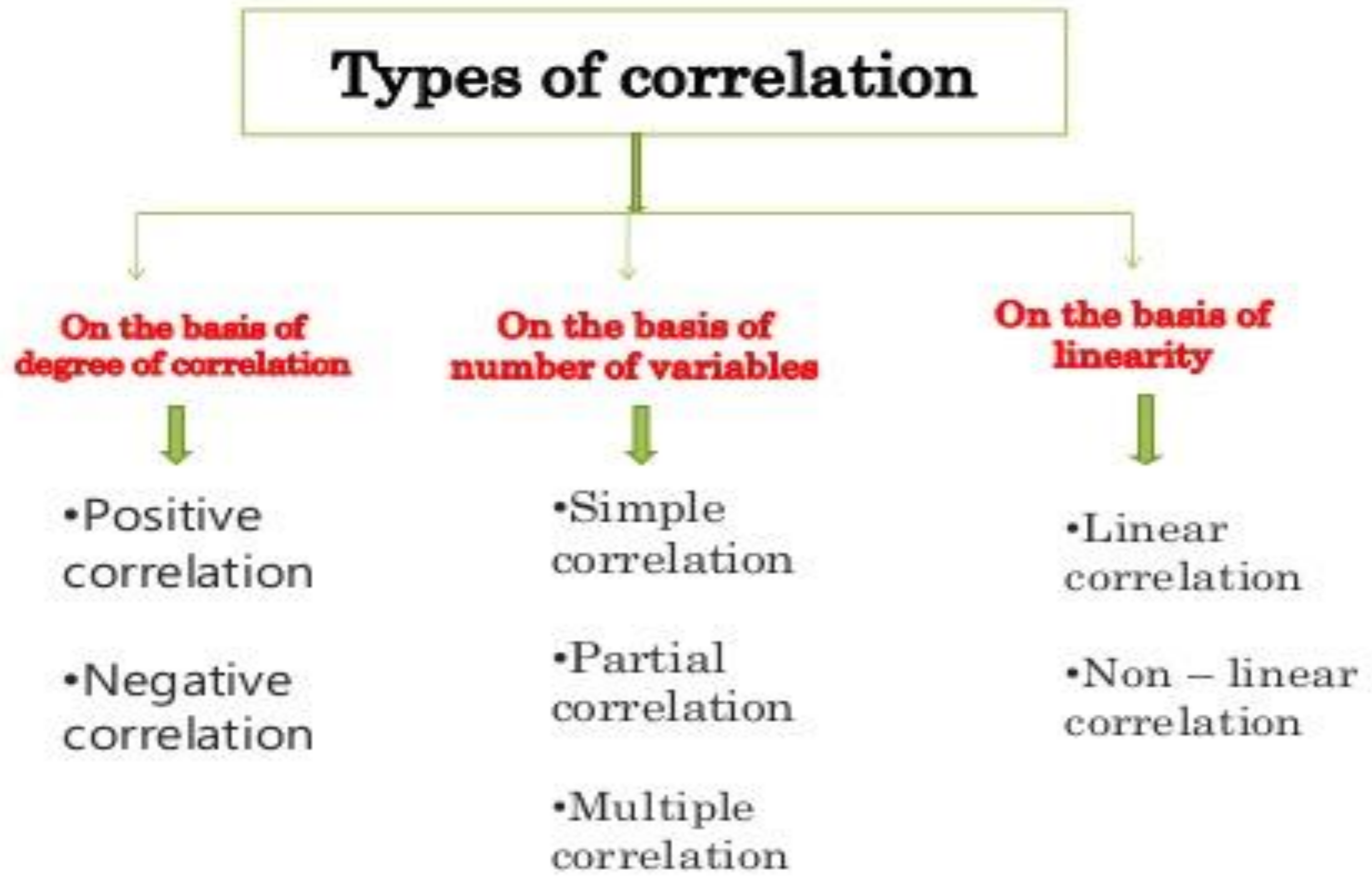
Unit IV

- Correlation analysis- Partial correlation – partial correlation coefficient – zero order - first order and second order coefficients - uses and limitations – multiple correlation : coefficient of multiple correlation –advantages and limitations

Meaning

- Correlation analysis- is a statistical method used to evaluate the strength of relationship between two quantitative variables.
- A high **correlation** means that two or more variables have a strong relationship with each other, while a weak **correlation** means that the variables are hardly related.

Types


Types of correlation
On the basis of degree of correlation
•Positive correlation
•Negative correlation
On the basis of number of variables
•Simple correlation
•Partial correlation
•Multiple correlation
On the basis of linearity
•Linear correlation
•Non – linear correlation

Meaning

- **Correlation** is a bivariate analysis that measures the strength of association between two variables and the direction of the relationship. In terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1.
- A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker.
- The direction of the relationship is indicated by the sign of the coefficient; a + sign indicates a positive relationship and a – sign indicates a negative relationship.
- Usually, in statistics, we measure four types of correlations: Pearson correlation, Kendall rank correlation, Spearman correlation, and the Point-Biserial correlation.

Partial correlation

- Partial correlation – measures the degree of association between two random variables, with the effect of a set of controlling random variables removed.
- **Partial correlation** is the measure of association between two variables, while controlling or adjusting the effect of one or more additional variables.
- Partial correlations can be used in many cases that assess for relationship, like whether or not the sale value of a particular commodity is related to the expenditure on advertising when the effect of price is controlled.

Partial correlation

- **Partial correlation coefficient** is a **coefficient** to describe the relationship between X and Y when taking away the effects of control variable Z , which can be used to test conditional independence.

Zero order, First order and Second order coefficients

- **Partial correlation such** as $r_{12.3}$, $r_{13.2}$ are often referred to as first order coefficients, since one variable has been held constant. Coefficients correlation between two variables only are called zero order coefficients.
- Since no variables are held constant $r_{12.34}$, $r_{13.24,}$ $^{etc.,}$ are called second order coefficients since two variables are kept constant.

Zero order, First order and Second order coefficients

Order of correlation

- Zero order correlation: $r_{12}, r_{13}, r_{23}, \ldots\ldots\ldots$
- First order correlation: $r_{12.3}, r_{13.2}, r_{23.1}, \ldots\ldots\ldots$
- Second order correlation: $r_{12.34}, r_{13.24}, \ldots\ldots\ldots$

Partial correlation

For Zero order correlation,

$$r_{12} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1{}^2 - (\sum X_1)^2}\sqrt{n\sum X_2{}^2 - (\sum X_2)^2}}$$

$$r_{13} = \frac{n\sum X_1X_3 - \sum X_1 \sum X_3}{\sqrt{n\sum X_1{}^2 - (\sum X_1)^2}\sqrt{n\sum X_3{}^2 - (\sum X_3)^2}}$$

And

$$r_{23} = \frac{n\sum X_2X_3 - \sum X_2 \sum X_3}{\sqrt{n\sum X_2{}^2 - (\sum X_2)^2}\sqrt{n\sum X_3{}^2 - (\sum X_3)^2}}$$

Partial correlation

What is partial correlation co-efficient?

➢ Relationship between two variables keeping the other variable constant/fixed

- Relation between X_1 and X_2 keeping X_3 constant is denoted by: $r_{12.3}$

- Similarly, $r_{13.2}$ means relation between X_1 and X_3 keeping X_2 and so on

Partial correlation

Calculating partial correlation

- Based on zero order correlation
- Formulas:

$$r_{12.3} = \frac{r_{12} - r_{13}.r_{23}}{\sqrt{1-(r_{13})^2}\sqrt{1-(r_{23})^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}.r_{23}}{\sqrt{1-(r_{12})^2}\sqrt{1-(r_{23})^2}}$$

$$r_{23.1} = \frac{r_{23} - r_{12}.r_{13}}{\sqrt{1-(r_{12})^2}\sqrt{1-(r_{13})^2}}$$

Contn…

Properties of Correlation co-efficient

- Its value lies between -1 to +1.
- $r_{12} = r_{21}, r_{13} = r_{31}$ and $r_{23} = r_{32}$
- $r_{12.3} = r_{21.3}, r_{13.2} = r_{31.2}$ and $r_{23.1} = r_{32.1}$

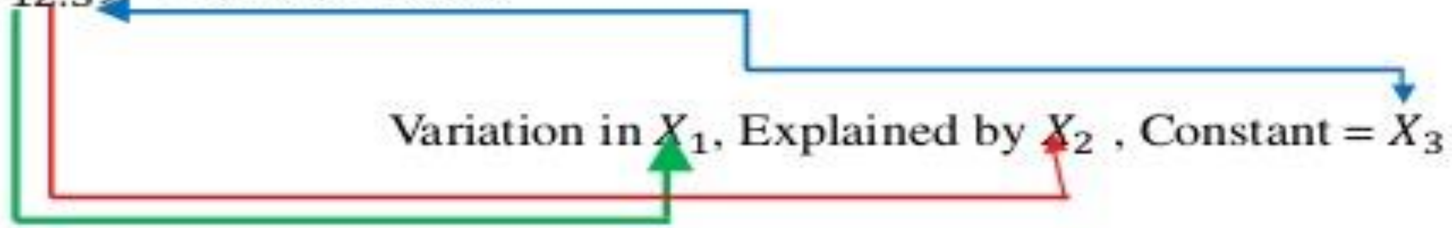
Partial correlation

Co-efficient of partial determination

- Square of partial correlation coefficient
- Also known as the percent of variation
- Used to measure variation in one variable explained by other variable keeping next variable constant

Example: If $r_{12.3} = 0.5$, then partial determination is:

$(r_{12.3})^2 = 0.25 = 25\%$



Multiple correlation

Multiple correlation

- Relation between three/more variables at the same time
- Denoted by R
- If $R<1$ ($r<1$), more consistent
- If $R>1$ ($r>1$), less consistent

Multiple correlation

Calculating Multiple Correlation

- If X_1, X_2 and X_3 are three variable then,

$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{23})^2}}$
$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{13})^2}}$
$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{13})^2}}$

Uses and limitations

- Partial correlation is a method used to describe the relationship between two variables while taking away the effects of another variable, or several other variables, on this relationship.
- Partial correlation is best thought of in terms of multiple regression;
- Limitations- the gross or zero order correlation must have linear regressions
- The effect of the independent variables must be additively and not jointly related.

Multiple correlation

- In statistics, the coefficient of **multiple correlation** is a measure of how well a given variable can be predicted using a linear function of a set of other variables.
- It is the **correlation** between the variable's values and the best predictions that can be computed linearly from the predictive variables.

Multiple correlation

Calculating Multiple Correlation

➢ If X_1, X_2 and X_3 are three variable then,

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{23})^2}}$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{13})^2}}$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{13}.r_{23}r_{12}}{1-(r_{13})^2}}$$

Types of Correlation

- **Types**
- Positive, Negative or Zero Correlation:
- Linear or Curvilinear Correlation:
- Scatter Diagram Method:
- Pearson's Product Moment Co-efficient of Correlation:
- **Spearman's Rank Correlation Coefficient**:

Multiple correlation

Advantages

- The **multiple correlation** is used to find the **impact** of **more than one variable** on the **output value**
- The other advantage of **multiple correlation** is that it helps in finding the **faults** or errors in the **data**
- The **multiple correlation** helps in comparing the d**ifferent "variable"** with **another" variable"** and gives a **better result**.

Disadvantages

- The problem of using **multiple correlation** is when the data is **"incomplete"** or it has **some errors**
- The **calculation involved** in multiple correlation is a **complicated process** and **not easy** to being used
- The problem with **multiple correlation** is that when the **correlation increases** then it is **not** very **effective**

Refereces

- https://www.slideshare.net/UdaybhaskarMogallapu/correlation-and-partial-correlation