beta_1 \sum_i x_i(x_i - \bar{x})$$

$$\boxed{\beta_1 = \frac{\sum_i x_i(y_i - \bar{y})}{\sum_i x_i(x_i - \bar{x})}} \tag{4}$$

There is another form of β_1 which is more frequently presented that is equivalent to equation 4. To find that out lets consider,

$$
\begin{aligned}
\sum_i (x_i - \bar{x})(y_i - \bar{y}) &= (x_i(y_i - \bar{y}) - \bar{x}(y_i - \bar{y})) \\
&= \sum_i x_i(y_i - \bar{y}) - \bar{x} \underbrace{\sum_i (y_i - \bar{y})}_{\text{this is always zero}} \\
&= \sum_i x_i(y_i - \bar{y}) \\
&= \sum_i y_i(x_i - \bar{x}) && \text{(numerator in 4)} \\
\sum_i (x_i - \bar{x})^2 &= \sum_i (x_i - \bar{x})(x_i - \bar{x}) \\
&= \sum_i x_i(x_i - \bar{x}) - \bar{x}(x_i - \bar{x}) \\
&= \sum_i x_i(x_i - \bar{x}) - \bar{x} - \underbrace{\sum_i (x_i - \bar{x})}_{\text{this is zero}} \\
&= \sum_i x_i(x_i - \bar{x}) && \text{(denominator in 4)}
\end{aligned}
$$

Therefore,

$$
\boxed{\beta_1 = \frac{\sum_i x_i(y_i - \bar{y})}{\sum_i x_i(x_i - \bar{x})} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}} \tag{4}
$$

$$
\begin{aligned}
\beta_0 &= \bar{y} - \beta_1 \bar{x} \\
\beta_1 &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}
\end{aligned}
$$

3 Multiple Linear Regression

Let us consider the regression equation for m predictors is

$$
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_m x_m + \epsilon
$$

where we have $(n+1)$ examples in our dataset.

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2m} \\ 1 & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The above equation in matrix form is $Y = X\beta + \epsilon$,

Note that Y is matrix of shape $(n+1) \times 1$ and X is of shape $(n+1) \times (m+1)$ and it is called **design matrix**, β is **coefficient matrix** of form $(m+1) \times 1$ and ϵ is of shape $(n+1) \times 1$ and called **noise** or disturbance.

$Y = X\beta + \epsilon$, and $\epsilon = Y - X\beta$. To **minimize** ϵ we take [1]inner product.

$$\begin{aligned} \epsilon^T \epsilon &= (Y - X\beta)^T (Y - X\beta) \\ &= \left(Y^T - (X\beta)^T\right)(Y - X\beta) \\ &= \left(Y^T - \beta^T X^T\right)(Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \end{aligned}$$

Note that $\beta^T X^T Y$ is a scaler, and verify that shape of

$\beta^T X^T Y$ is $\underbrace{1 \times (m+1)}_{\beta^T} \times \underbrace{(m+1) \times (n+1)}_{X^T} \times \underbrace{(n+1) \times 1}_{Y} = 1$ also traspose of a scalar is also scaler, that is $(Y^T X\beta)^T = \beta^T X^T Y$.

$$\epsilon^T \epsilon = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

To minimise[2],

[1]Inner product tells you how much of one vector is pointing in the direction of another one. Inner product, of two vectors, is the sum of the products of corresponding components

Matrix Calculus Review

[2] $\frac{\partial a^T b}{\partial b} = \frac{\partial b^T a}{\partial b} = a$

$\frac{\partial b^T A b}{\partial b} = 2Ab = 2b^T A$

$$\frac{\partial \epsilon^T \epsilon}{\partial \beta} = -2X^T Y + 2X^T X\beta$$

$$0 = -2X^T Y + 2X^T X\beta$$

$$X^T X\beta = X^T Y$$

The above form is called normal equations.

If $X^T X$ is **non-singular** (that is, inverse exists) then

$$\left(X^T X\right)^{-1} X^T X\beta = \left(X^T X\right)^{-1} X^T Y$$

$$\boxed{\beta = \left(X^T X\right)^{-1} X^T Y}$$

Implementing in Python `X` is the data matrix and `Y` is response. Using native `numpy` methods we can find out estimates of β

```
betas = (np.linalg.inv(X.T@X))@X.T@Y
```

4 Standard Error & Hypothesis Testing