Practical.10
Fitting of simple linear regression of y on x

Testing the significance of regression co-efficient

To test the significance of the regression coefficient we can apply either a t test or analysis of variance (F test). The ANOVA table for testing the regression coefficient will be as follows:

Sources of variation	**DF**	**SS**	**MS**	**F ratio**
Due to regression	1	SS(b)	S_b^2	S_b^2 / S_e^2
Deviation from regression	n-2	SS(Y)-SS(b)	S_e^2	
Total	n-1	SS(Y)		

In case of t test the test statistic is given by

$$t = b / SE(b) \text{ where } SE(b) = s_e^2 / SS(X)$$

Example 1

Form a paddy field, 36 plants were selected at random. The length of panicles(x) and the number of grains per panicle (y) of the selected plants were recorded. The results are given below. Fit a regression line y on x. Also test the significance (or) regression coefficient.

The length of panicles in cm (x) and the number of grains per panicle (y) of paddy plants.

S.No.	Y	X	S.No.	Y	X	S.No.	Y	X
1	95	22.4	13	143	24.5	25	112	22.9
2	109	23.3	14	127	23.6	26	131	23.9
3	133	24.1	15	92	21.1	27	147	24.8
4	132	24.3	16	88	21.4	28	90	21.2
5	136	23.5	17	99	23.4	29	110	22.2
6	116	22.3	18	129	23.4	30	106	22.7
7	126	23.9	19	91	21.6	31	127	23.0
8	124	24.0	20	103	21.4	32	145	24.0
9	137	24.9	21	114	23.3	33	85	20.6
10	90	20.0	22	124	24.4	34	94	21.0

11	107	19.8	23	143	24.4	35	142	24.0
12	108	22.0	24	108	22.5	36	111	23.1

Null Hypothesis H_o: regression coefficient is not significant.

Alternative Hypothesis H_1: regression coefficient is significant.

$$\sum y = 4174 \qquad \sum y^2 = 496258 \qquad \bar{y} = \frac{\sum y}{n} = 115.94$$

$$\sum x = 822.9 \qquad \sum x^2 = 18876.83 \qquad \bar{x} = \frac{\sum x}{n} = 22.86$$

$$\sum xy = 96183.4$$

$$\text{SS(Y)} = \sum y^2 - \frac{(\sum y)^2}{n} = 496258 - \frac{(4174)^2}{36} = 12305.8889$$

$$\text{SS(X)} = \sum x^2 - \frac{(\sum x)^2}{n} = 18876.83 - \frac{(822.9)^2}{36} = 66.7075$$

The regression line y on x is $\bar{y} = a_1 + b_1 \bar{x}$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{96183.4 - \frac{(822.9)(4174)}{36}}{66.7075} = 11.5837$$

$\bar{y} = a_1 + b_1 \bar{x}$

115.94 = a + (11.5837)(22.86)

A=115.94-264.8034

A=-148.8633

The fitted regression line is y =-148.8633+11.5837x

$$SS(b) = \frac{\left(\sum xy - \frac{\sum x \sum y}{n}\right)^2}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{(722.7167)^2}{66.7075} = 8950.8841$$

ANOVA Table:

Sources of Variation	d.f	SS	MSS	F-value
Replication	1	8950.8841	8950.8841	90.7093
Error	36-2=34	3355.0048	98.6766	
Total	35	12305.8889		

For t-test

$$t = \frac{b}{SE(b)} \sim t_{(n-2)} d.f$$

$$SE(b = \sqrt{\frac{Se^2}{SS(X)}} = \sqrt{\frac{98.6776}{66.7075}} = 1.2162$$

$$t = \frac{11.5837}{1.2162} = 9.5245$$

Table Value:

$t_{(n-2)}$ d.f=t_{34} d.f at 5% level=2.032

$t_{cal} > t_{tab}$. we reject H_o.

Hence t is significant.

Learning Exercise

1. The following data are using length of 13 sparrows of various ages.

Age (days) (x)	3	4	5	6	8	9	10	11	12	14	15	16	17
Wing length (cm) Y	1.4	1.5	2.2	2.4	3.1	3.2	3.2	3.9	4.1	4.7	4.5	5.2	5.0

Fit the regression line of y on x. Also test the significance of the regression coefficient.

2. Obtain the regression line of the form y = a + bx between the average number of tilters (x) and the yield in kgs (y) of turmeric crop from the following data

Average number of tillers(x)	3.5	3.2	3.5	3.8	3.6	3.74	2.8	4.2	4.0	4.5
Yield in (Kgs) (y)	2.0	1.8	1.9	2.1	2.0	2.3	1.7	2.5	2.6	3.0

3. Find out the regression equations y on x between the number of root fibers(x) and yields in kgs(y) of ginger crop from the following data. Also test the significance of regression coefficient.

Number of root fibers (x)	6	5	3	7	4	9	10	11	15	8
Yield in kgs(y)	1.0	0.8	0.5	1.1	0.6	1.2	1.5	1.6	1.9	0.9

4. Find out the regression equation of y on x and test the significance of regression coefficient.

X	2.1	2.7	3.5	4.6	4.9	5.7	6.0	7.4	8.3	4.8
Y	11.2	12.8	16.4	17.8	19.5	21.1	22.8	26.3	31.0	21.1