Related Concept, Formula/Equations and Solved Examples

(1). Concept of Variance

- Var (X) = $[(x_1-\bar{x})^2+ (x_2-\bar{x})^2+ (x_3-\bar{x})^2+.......+(x_n-\bar{x})^2]/n$ ....(1)
- **Types**: See the below sheet

Population Variance	Sample Variance
$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
σ^2 = population variance x_i = value of i^{th} element μ = population mean N = population size	s^2 = sample variance x_i = value of i^{th} element $\bar{x}$ = sample mean n = sample size

Example: Let's say the heights (in mm) are 610, 450, 160, 420, 310.
Step-I : compute the Mean values:: Mean = (610+450+160+420+310)/ 5 = 390
Step-II: to calculate the Variance, compute the difference of each from the mean, square it and find then find the average once again.
So for this particular case the variance is := $(220^2 + 60^2 + (-230)^2 +30^2 + (-80)^2)/5$= (48400 + 3600 + 52900 + 900 + 6400)/5::
Variance = 22440

(2) Concept of Co-variance:

- **Type::**

Population Covariance

$$Cov(X,Y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}$$

Sample Covariance

$$Cov(X,Y) = \frac{\Sigma(x_i-\bar{x})(y_i-y)}{N-1}$$

These are the formula for finding Population and Sample Covariance.

where,

- x_i = data value of x
- y_i = data value of y
- $\bar{x}$ = mean of x
- $\bar{y}$ = mean of y
- N = number of data values.

Example:

How to Calculate the Covariance?

2 data sets X = (2, 4, 6 ,8, 10) Y = (12, 11, 8, 3, 1)

step 1: find the mean (average) of both sets

$$\bar{x} = \frac{\Sigma x_i}{n} = 6 \qquad \bar{y} = \frac{\Sigma y_i}{n} = 7$$

step 2: find the variance of both sets

$$\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = 8 \qquad \sigma_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n} = 18.8$$

step 3: find the covariace

$$\text{cov}(x,y) = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n} = -12$$

(3) Concept of Correlations:

Related Concept, Formula/Equations and Solved Examples

Correlation Coefficient Formula

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

……………………..(2)

- **Types**

Correlation coefficient	Type of relationship	Levels of measurement
Pearson's r	Linear	Two quantitative (interval or ratio) variables
Spearman's rho	Non-linear	Two ordinal, interval or ratio variables
Point-biserial	Linear	One dichotomous (binary) variable and one quantitative (interval or ratio) variable
Cramér's V (Cramér's φ)	Non-linear	Two nominal variables
Kendall's tau	Non-linear	Two ordinal, interval or ratio variables

- **Examples:**

A. Direct Method

Type I : This method is used when given variables are small in magnitude.

Formula : $r = \dfrac{N\,\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{N\,\Sigma X^2 - (\Sigma X)^2}\,\sqrt{N\,\Sigma Y^2 - (\Sigma Y)^2}}$

Example 1. Calculate Karl Pearson's coefficient of correlation between the age and weight of the children :

Age (years) :	1	2	3	4	5
Weight (kg.) :	3	4	6	7	12

Solution : $\Sigma X = 15$; $\Sigma Y = 32$; $\Sigma X^2 = 55$; $\Sigma Y^2 = 254$; $\Sigma XY = 117$

Age (X)	Weight (Y)	X^2	Y^2	XY
1	3	1	9	3
2	4	4	16	8
3	6	9	36	18
4	7	16	49	28
5	12	29	144	60
15	32	55	254	117

As $r = \dfrac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\,\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$

$\therefore\ r = \dfrac{5 \times 117 - 15 \times 32}{\sqrt{5 \times 55 - (15)^2}\,\sqrt{5 \times 254 - (32)^2}}$

$= \dfrac{585 - 480}{\sqrt{275 - 225}\,\sqrt{1270 - 1024}} = \dfrac{105}{\sqrt{50 \times 246}} = \dfrac{105}{\sqrt{12300}} = \dfrac{105}{110.90} = 0.9467$ **Ans.**

Type II : It is direct formula to find r. This formula can effectively be used where $\overline{X}$ and $\overline{Y}$ is not in fractions. The formula is

$r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}}$; where dx is the deviation of X variable from its $\overline{X}$.

y is the deviation of Y variable from its $\overline{Y}$. ; xy is the product of the two above dx^2 is the square of x ; y^2 is the square of dy.

Example 2. Calculate coefficient of correlation between death and birth rate for the following data.

Birth Rate	24	26	32'	33	35	30
Death Rate	15	20	22	24	27	24

Solution

Birth Rate X	Death Rate Y	$(X - \overline{X}) = x$	$(Y - \overline{Y}) = y$	$(X - \overline{X})^2 = x^2$	$(Y - \overline{Y})^2 = y^2$	$(X - \overline{X})(Y - \overline{Y}) = xy$
24	15	–6	–7	36	49	42
26	20	–4	–2	16	4	8
32	22	2	0	4	0	0
33	24	3	2	9	4	6
35	27	5	5	25	25	25
30	24.	0	2	0	4	0
$\Sigma X = 180$ $\overline{X} = \frac{180}{6} = 30$	$\Sigma Y = 132$ $\overline{Y} = \frac{132}{6} = 22$	$\Sigma x = 0$	$\Sigma y = 0$	$\Sigma x^2 = 90$	$\Sigma y^2 = 86$	$\Sigma xy = 81$

$r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}} = \dfrac{(81)}{\sqrt{90 \times 86}} = \dfrac{81}{\sqrt{7740}} = \dfrac{81}{87.98} = .92$

(4) Chi-Square hypothesis testing:

Related Concept, Formula/Equations and Solved Examples

$$X^2 = \sum \frac{(\text{Observed value - Expected value})^2}{\text{Expected value}} \qquad \text{.....................(3)}$$

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i} \qquad \text{............................(4)}$$

➢ **Significance level for specific Degree of Freedom on Chi-square Test:**

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

INTRODUCTION TO POPULATION GENETICS, Table D.1

➢ **Example:**

	Observed Value			
	High income	Med income	Low income	
B.Tech	20	25	15	60
M.Tech	15	20	5	40
	35	45	20	100

	Expected Value		
	High income	Med income	Low income
B.Tech	21	27	12
M.Tech	14	18	8

Chi-Sq Stats = ∑[(Observed - Expected)^2] / Expected			
	High income	Med income	Low income
B.Tech	0.0476	0.1481	0.75
M.Tech	0.0714	0.2222	1.125

Chi-Square Statistics =	2.3643
Degree of freedom =	2
P-Value =	0.3066