Exercise.9

Calculation of Karl Pearson's correlation coefficient

Pearsons Correlation coefficient

. The correlation coefficient r is known as Pearson's correlation coefficient as it was discovered by Karl Pearson.

$$r = \frac{\frac{1}{n-1}\left(\sum (x-\bar{x})(y-\bar{y})\right)}{\sqrt{\frac{1}{n-1}\sum (x-\bar{x})^2}\sqrt{\frac{1}{n-1}\sum (y-\bar{y})^2}}$$

Which can be simplified as

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

Testing the significance of r

The significance of r can be tested by Student's t test. The test statistics is given by

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}}$$

Example.1

Compute Pearsons coefficient of correlation between advertisement cost and sales as per the data given below:

Advertisement Cost in 1000's	39	65	62	90	82	75	25	98	36	78
Sales in lakhs	47	53	58	86	62	68	60	91	51	84

Solution

H_o: The correlation coefficient r is not significant

H_1: The correlation coefficient r is significant.

Level of significance 5%

From the data

n = 10

$$\sum x = 650 \quad \sum y = 660 \quad \sum xy = 45604 \quad \sum x^2 = 47648 \quad \sum y^2 = 45784$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

$$= \frac{45604 - \frac{(650)(660)}{10}}{\sqrt{47648 - \frac{(650)^2}{10}} \sqrt{45784 - \frac{(660)^2}{10}}}$$

$$= \frac{45604 - 42900}{(73.47)(47.1)} = 0.7804$$

Correlation coefficient is positively correlated.

Test Statistic

$$t = \frac{|r|}{\sqrt{\frac{1 - r^2}{n - 2}}} \sim \text{(n-2) d.f}$$

$$t = \frac{0.7804}{\sqrt{\frac{1-(0.7804)^2}{10-2}}} = 3.530$$

$t_{tab} = t_{(10-2,\ 5\%los)} = 2.306$

Inference

$t_{cal} > t_{tab}$, we reject null hypothesis.

∴The correlation coefficient r is significant. (i.e) There is a relation between advertisement company and the sales.

Learning Exercise

1. Calculate the simple correlation coefficient between wing length & tail length of the following 12 birds of a particular species. Also test its significant.

Wing length (cm)x	1	2	3	4	5	6	7	8	9	10	11	12
	10.4	10.8	11.1	10.2	10.3	10.2	10.7	10.5	10.8	11.2	10.6	11.4
Tail length (cm)y	7.4	7.6	7.9	7.2	7.4	7.1	7.4	7.2	7.8	7.7	7.8	8.3

2. The date refer to the yield of grain in gms|plant(y) and the number of productive tillers (x) and 15 paddy plants

Y	37	20	42	36	20	30	26	21	43	44	22	31	26	37	26
X	15	12	17	14	12	13	12	9	24	20	14	18	13	15	7

Find the correlation

3. The following data relates to the yield in grams(y) and the matured pods (x) of 10 groundnut plants. Work out the correlation coefficient and test its significance.

X:	14	34	20	16	11	11	20	17	22	17
Y:	16	40	21	18	14	13	20	35	17	27

4. Find the persons coefficient of correlation between price and demand from the following data.

Price	11	13	15	17	18	19	20
Demand	30	29	24	24	21	18	15