Karl's Pearson Correlation

Dr. Anshul Singh Thapa

Karl's Pearson Correlation

- The Pearson's Correlation coefficient is usually calculated for two continuous variables. If either or both the variables are not continuous than some other statistical procedure are to be used.

- Understanding product moment correlation requires understanding of mean, variance and covariance.

- **Mean:** Mean of variable X and variable Y have to be calculated separately. Mean of variable is sum of scores divided by the number of observations.
- **Variance:** The variance of variable X and variable Y have to be calculated separately. The variance of the variable is the sum of squares of the deviations of each score from the mean and divided by number of observation.
- **Covariance:** the covariance is the number that indicates the association between the two variables. To compute covariance, deviation of each score of the variable from its mean and deviation of each score of the variable from its mean is initially calculated. Then product of these deviations are obtained. Then, these products are summated. This sum gives the numerator for covariance. Divide this sum by the number of observation. The resulting value is covariance.

Example

No. of years of schooling of farmers X	Annual yield per acre in, 000 (Rs) Y
0	4
2	4
4	6
6	10
8	10
10	8
12	7

$$r = \frac{\Sigma\,(X - \overline{X})\,(Y - \overline{Y})}{\sqrt{\Sigma\,(X - \overline{X})^2}\,\sqrt{(Y - \overline{Y})^2}}$$

Mean (X)
Education X
0
2
4
6
8
10
12

	Mean (X)
Education X	(X - X̄)
0	-6
2	-4
4	-2
6	0
8	2
10	4
12	6

Mean (X)		Variance (X)
Education X	$(X - \bar{X})$	$(X - \bar{X})^2$
0	-6	36
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16
12	6	36

Mean (X)		Variance (X)	Mean (Y)
Education X	(X - X̄)	(X - X̄)²	Annual Yield Y
0	-6	36	4
2	-4	16	4
4	-2	4	6
6	0	0	10
8	2	4	10
10	4	16	8
12	6	36	7

Mean (X)		Variance (X)	Mean (Y)	
Education X	(X - X)	(X - X)^2	Annual Yield Y	(Y - Y)
0	-6	36	4	-3
2	-4	16	4	-3
4	-2	4	6	-1
6	0	0	10	3
8	2	4	10	3
10	4	16	8	1
12	6	36	7	0

Mean (X)		Variance (X)	Mean (Y)		Variance (Y)
Education X	(X - X)	(X - X)^2	Annual Yield Y	(Y - Y)	(Y – Y)^2
0	-6	36	4	-3	9
2	-4	16	4	-3	9
4	-2	4	6	-1	1
6	0	0	10	3	9
8	2	4	10	3	9
10	4	16	8	1	1
12	6	36	7	0	0

Mean (X)		Variance (X)	Mean (Y)		Variance (Y)	Covariance (XY)
Education X	(X - X)	(X - X)^2	Annual Yield Y	(Y - Y)	(Y – Y)^2	(X - X) (Y - Y)
0	-6	36	4	-3	9	18
2	-4	16	4	-3	9	12
4	-2	4	6	-1	1	2
6	0	0	10	3	9	0
8	2	4	10	3	9	6
10	4	16	8	1	1	4
12	6	36	7	0	0	0

Mean (X)		Variance (X)	Mean (Y)		Variance (Y)	Covariance (XY)
Education X	(X - X)	(X - X)^2	Annual Yield Y	(Y - Y)	(Y – Y)^2	(X - X) (Y - Y)
0	-6	36	4	-3	9	18
2	-4	16	4	-3	9	12
4	-2	4	6	-1	1	2
6	0	0	10	3	9	0
8	2	4	10	3	9	6
10	4	16	8	1	1	4
12	6	36	7	0	0	0
ΣX = 42		**Σ(X - X)^2 = 112**	**ΣY = 49**		**Σ(Y – Y)^2 = 38**	**Σ(X - X) (Y - Y) = 42**

Mean (X)		Variance (X)	Mean (Y)		Variance (Y)	Covariance (XY)
Education X	(X - X)	(X - X)²	Annual Yield Y	(Y - Y)	(Y − Y)²	(X - X) (Y - Y)
0	-6	36	4	-3	9	18
2	-4	16	4	-3	9	12
4	-2	4	6	-1	1	2
6	0	0	10	3	9	0
8	2	4	10	3	9	6
10	4	16	8	1	1	4
12	6	36	7	0	0	0
ΣX = 42		Σ(X - X)² = 112	ΣY = 49		Σ(Y − Y)² = 38	Σ(X - X) (Y - Y) = 42

$$r = \frac{\Sigma\,(X - \overline{X})(Y - \overline{Y})}{\sqrt{\Sigma\,(X - \overline{X})^2}\;\;\sqrt{\Sigma\,(Y - \overline{Y})^2}}$$

Mean (X)		Variance (X)	Mean (Y)		Variance (Y)	Covariance (XY)
Education X	$(X - \overline{X})$	$(X - \overline{X})^2$	Annual Yield Y	$(Y - \overline{Y})$	$(Y - \overline{Y})^2$	$(X - \overline{X})(Y - \overline{Y})$
0	-6	36	4	-3	9	18
2	-4	16	4	-3	9	12
4	-2	4	6	-1	1	2
6	0	0	10	3	9	0
8	2	4	10	3	9	6
10	4	16	8	1	1	4
12	6	36	7	0	0	0
$\Sigma X = 42$		$\Sigma(X - \overline{X})^2 = 112$	$\Sigma Y = 49$		$\Sigma(Y - \overline{Y})^2 = 38$	$\Sigma(X - \overline{X})(Y - \overline{Y}) = 42$

$$r = \frac{\Sigma (X - \overline{X})(Y - \overline{Y})}{\sqrt{\Sigma (X - \overline{X})^2} \; \sqrt{\Sigma (Y - \overline{Y})^2}}$$

$$r = \frac{42}{\sqrt{112} \; \sqrt{38}} = 0.644$$

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

- Where ;

    $\sum X$ = sum of all x scores

    $\sum Y$ = sum of all y scores

    $\sum X^2$ = sum of squared of all x scores

    $\sum Y^2$ = sum of squared of all y scores

    $\sum XY$ = sum of multiplication of each x score by the corresponding y score

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

No. of years of schooling of farmers X	Annual yield per acre in, 000 (Rs) Y
0	4
2	4
4	6
6	10
8	10
10	8
12	7

X	Y	X²
0	4	0
2	4	4
4	6	16
6	10	36
8	10	64
10	8	100
12	7	144

X	Y	X²	Y²
0	4	0	16
2	4	4	16
4	6	16	36
6	10	36	100
8	10	64	100
10	8	100	64
12	7	144	49

X	Y	X²	Y²	XY
0	4	0	16	0
2	4	4	16	8
4	6	16	36	24
6	10	36	100	60
8	10	64	100	80
10	8	100	64	80
12	7	144	49	84

X	Y	X²	Y²	XY
0	4	0	16	0
2	4	4	16	8
4	6	16	36	24
6	10	36	100	60
8	10	64	100	80
10	8	100	64	80
12	7	144	49	84
ΣX = 42	ΣY = 49	ΣX² = 364	ΣY² = 381	ΣXY = 336

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Calculate Correlation of Coefficient

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

Calculate Correlation of Coefficient

X	Y
1	5
5	6
6	7
8	9
10	12