Probability and Statistics

Unit VI: Linear Statistical Models

Sachin Verma[Visiting faculty]

Mukesh Patel School of Technology Management & Engineering,Mumbai, India

February 24, 2025

Contents

SVKM'S NMIMS Deemed-to-be UNIVERSITY

Correlation

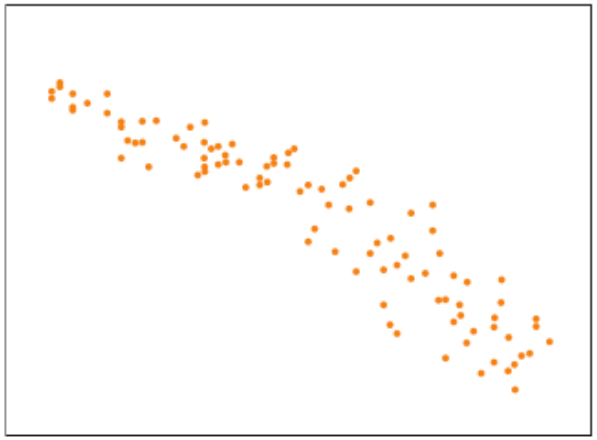
Correlation

- Correlation is a statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables.
- Two variables are said to be correlated if change in one variable affects the change in other variable, and the relation between them is known as correlation.

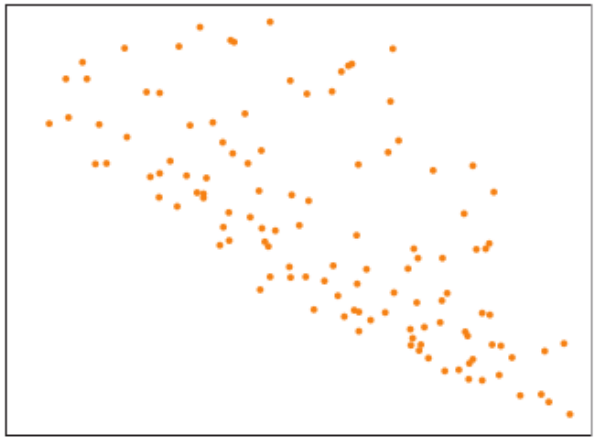
Correlation

Five Correlations

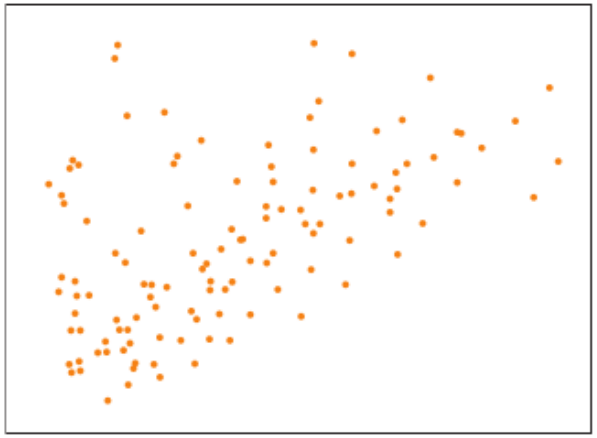
(a) Strong Negative Correlation ($r = -.933$)



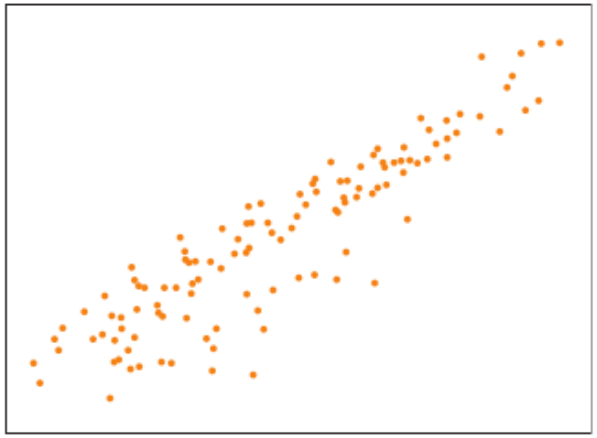
(b) Moderate Negative Correlation ($r = -.674$)



(c) Moderate Positive Correlation ($r = .518$)



(d) Strong Positive Correlation ($r = .909$)



Karl Pearson

Karl Pearson's Product Moment coefficient of correlation: Correlation coefficient between two random variables X and Y, usually denoted by $r(X, Y)$ or r_{xy} and defined as

$$r(X, Y) = \frac{\frac{\sum x_i y_i}{n} - \overline{x}.\overline{y}}{\sqrt{\frac{\sum x_i^2}{n} - \overline{x}^2}.\sqrt{\frac{\sum y_i^2}{n} - \overline{y}^2}}$$

- If r = + 1 then correlation is perfectly positive,
- If r = - 1 then correlation is perfectly negative,
- If r = 0 then variables are uncorrelated.

$$r(X,Y) = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}.\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

Question

Determine the value of the coefficient of correlation r, for the following data:

X	4	6	7	11	14	17	21
Y	18	12	13	8	7	7	4

Karl Pearson Coefficient of correlation is given by:

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}.\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$n = 7, \sum X = 80, \sum Y = 69, \sum XY = 624,$
$\sum X^2 = 1148, \sum Y^2 = 815,$
$r = -0.92698$

Question

Calculate the correlation coefficient for the following heights (in inches) of fathers (X) and their sons(Y)

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

$$r(X,Y) = \frac{\frac{\sum x_i y_i}{n} - \overline{x}.\overline{y}}{\sqrt{\frac{\sum x_i^2}{n} - \overline{x}^2}.\sqrt{\frac{\sum y_i^2}{n} - \overline{y}^2}}$$

Question

A computer while calculating correction coefficient between two variables X and Y from 25 pairs of observations obtained the following results:

$n = 25, \sum X = 125, \sum X^2 = 650,$
$\sum Y = 100, \sum Y^2 = 460, \sum XY = 508$

$$r(X,Y) = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}.\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

Spearman's Rank Correlation:

The method develop by Spearman is simpler than Karl Pearson's method since, it depends upon ranks of the items and actual values of the items are not required.
Hence this can be used to study correlation even when actual values are not known. For instance, we can study correlation between intelligence and honesty by this method.

$$R = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

$d_i = R_1 - R_2$
R_1 : rank of X
R_2 : rank of Y

Question

Calculate the rank correlation coefficient from the following data.

X	1	3	7	5	4	6	2	10	9	8
Y	3	1	4	5	6	9	7	8	10	2

X	1	3	7	5	4	6	2	10	9	8
Y	3	1	4	5	6	9	7	8	10	2
d_i										
d_i^2										

We know,

$$R = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

SVKM'S NMIMS Deemed-to-be UNIVERSITY

Question

Six students got the following Mathematics and Physics:

Maths	*78*	*36*	*98*	*25*	*75*	*82*
Physics	*84*	*51*	*91*	*60*	*68*	*62*

Calculate the rank correlation coefficient.

SVKM'S NMIMS Deemed-to-be UNIVERSITY

Maths	R_1	Physics	R_2	d_i	d_i^2

SVKM'S NMIMS Deemed-to-be UNIVERSITY

Ranks are repeated

If ranks are repeated then the Spearman' s Rank corrrelation formula becomes

$$R = 1 - \frac{6[\sum_i^n d_i^2 + \sum_{i=1}^n \frac{m_i^3 - m_i}{12}]}{n^3 - n}$$

Where **m** is the number of times an item is repeated

SVKM'S NMIMS Deemed-to-be UNIVERSITY

Question

The following table shows the marks obtained by 10 students in Accountancy and Statistics. Find the Spearman's coefficient of rank correlation.

No.	1	2	3	4	5	6	7	8	9	10
Acc	45	70	65	30	90	40	50	57	85	60
Stat	35	90	70	40	95	40	60	80	80	50

$$R = 1 - \frac{6[\sum d_i^2 + \sum_{i=1}^{n} \frac{m_i^3 - m_i}{12}]}{n^3 - n}$$

Acc	R_1	Stats	R_2	d_i	d_i^2
45		35			
70		90			
65		70			
30		40			
90		95			
40		40			
50		60			
57		80			
85		80			
60		50			