LR - CO6 - 2024-09-02 - Analyze the data to perform linear regression

02 September 2024 09:18

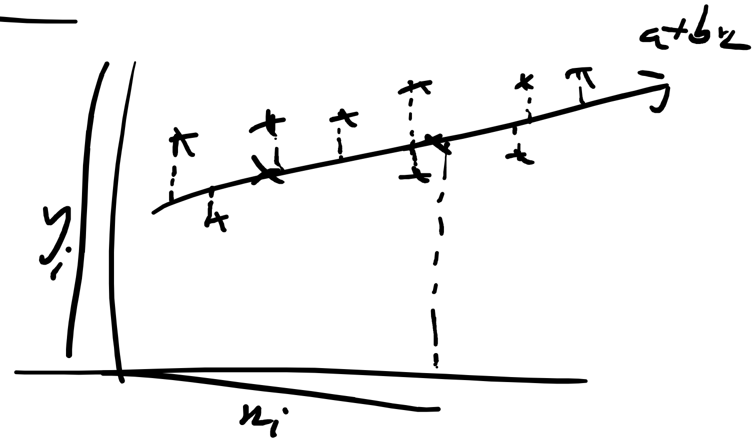
Simple Linear Regression:

$$y = a + bx$$

where a is intercept & b is slope of line

x is independent & y is dependent.

-> Least Squares Estimation:



Approximation is done by

$$y' = a + bx.$$

Exact representation

$$y = a + bx + \text{error}$$

-> $$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{SSxy}{SSX}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad SSX$$

→ Correlation coefficient b/w x & y

$$r = \hat{b} \cdot \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \hat{b}\sqrt{\frac{SSX}{SSY}}$$

TV x	overweight y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
42	18	10.533	12.26	110.9511	129.2089
34	6	2.533	0.26	6.4178	0.6756
25	0	-6.4667	-5.733	41.8178	37.0756
35	-1	3.533	-6.733	12.4844	-23.7911
37	13	5.533	7.2667	30.6178	40.2089
38	14	6.533	8.2667	42.6844	54.0089
31	7	-0.4667	1.2667	0.2178	-0.5911
33	7	1.5333	1.2667	2.3511	1.9422
19	-9	-12.4667	-14.733	155.4178	163.6756
29	8	-2.4667	2.2667	6.0844	-5.5911
38	8	6.533	2.2667	42.6844	14.8089
28	5	-3.4667	-0.7333	12.0178	2.5422
29	3	-2.4667	-2.7333	6.0844	6.7422
36	14	4.5333	8.2667	20.5511	37.4756
18	-7	-13.4667	-12.7333	181.3511	171.4756

$-7 \quad -13.4667 \quad -12.1333 \quad 181.3511 \quad 111.7466$

$\bar{x} = \underline{31.4667} \qquad \bar{y} = \underline{5.733} \qquad SSx = 671.733 \qquad SSxy = 649.8667$

$$\hat{b} = \frac{SSxy}{SSx}$$

$$= \frac{649.8667}{671.733} = \underline{\underline{0.9674}}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 5.733 - (0.9674 \times 31.4667)$$

$$= \underline{\underline{-24.709}}$$

$$\hat{y} = \underline{\underline{a + bx}}$$

$x = 30$

$$\hat{y} = a + bx$$

$$= -24.709 + (0.9674 \times 30)$$

$$= \underline{\underline{4.315}}$$

→ Metric – Accuracy ~ Goodness of Fit

(R^2 – Score)

$\rightarrow \sum_{i=1}^{n} (y_i - \hat{y})^2 \rightarrow$ Sum of Squares of Error (SSE)

$\rightarrow \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \rightarrow$ Sum of Squares due to Regression (SSR)

$\rightarrow SST = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$

$SST = SSE + SSR$

$\rightarrow R^2 = \frac{SSR}{SST} = 1 - \left(\frac{SSE}{SST} \right)$

— $SST = 818.9333$

— $SSE = 190.2216$

$R^2 = 1 - \left(SSE / SST \right) = 1 - \left(\frac{190.2216}{818.933} \right)$

$= \underline{\underline{0.7677}}$