Moments

In statistics, moments are used to describe the characteristic of a distribution.

The arithmetic mean of the r^{th} powers of the deviations from any point is called as the r^{th} moment about 'a' and is denoted by μ'. (Raw moments)

$$\mu_r' = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - a)^r \quad , \quad N = \sum_{i=1}^{n} f_i \rightarrow (1)$$

If the deviations are taken from the mean ($\bar{x}$), then the moments are called as Central moments (↳ moments about the mean) and are defined as

$$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^r \rightarrow (2)$$

If the deviations are taken from the origin (i.e. 0) the moments are called as Raw moments (↳ moments about any point or origin) and are defined as

$$\mu_r' = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^r$$

There are 4 commonly used moments in statistics are:
Mean, variance, skewness and kurtosis.

First moment : Mean or average (measures the location of central point)

Second central moment: variance (measures spread of values in distribution or how set of data points are spread out from the mean)

Third moment: skewness (measures how asymmetric distribution is about its mean)

3 types:
1) symmetric
2) +ve skewed dis.
3) -ve -11-

Fourth moment: Kurtosis (explains if the distribution is flat or with a high peak)

Mesokurtic,
Platykurtic,
leptokurtic

In (2), if $r = 0$,

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^0 = \frac{1}{N} \sum_{i=1}^{n} f_i = \frac{N}{N} = 1$$

$$\therefore \boxed{\mu_0 = 1} \rightarrow \text{zeroth moment talks about total probability}$$

$r = 1$,

$$\mu_1 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^1$$

$$= \frac{1}{N} \left[\sum_{i=1}^{n} f_i x_i - \sum_{i=1}^{n} f_i \bar{x} \right]$$

$$= \frac{1}{N} \Sigma f_i x_i - \frac{1}{N} \bar{x} \Sigma f_i$$

$$\therefore \mu_1 = \bar{x} - \bar{x} \left(\frac{N}{N} \right) = 0$$

Similarly μ_2, μ_3 ... so on.

$r = 2$, $\mu_2 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \sigma^2$ (i.e. variance)

Also $\mu_0' = 1$, $\mu_1' = \frac{1}{N}\sum f_i(x_i - a) = \frac{1}{N}\sum f_i x_i - a = \bar{x} - a$

Moments about the mean in terms of moments about any point 'a'.

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$

$x_i - \bar{x} = (x_i - a) - (\bar{x} - a)$

$x_i - \bar{x} = (x_i - a) - \mu_1'$

$\mu_n = \frac{1}{N}\sum f_i (x_i - \bar{x})^n$

$\mu_n = \frac{1}{N}\sum f_i [(x_i - a) - \mu_1']^n$

Moments about any point 'a' in terms of moments about mean

$$\mu_2' = \mu_2 + (\mu_1')^2$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + (\mu_1')^3$$

$$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2(\mu_1')^2 + (\mu_1')^4$$

$x_i - a = (x_i - \bar{x}) + (\bar{x} - a)$

$= (x_i - \bar{x}) + \mu_1'$

$\mu_n' = \frac{1}{N}\sum f_i (x_i - a)^n$

$\mu_n' = \frac{1}{N}\sum f_i [(x_i - \bar{x}) + \mu_1']^n$

measure of degree of asymmetry in distribution

Skewness and Kurtosis:

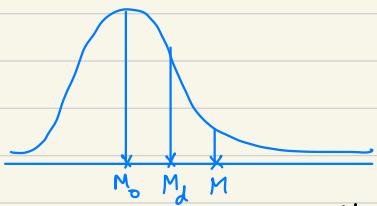
Skewness means non-symmetry or lack of symmetry. A frequency distribution is said to be symmetrical when the values equidistant from the mean have equal frequencies and for such a frequency distribution; arithmetic mean, median and mode always coincide.

A distribution is said to be skewed if

1) mean, median and mode fall at different points (unequal)
2) the sketched curve is stretched more to one side

than to the other side.

Skewness can be positive as well as negative.

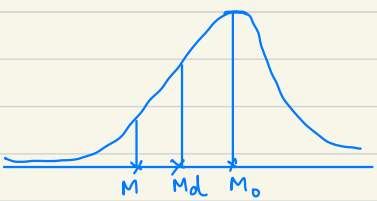


(Positive skewness) right tail longer than left tail

Mode < Median < Mean

(Symmetrical)

Mode = Median = Mean



(Negative skewness) → left tail is longer than right tail

Mean < Median < Mode

Measure of skewness with help of moments

$\mu_3 = 0 \Rightarrow$ no skewness

$\mu_3 > 0 \Rightarrow$ +ve skewness

$\mu_3 < 0 \Rightarrow$ -ve skewness

Pearson's β and γ coefficients

4 coefficients based upon first 4 central moments :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad ; \quad \gamma_1 = \pm\sqrt{\beta_1}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad ; \quad \gamma_2 = \beta_2 - 3$$

These coefficients have applications in the measurement of skewness and kurtosis.

The Karl Pearson's coefficient of skewness is defined as follows:

$$S_K = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2\,(5\beta_2 - 6\beta_1 - 9)}$$

Kurtosis : gives degree of flatness or peakedness of curve

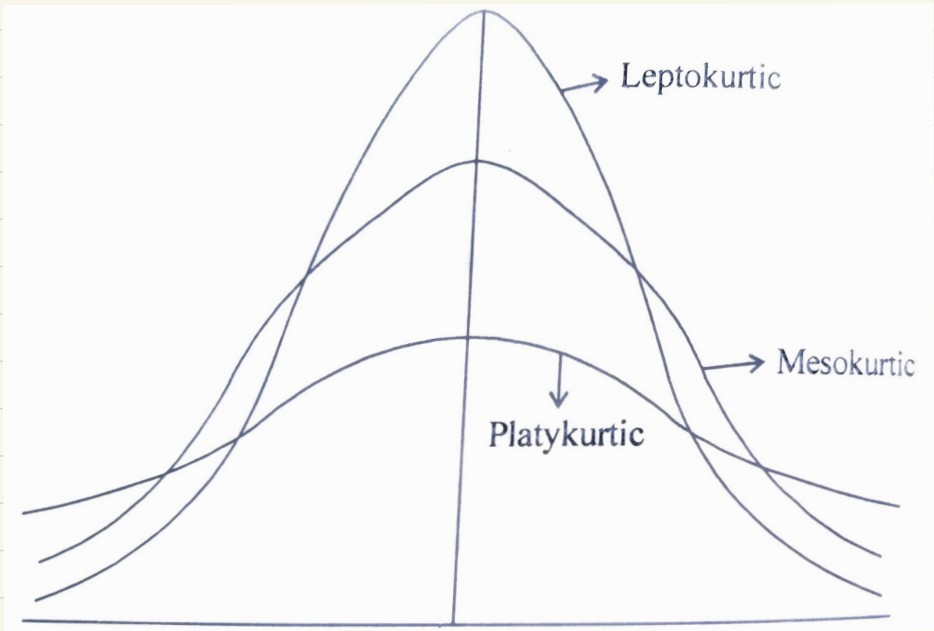
It is a characteristic related with nature of concentration of observations in the middle part of frequency distribution. Kurtosis helps us to have an idea about flatness or peakedness of curve. It is measured by Pearson's coefficients β_2 and γ_2.

3 forms :

1) Mesokurtic : A curve which is not very peaked or very flat is called normal or Mesokurtic curve. For such a curve $\beta_2 = 3$ and $\gamma_2 = 0$.

2) Platykurtic : A curve which is flatter than Mesokurtic curve. For such a curve $\beta_2 < 3$ and $\gamma_2 < 0$.

3) Leptokurtic : A curve which is more peaked than Mesokurtic curve. For such a curve $\beta_2 > 3$ and $\gamma_2 > 0$.



1) The first four moments of a distribution about the value 4 of the variable are -1.5, 17, -30 and 108. Find the moments about mean, β_1 and β_2. Find also moments about i) origin and ii) the point $x=2$.

Sol: Given $a=4$, $\mu_1' = -1.5$, $\mu_2' = 17$, $\mu_3' = -30$, $\mu_4' = 108$

Moments about mean:

$$\mu_2 = \mu_2' - (\mu_1')^2 = 17 - (-1.5)^2 = 14.75$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = -30 - 3(17)(-1.5) + 2(-1.5)^2$$
$$= 39.75$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 = 142.3125$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.4924$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2} = 0.6541$$

Also $\mu_1' = \bar{x} - a \Rightarrow \bar{x} = a + \mu_1' = 4 + (-1.5) = 2.5$

Moments about origin : We have $\bar{x} = 2.5$, $\mu_2 = 14.75$,

$\mu_3 = 39.75$, $\mu_4 = 142.3125$

$\mu_1' = \bar{x} - a = \bar{x} - 0 = \bar{x} = 2.5$ (first moment about origin)

$\mu_2' = \mu_2 + \mu_1'^2 = 14.75 + (2.5)^2 = 21$

$\mu_3' = \mu_3 + 3\mu_2\mu_1' + (\mu_1')^3 = 39.75 + 3(14.75)(2.5) + (2.5)^3$
$= 166$

$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2(\mu_1')^2 + (\mu_1')^4$

$\therefore \mu_4' = 1132$

Moments about the point $a = 2$

We have $\bar{x} = a + \mu_1'$. The first moment about the point $a = 2$ is

$$\mu_1' = \bar{x} - a = 2.5 - 2 = 0.5$$

$$\mu_2' = \mu_2 + (\mu_1')^2 = 14.75 + 0.25 = 15$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + (\mu_1')^3 = 39.75 + 3(14.75)(0.5) + (0.5)^3$$
$$= 62$$

$$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2(\mu_1')^2 + (\mu_1')^4$$

$$\therefore \mu_4' = 244$$

Q) Calculate the first 4 moments of the following distribution about the mean and hence find β_1 and β_2.

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

Sol:

x	f	fx	$d = (x-4)$	fd	fd^2	fd^3	fd^4
0	1	0	-4	-4	16	-64	256
1	8	8	-3	-24	72	-216	648
2	28	56	-2	-56	112	-224	448
3	56	168	-1	-56	56	-56	56
4	70	280	0	0	0	0	0
5	56	280	1	56	56	56	56
6	28	168	2	56	112	224	448
7	8	56	3	24	72	216	648
8	1	8	4	4	16	64	256
	$\Sigma f = 256$	1024	0	0	512	0	2816

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{1024}{256} = 4$$

Moments about the point $x = 4$ are.

$$\mu_1' = \frac{\Sigma fd}{N} = 0, \quad \mu_2' = \frac{\Sigma fd^2}{N} = \frac{512}{256} = 2, \quad \mu_3' = \frac{\Sigma fd^3}{N} = 0$$

$$\mu_4' = \frac{\Sigma fd^4}{N} = \frac{2816}{N} = 11$$

Moments about mean are:

always 0 $\leftarrow$ $\mu_1 = 0, \quad \mu_2 = \mu_2' - (\mu_1')^2 = 2, \quad \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 0$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 11$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11}{4} = 2.75.$$

3) Wages of workers are given in the following table:

10-12	12-14	14-16	16-18	18-20	20-22	22-24
1	3	7	20	12	4	3

Calculate the first 4 moments of the ~~foll~~ distribution and β_1 and β_2.

Sol:

Wages	f	x	fx	$d=(x_i-17.52)$	fd	fd^2	fd^3
10-12	1	11	11	-6.52	-6.52	42.5104	-277.168
12-14	3	13	39	-4.52	-13.56	61.2912	-277.036
14-16	7	15	105	-2.52	-17.64	44.4528	-112.021
16-18	20	17	340	-0.52	-10.4	5.408	-2.8122
18-20	12	19	228	1.48	17.76	26.2848	38.9015
20-22	4	21	84	3.48	13.92	48.4416	168.5768
22-24	3	23	69	5.48	16.44	90.0912	493.6998
	50		876		0	318.48	32.1409

$$\bar{x} = \frac{1}{N}\sum f_i x_i = \frac{876}{50} = 17.52$$

$$\mu_1 = \frac{1}{N}\sum f_i(x_i - \bar{x}) = \frac{1}{N}\sum fd = 0, \quad \mu_2 = \frac{1}{N}\sum fd^2 = \frac{318.48}{50} = 6.3696$$

$$\mu_3 = \frac{1}{N}\sum fd^3 = 0.6428$$

$$\mu_4 = \frac{1}{N}\sum fd^4 = 133.8558.$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.00159, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = 3.2992$$

3) Find the second, third and fourth central moments of the frequency distribution given below. Hence, find i) a measure of skewness (γ_1) ii) a measure of kurtosis (γ_2)

Class	f	Mid-point x	fx	$d = x - \bar{x}$	fd	fd^2	fd^3	fd^4
110 – 114.9	5	112.45	562.25	−14	−70	980	−13720	192080
115 – 119.9	15	117.45	1761.75	−9	−135	1215	−10935	98415
120 – 124.9	20	122.45	2449	−4	−80	320	−1280	5120
125 – 129.9	35	127.45	4460.75	1	35	35	35	35
130 – 134.9	10	132.45	1324.5	6	60	360	2160	12960
135 – 139.9	10	137.45	1374.5	11	110	1210	13310	146410
140 – 144.9	5	142.45	712.25	16	80	1280	20480	327680
	100		12645	7	0	5400	10050	782700

$$\bar{x} = \frac{\sum f_i x_i}{N} = \frac{12645}{100} = 126.45$$

$$\mu_1 = \frac{1}{N}\sum fd = 0, \quad \mu_2 = \frac{1}{N}\sum fd^2 = \frac{5400}{100} = 54,$$

$$\mu_3 = \frac{1}{N}\sum fd^3 = 100.5, \quad \mu_4 = \frac{1}{N}\sum fd^4 = 7827$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\mu_2\sqrt{\mu_2}} = \frac{100.5}{54\sqrt{54}} = 0.2533$$

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{7827}{54^2} - 3 = -0.3158$$

* If the moments are asked about the point 127.45

$$\mu_1' = \bar{x} - a = 126.45 - 127.45 = -1, \quad \mu_2' = \mu_2 + (\mu_1')^2 = 55$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + 3\mu_1'(\mu_1')^2 + (\mu_1')^3 = 100.5 + 3(54)(-1) +$$

$$3(0) + (-1)^3 = -62.5, \quad \mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2(\mu_1')^2 + (\mu_1')^4$$

$$\mu_4' = 7827 + 4(100.5)(-1) + 6(54)(1) + 1$$

$$\therefore \mu_4' = 7750$$

Correlation

Correlation is a statistical method to determine whether a linear relationship exists between the variables.

A positive relationship exists when both variables increase or decrease at the same time.

ex: The no of gallons of gas pumped is positively correlated to amount spent on gas.

In a negative relationship, as one variable increases the other variable decreases and vice versa.

ex: Miles travelled is negatively correlated to amount of gas left in your tank.

If there exists no relationship b/w 2 variables then they are said to be non correlated

ex: amount of chocolate someone eats and how many hours they spend on work.

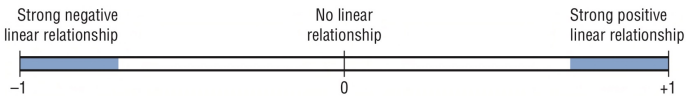
Correlation coefficient (Karl Pearson correlation coefficient) measure to determine the strength of linear relationship b/w 2 variables. It is denoted by r.

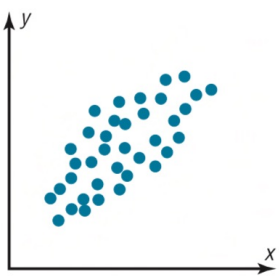
The range of r is from -1 to 1. If there is a strong linear relationship b/w variables the value of r will be close to 1. If there is a strong negative linear relationship b/w variables, the value of r will be close to -1. When there is no linear relationship b/w variables or only weak relationship, the value of r will be close to 0.

Formula for the Correlation Coefficient *r*

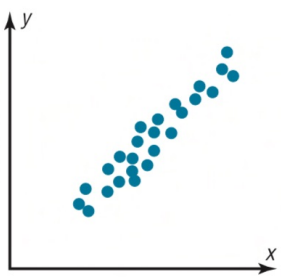
$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

where n is the number of data pairs.

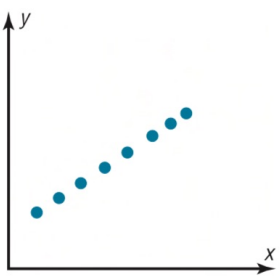




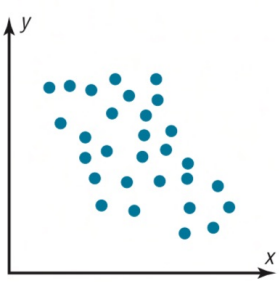
(a) $r = 0.50$



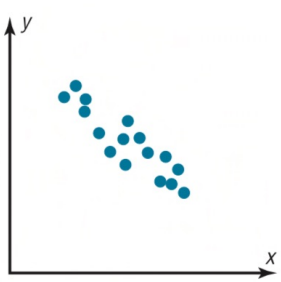
(b) $r = 0.90$



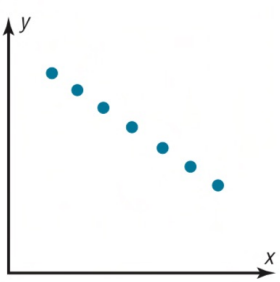
(c) $r = 1.00$



(d) $r = -0.50$



(e) $r = -0.90$



(f) $r = -1.00$

Alternate formula for r:

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}}$$

where $X_i = x_i - \bar{x}$, $Y_i = y_i - \bar{y}$

Find the correlation coefficient for following data

x	1	2	3	4	5
y	2	5	3	8	7

Sol. $n = 5$

$$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{2+5+3+8+7}{5} = \frac{25}{5} = 5$$

x_i	$X_i = x_i - \bar{x}$	X_i^2	y_i	$Y_i = y_i - \bar{y}$	Y_i^2	$X_i Y_i$
1	-2	4	2	-3	9	6
2	-1	1	5	0	0	0
3	0	0	3	-2	4	0
4	1	1	8	3	9	3
5	2	4	7	2	4	4
		$\sum X_i^2 = 10$			$\sum Y_i^2 = 26$	$\sum X_i Y_i = 13$

$$\therefore r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}} = \frac{13}{\sqrt{10 \times 26}} = \frac{13}{\sqrt{260}} = 0.8060$$

Alternate method:

x	y	x^2	y^2	xy
1	2	1	4	2
2	5	4	25	10
3	3	9	9	9
4	8	16	64	32
5	7	25	49	35
$\Sigma x = 15$	$\Sigma y = 25$	55	151	88

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{(n\Sigma x^2 - (\Sigma x)^2)(n\Sigma y^2 - (\Sigma y)^2)}} = 0.8060$$

There is a strong positive relationship b/w x & y.

2)

Company	Cars (in ten thousands)	Revenue (in billions)
A	63.0	$7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

Sol.

Company	Cars x (in 10,000s)	Revenue y (in billions)	xy	x^2	y^2
A	63.0	7.0	441.00	3969.00	49.00
B	29.0	3.9	113.10	841.00	15.21
C	20.8	2.1	43.68	432.64	4.41
D	19.1	2.8	53.48	364.81	7.84
E	13.4	1.4	18.76	179.56	1.96
F	8.5	1.5	12.75	72.25	2.25
	$\Sigma x = 153.8$	$\Sigma y = 18.7$	$\Sigma xy = 682.77$	$\Sigma x^2 = 5859.26$	$\Sigma y^2 = 80.67$

Step 3 Substitute in the formula and solve for r.

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

$$= \frac{(6)(682.77) - (153.8)(18.7)}{\sqrt{[(6)(5859.26) - (153.8)^2][(6)(80.67) - (18.7)^2]}} = 0.982$$

The correlation coefficient suggests a strong relationship between the number of cars a rental agency has and its annual revenue.

3)

Compute the value of the correlation coefficient for the data obtained in the study of the number of absences and the final grade of the seven students in the statistics class given in table.

Student	Number of absences x	Final grade y (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78

Sol :

Student	Number of absences x	Final grade y (%)	xy	x^2	y^2
A	6	82	492	36	6,724
B	2	86	172	4	7,396
C	15	43	645	225	1,849
D	9	74	666	81	5,476
E	12	58	696	144	3,364
F	5	90	450	25	8,100
G	8	78	624	64	6,084
	$\Sigma x = 57$	$\Sigma y = 511$	$\Sigma xy = 3745$	$\Sigma x^2 = 579$	$\Sigma y^2 = 38{,}993$

Step 3 Substitute in the formula and solve for r.

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

$$= \frac{(7)(3745) - (57)(511)}{\sqrt{[(7)(579) - (57)^2][(7)(38{,}993) - (511)^2]}} = -0.944$$

The value of r suggests a strong negative relationship between a student's final grade and the number of absences a student has. That is, the more absences a student has, the lower is his or her grade.

4) Compute the value of the correlation coefficient for the data given in Example for the age and wealth of the richest persons in the United States.

Person	Age x	Net wealth y ($ billions)
A	73	16
B	65	26
C	53	50
D	54	21.5
E	79	40
F	69	16
G	61	19.6
H	65	19

Sol:

Person	Age x	Net wealth y	xy	x^2	y^2
A	73	16	1,168	5,329	256
B	65	26	1,690	4,225	676
C	53	50	2,650	2,809	2,500
D	54	21.5	1,161	2,916	462.25
E	79	40	3,160	6,241	1,600
F	69	16	1,104	4,761	256
G	61	19.6	1,195.6	3,721	384.16
H	65	19	1,235	4,225	361
	$\Sigma x = 519$	$\Sigma y = 208.1$	$\Sigma xy = 13{,}363.6$	$\Sigma x^2 = 34{,}227$	$\Sigma y^2 = 6{,}495.41$

Step 3 Substitute in the formula and solve for r.

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

$$= \frac{8(13{,}363.6) - (519)(208.1)}{\sqrt{[8(34{,}227) - (519)^2][8(6495.41) - (208.1)^2]}}$$

$$= \frac{-1095.1}{\sqrt{(4455)(8657.67)}}$$

$$= \frac{-1095.1}{6210.469}$$

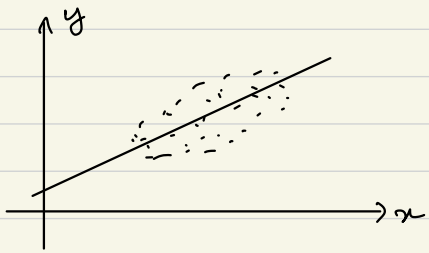
$$= -0.176$$

The value of r indicates a very weak negative relationship between the variables.

Regression

Measures nature and extent of correlation.

Linear regression:



Regression line is the data's line of best fit.

Line of regression of y on x

$y = a + bx$

- y: dep. var.
- x: independent variable
- b: regression coefficient of y on x

Line of regression of x on y

$x = a + by$

- x: dep. var
- y: ind. var
- b: regression coefficient of x on y

Regression line of y on x

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad ; \quad \sigma_y \to \text{S.D. of } y, \ \sigma_x \to \text{S.D. of } x$$

$$\boxed{y - \bar{y} = b_{yx}(x - \bar{x})}$$

where

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma XY}{\Sigma X^2} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

Regression line of x on y

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y}(y - \bar{y}) = \boxed{x - \bar{x} = b_{xy}(y - \bar{y})}$$

where $b_{xy} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum XY}{\sum Y^2} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (y-\bar{y})^2}$

1) Obtain the correlation coefficient and regression lines of y on x and x on y for the following data

x	1	3	4	2	5	8	9	10	13	15
y	8	6	10	8	12	16	16	10	32	32

Sol:

$$\bar{x} = \frac{\sum x}{n} = \frac{70}{10} = 7, \quad \bar{y} = \frac{\sum y}{n} = \frac{150}{10} = 15$$

x	$X = x - \bar{x}$	X^2	y	$Y = y - \bar{y}$	Y^2	XY
1	−6	36	8	−7	49	42
3	−4	16	6	−9	81	36
4	−3	9	10	−5	25	15
2	−5	25	8	−7	49	35
5	2	4	12	−3	9	6
8	1	1	16	1	1	1
9	2	4	16	1	1	2
10	3	9	10	−5	25	−15
13	6	36	32	17	289	102
15	8	64	32	17	289	136
		204			818	360

$$r = \frac{\sum XY}{\sqrt{\sum x^2 \sum Y^2}} = \frac{360}{\sqrt{204 \times 818}} = 0.8812$$

Regression coefficients :

$$b_{yx} = \frac{\sum XY}{\sum X^2} \quad , \quad b_{xy} = \frac{\sum XY}{\sum Y^2}$$

$$b_{yx} = \frac{360}{204} = 1.7647 \quad , \quad b_{xy} = \frac{360}{818} = 0.4401$$

The regression liner of y on x :

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 15 = 1.7647(x - 7)$$

$$\therefore y = 1.7647x + 2.6471$$

The regression liner of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 7 = 0.4401(y - 15)$$

$$\therefore x = 0.4401y + 0.3985$$

2) In a partially destroyed lab, record of analysis of correlation data, the following results only are legible. variance of $x = 9$, Regression equations $8x - 10y + 66 = 0$, $40x - 18y = 214$. What are i) mean values of x & y ii) correlation coefficient b/w x & y iii) S.D. of y

Sol:

i) Since both lines pass through $(\bar{x}, \bar{y})$

$$8\bar{x} - 10\bar{y} + 66 = 0,$$
$$40\bar{x} - 18\bar{y} - 214 = 0$$

Solving, $\bar{x} = 13$, $\bar{y} = 17$

ii) $\sigma_x^2 = 9$, $\sigma_x = 3$

$8x - 10y + 66 = 0$, $40x - 18y = 214$

$y = 0.8x + 6.6$, $x = 0.45y + 5.35$

$b_{yx} = 0.8$, $b_{xy} = 0.45$

$$r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{0.36} = \pm 0.6$$

Since both regression coefficients are +ve, we take $r = 0.6$.

iii) $b_{yx} = r \frac{\sigma_y}{\sigma_x}$

$$\therefore \sigma_y = \frac{b_{yx} \times \sigma_x}{r} = \frac{0.8 \times 3}{0.6} = 4$$

$\therefore$ S.D. of $y = 4$

3) Find the equation of regression line, of y on x for the data

Student	Number of absences x	Final grade y (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78

Ans: $y = 102.493 - 3.622x$

4) Find the equation of regression line, of y on x for the data

Company	Cars (in ten thousands)	Revenue (in billions)
A	63.0	$7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

Ans: $y = 0.396 + 0.106x$