# BLOCK 3
# CORRELATION

# BLOCK INTRODUCTION

Block 3 of this course mainly focuses on the topic correlation. The block has been divided in to two units unit 6 and unit 7. **Unit six** focuses on the theoretical information about correlation, including the concept, direction and magnitude of correlation. Properties, uses and limitations of correlation will also be covered. **Unit seven**, on the other hand, deals with the computation of correlation where you will learn to compute correlation with the help of Pearson's product moment correlation and Spearman's rank order correlation.

# UNIT 6 CORRELATION: AN INTRODUCTION*

**Structure**



## 6.0 OBJECTIVES

After reading this unit, you will be able to:

- explain the concept, direction and magnitude of correlation;
- discuss the properties, uses and limitations of correlation; and
- describe other methods of correlation.

## 6.1 INTRODUCTION

Let us focus on some of the statements below:

As there is rise in temperature, there is increase in sale of ice creams.

As there is increase in occupational stress experienced by employees, their performance may go down.

Sale of air purifiers can be linked to the increase in pollution.

In our day to day life, we may experience such associations, that is, when one aspect increases or decreases, another aspect also increases or decreases or when one aspect increases, the other decreases and vice versa.

To take more examples, a researcher may be interested in studying whether there exists a relationship between self esteem and achievement motivation of

* Prof. Suhas Shetgovekar, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi

adolescents. Or a researcher may be interested in finding out if there exists any relationship between psychological wellbeing and job satisfaction of employees in an organisation. In such cases, correlation will help the researcher study the relationship and also understand its direction and magnitude.

In research studying relationship between two or more variables could be an important objective and this can be studied with the help of correlation.

In the present unit, we will focus on the concept of correlation, its direction and magnitude. The properties, uses and limitations of correlation will also be covered besides other methods of correlation.

## 6.2 CONCEPT OF CORRELATION, DIRECTION AND MAGNITUDE OF CORRELATION

Let us take another example, to understand the concept of correlation.

A research was carried out by a researcher on relationship between years of experience and monthly income earned by junior managers in an organisation (hypothetical data). The data given is as follows:

**Table 6.1: Years of experience and monthly income of junior managers**

| Junior Managers | Years of Experience | Monthly income |
|---|---|---|
| John | 1 | 25, 000 |
| Ravi | 2 | 30, 000 |
| Maria | 3 | 35, 000 |
| Kuldeep | 4 | 40, 000 |
| Salma | 5 | 45, 000 |

Looking at this data, what do you understand? You may notice that as the years of experience is increasing, the monthly income earned by the junior managers is also increasing. Thus, it can be inferred that in case of these junior managers, there is a positive relationship between the years of experience and the monthly income.

Let us take another example,

An experimenter was carrying out a study on relationship between hours of practice in a certain task and number of errors committed by the participants. The data obtained from the same is given as follows:

**Table 6.2: Hours of practice in certain task and number of errors**

| Participants | Hours of Practice | Number of errors |
|---|---|---|
| Rehman | 4 | 20 |
| Sophia | 5 | 12 |
| Navjyot | 7 | 8 |
| Anjali | 1 | 30 |
| Rahul | 2 | 24 |

As we look at this data, it can be seen that as the hours of practice is increasing, the number of errors committed by the participants is decreasing. Thus, it can be said that there is a negative relationship between hours of practice and number of errors committed.

With the above examples, you must have developed some idea about what is correlation. Let us now look at the concept of correlation and also focus on its direction and magnitude.

Correlations can be used to study relationship between two or more variables. It is a measure of association between two or more variables and this relationship is determined not only in terms of direction, whether negative or positive but also in terms of its magnitude, whether high or low. However, it will not provide information about any causal relationship between the variables.

Sir Francis Galton's contribution to development of correlation is noteworthy. He carried out studies on individual differences and also studies on the influence of heredity. He studied the association between the height of parents and that of their children with the help of bivariate distribution (that studies relationship between two variables) and found that the parents who are tall have children who are also tall (Veeraraghavan and Shetgovekar, 2016). Further, in 1986, Karl Pearson put forth mathematical procedure for correlation.

Correlation can be categorised in to linear and nonlinear correlation. These are discussed as follows:

**Linear Correlation:** Linear correlation is denoted by a single straight line in a graph that denotes linear relationship between given two variables. Such a graph indicates whether increase in one variable leads to increase in another variable and vice versa, or decrease in one variable leads to increase in another variable and vice versa. For example, if the scores on emotional intelligence increase or decrease, the scores on self esteem also increase or decrease. Linear relationship is graphically represented in figure 6.1.

**Nonlinear correlation:** As opposed to linear relationship, in nonlinear relationship. The relationship between two given variables is not denoted by a straight line. Thus, the relationship is curvilinear as denoted in figure 6.2.

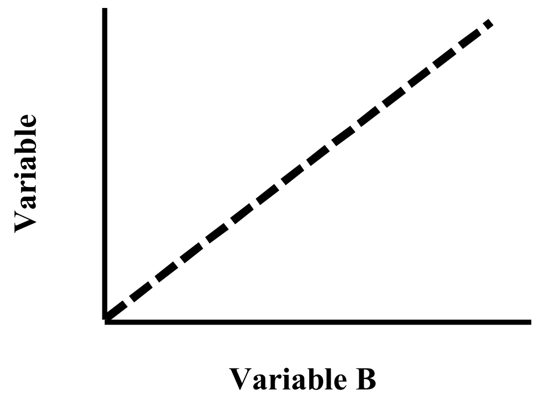


**Fig. 6.1: Linear Correlation**

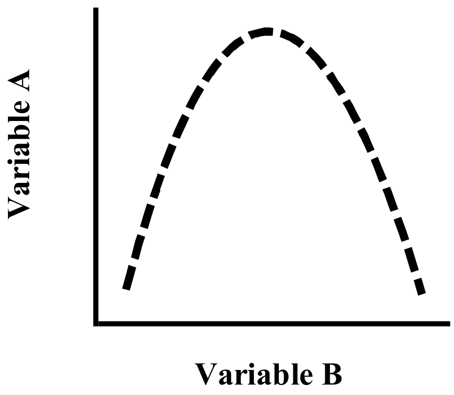


**Fig. 6.2: Nonlinear Correlation**

### 6.2.1 Direction and Magnitude of Correlation

With the help of examples that we discussed at the start of this section, it must be clear that correlation can be either positive or negative (though there could also be no relationship between the given two variables). This can be described as direction of correlation. Let us now discuss these in detail.

**Positive correlation:** Positive correlation denotes that increase in one variable leads to increase in another variable and decrease in one variable leads to decrease in another variable. For example, if the scores on emotional intelligence obtained by adolescents increase, then the scores obtained by them on achievement motivation will also increase or if the scores on emotional intelligence obtained by adolescents decrease then the scores obtained by them on achievement motivation will also decrease. Positive correlation indicates that both the variables are moving in same direction (refer to figure 6.3).

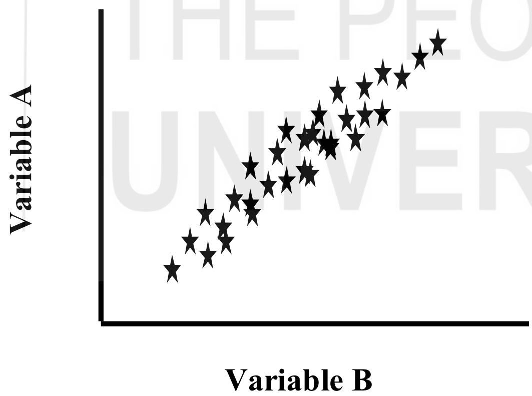


**Fig. 6.3: Positive Correlation**

Figure 6.3 is a scatter diagram denoting positive relationship between two variables, A and B. Scatter diagram can be effectively used to present a bivariate distribution that denotes relationship between the two variables.

**Negative Correlation:** Negative correlation denotes that increase in one variable leads to decrease in another variable or decrease in one variable leads to increase in another variable. For example, if the scores on occupational stress obtained by employees increase, then the scores obtained by them on work motivation will decrease or if the scores on occupational stress obtained by employees decrease, then the scores obtained by them on work motivation will increase. In this case, the two variables are not moving in same direction. Figure 6.4 is a diagrammatic representation of negative correlation.

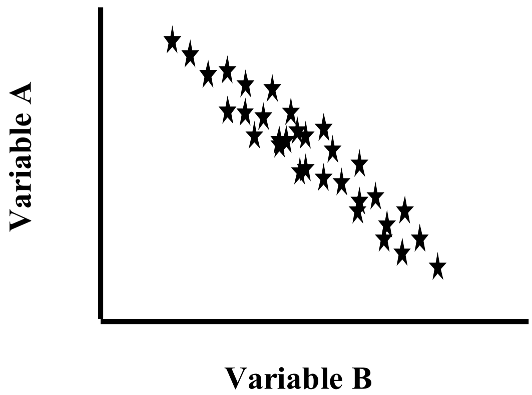


Fig. 6.4: Negative Correlation

**No Correlation or Zero Correlation:** It may so happen that there is no relationship between the two variables. In such a case the correlation will be zero (this will be further clear as we discuss the magnitude of correlation). Thus, in this case the relationship is neither positive or negative. There are such variables where there might be no relationship, for example, there may be no correlationship between height of persons and years of their work experience or there may exist no relationship between weight of persons and attitude towards environment. No correlation is represented in form of scatter diagram in figure 6.5.

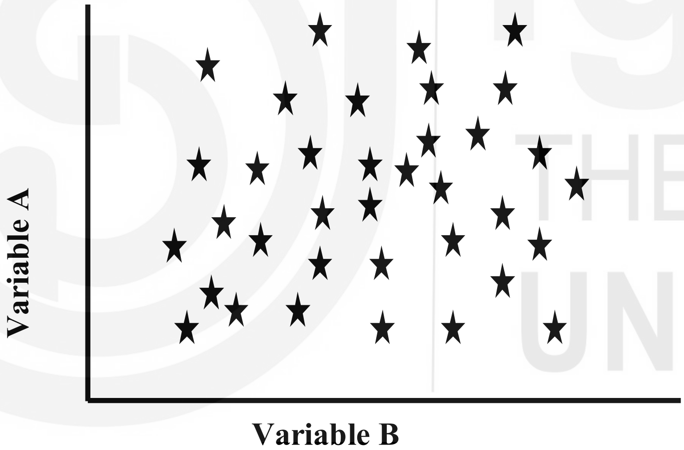


Fig. 6.5: No or Zero Correlation

Besides the direction of the correlation, it is also significant to understand the magnitude or strength of the correlation. Magnitude is denoted by the degree of linearity of the correlationship. Correlationship between any two variables is Coefficient of Correlation that is quantitatively represented. And the range for a coefficient of correlation is between -1 to +1. Thus, coefficient of correlation can be obtained as 0.28 or -0.09 or 0.75 and so on. The number will lie between -1 to +1 and the + and - signs denote the direction of correlation, whether it is positive or negative. The obtained coefficient of correlation can be interpreted with the help of the table 6.3 given as follows (Mangal, 2002, page 105):

| Table 6.3: Interpretation of Coefficient of Correlation | |
|---|---|
| **Coefficient of Correlation Range** | **Interpretation** |
| **+ 1 or - 1** | This can be interpreted as a correlation that is perfect, though the direction could be positive or negative. |
| **± 0.91 to 0.99** | Correlation is very high |
| **± 0.71 to 0.90** | Correlation is high |
| **± 0.41 to 0.70** | Correlation is moderate |
| **± 0.21 to 0.40** | Correlation is low |
| **0 to ± 0.20** | Correlation is negligible |
| **0** | No correlation |

While interpreting coefficient of correlation it is important to keep in mind the direction based on the positive and negative signs.

### 6.2.2 Scatter Diagram

One way in which relationship between two variables can be denoted is by using scatter diagram. Scatter diagram is also called as scatter plot or scatter gram. It is drawn by plotting the two variables in same graph, that is variable A on y axis and variable B on x axis (as shown in figures 6.3, 6.4 and 6.5).

| Table 6.4: Marks obtained by students in Psychology and Sociology class test | | |
|---|---|---|
| **Students** | **Marks in Psychology** | **Marks in Sociology** |
| **Peter** | 25 | 26 |
| **Sarika** | 45 | 42 |
| **Kamaldeep** | 78 | 73 |
| **Salman** | 5 | 4 |
| **Arvind** | 89 | 84 |

For example, we want to study the relationship between marks obtained by five students in Psychology and in Sociology in class test. The data is given in table 6.4.

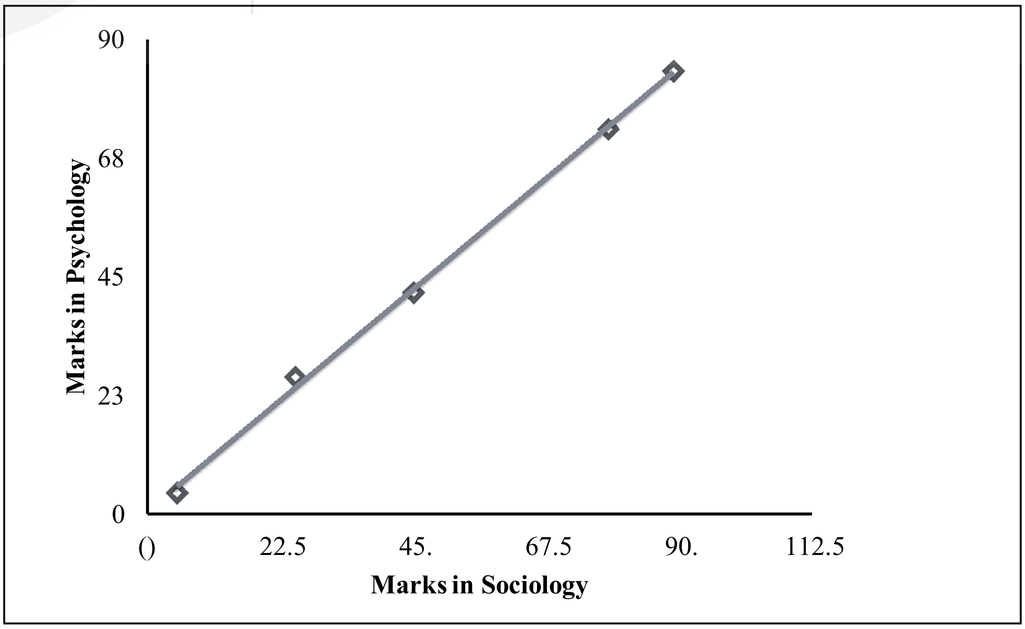


**Fig. 6.6: Scatter diagram based on table 6.4**

As can be seen from the graph, there is a linear relationship between the marks obtained in psychology and marks obtained in sociology.

Coefficient of Correlation can be computed with the help of Pearson's product moment correlation and Spearman's rank order correlation that will be discussed in the next unit.

**Check Your Progress I**

1) What is Correlation?

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

## 6.3 PROPERTIES, USES AND LIMITATIONS OF CORRELATION

As the concept of correlation is now clear, we will discuss about its properties, uses and limitations.

### 6.3.1 Properties of Correlation

1) The variables used in correlation are quantitative in nature.

2) The value of coefficient of correlation will range from -1 to +1 and correlation can be positive or negative or there can also be zero correlation.

3) As we have discussed before, correlation will provide information about relationship between the variables, but it will not denote whether a cause and effect relationship exists between the variables. For example, we may obtain a positive correlation between organisational culture and job satisfaction of employees, but correlation will not denote if there is a cause and effect relationship between these two variables.

4) It is not possible to make any predictions based on correlation. For example, there may be positive correlation between rise in temperature and sale of ice cream or cold drinks but the sale of ice cream or cold drinks cannot be predicted based on the temperature with the help of correlation.

5) Even if variables randomly vary, correlation can be used.

6) Sampling errors can have an effect on correlation.

### 6.3.2 Uses of Correlation

Correlation can be used for varied purposes that have been discussed as follows (Mohanty and Misra, 2016).

1) **Validity and reliability:** Validity and reliability are important aspects of psychological testing and correlation can be used to obtain validity and

reliability of a psychological test. Validity is whether a test is measuring what it is supposed to measure and reliability provides information about consistency of a test.

2) **Verification of theory:** Correlation can also be used to verify or test certain theories by denoting whether relationship exists between the variables. For example, if a theory states that there is a relationship between parenting style and resilience, the same can be tested by computing correlation for the two variables.

3) **Putting variables in groups:** Variables that show positive correlation with each other can be grouped together and variables that show negative correlation can be grouped separately based on the coefficient of correlation obtained.

4) **Computation of further statistical analysis:** Based on the results obtained after computing correlation, various statistical techniques can be used like regression. Further, correlation is also used for multivariate statistical analysis, especially for techniques like Multivariate Analysis of Variance (MANOVA), Multivariate Analysis of Covariance (MANCOVA), Discriminant Analysis, Factor analysis and so on (Mohanty and Misra, 2016).

5) **Based on correlation, one can decide whether or not to determine prediction:** By computing correlation, it is not possible to predict one variable based on another variable, but based on the information that two or more variables are significantly related to each other, further statistical techniques can be used to make predictions. For example, if we obtain a positive correlation between family environment and adjustment of children, then further statistical techniques can be employed to find if adjustment of children can be predicted based on family environment.

### 6.3.3 Limitations of Correlation

Some of the limitations of correlation have been discussed as follows:

1) As was stated earlier, correlation will not provide any information about cause and effect relationship or causation.

2) The coefficient of correlation, mainly, Pearson's product moment correlation and Spearman's rank order correlation are suitable, when there is a linear relationship between the variables.

3) With regard to distributions that are discontinuous, the coefficient of correlation obtained may be overestimated or higher.

4) Sample variations can have an effect on correlation (as is also true with other statistical techniques).

5) In case of pooled sample, the correlation will be determined by relative position of the scores in X and Y dimensions or variables.

**Check Your Progress II**

1) List the uses of Correlation.

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

## 6.4 OTHER METHODS OF CORRELATION

There are various other methods of correlation as well, that will be discussed in the present section of this unit.

1) **Partial Correlation:** In partial correlation, the relationship between two variables is studied by controlling the influence of a third variable. For example, if we are studying the relationship between emotional intelligence and self concept of adolescents, we may partial out or control the third variable, for instance, family environment .

2) **Part Correlation:** This is also called as Semipartial Correlation. This is in a way similar to partial correlation, but here as the correlationship between two variables is studied, the influence of third variable on one of the variables is controlled. Taking the example discussed under partial correlation, the influence of family environment only on emotional intelligence is controlled and not on self concept.

3) **Multiple correlation:** In multiple correlation, one variable is correlated with many other variables. For example, self concept will be correlated with various other variables like emotional intelligence, achievement motivation, quality of life and so on.

4) **Biserial Correlation:** In biserial correlation, the relationship is measured between a continuous variable and an artificially dichotomous variable. A dichotomous variable is a variable that can be categorised into two. For example, Socio-Economic Status could be high and low. An example of biserial correlation would be relationship between achievement motivation and high and low emotional intelligence. Here, achievement motivation is a continuous variable and emotional intelligence is a variable that is artificially dichotomous.

5) **Point-Biserial Correlation:** In point biserial correlation, the relationship is measured between a continuous variable and a naturally dichotomous variable. Examples of naturally dichotomous variables are gender (male and female), religion (Hindu and Muslim) and so on. For example, point biserial correlation can be used when we want to find out relationship between work motivation (continuous variable) and gender (naturally dichotomous variable).

6) **Tetrachoric Correlation:** When both the variables are artificially dichotomous, then tetrachoric correlation can be computed to study the relationship between the two variables. For example, tetrachoric

correlation can be used when we want to study the relationship between variables, emotional intelligence (that is categorised in to high and low) and adjustment (that is categorised in to well adjusted and maladjusted).

7) **Phi coefficient:** This is similar to tetrachoric correlation, but is used when both the variables are naturally dichotomous. For example, if we want to find relationship between gender, that is categorised as male and female and response to a statement in terms of agree and disagree, then phi coefficient can be computed.

**Check Your Progress III**

Give a brief description and nature of variables 1 and 2 for other methods of correlation

| Method of Correlation | Description | Variable 1 | Variable 2 |
|---|---|---|---|
| **Partial Correlation** | | | |
| **Part Correlation** | | | |
| **Multiple Correlation** | | | |

| | | | |
|---|---|---|---|
| **Biserial Correlation** | | | |
| **Point-Biserial** | | | |
| **Tetrachoric Correlation** | | | |
| **Phi Coefficient** | | | |

## 6.5 LET US SUM UP

In the present unit, we mainly discussed about the concept of correlation. Correlation can be used to study relationship between two or more variables. It is a measure of association between two or more variables and this relationship is determined not only in terms of direction, whether negative or positive, but also in terms of its magnitude, whether high or low. Correlation can be categorised in to linear and nonlinear correlation and these were discussed in the present unit with the help of figures. The concept of scatter diagram was also briefly discussed in this unit. Scatter diagram is also called as scatter plot or scatter gram and is drawn by plotting the two variables in same graph, that is, variable A on y axis and variable B on x axis. Further in the unit, we also discussed about the properties, uses and limitations of correlation that are relevant, so as to know when exactly to use correlation. In the last section of this unit, we focused on the other methods of correlation including partial correlation, part correlation, multiple correlation, biserial correlation, point biserial correlation, tetrachoric correlation and phi- coefficient. In the next unit, we will learn how to compute coefficient of correlation with the help of Pearson's product moment correlation and Spearman's rank order correlation.

## 6.6 KEY WORDS

**Biserial Correlation:** In biserial correlation, the relationship is measured between a continuous variable and an artificially dichotomous variable.

**Correlation:** It is a measure of association between two or more variables and this relationship is determined not only in terms of direction, whether negative or positive.

**Linear Correlation:** Linear correlation is denoted by a single straight line in a graph that denotes linear relationship between given two variables.

**Multiple correlation:** In multiple correlation, one variable is correlated with many other variables.

**Nonlinear correlation:** Here the relationship is not denoted by a straight line but it is curvilinear.

**Negative Correlation:** The negative correlation denotes that increase in one variable leads to decrease in another variable or decrease in one variable leads to increase in another variable.

**No Correlation or Zero Correlation:** When there is no relationship between the two variables, the correlation will be zero. Thus in this case the relationship is neither positive or negative.

**Part Correlation:** This is also called as Semipartial Correlation. Here as the relationship between two variables is studied, the influence of third variable on one of the variables is controlled.

**Partial Correlation:** In partial correlation, the relationship between two variables is studied by controlling the influence of a third variable.

**Point-Biserial Correlation:** In point biserial correlation, the relationship is measured between a continuous variable and naturally dichotomous variable.

**Phi coefficient:** Phi coefficient is used when both the variables are naturally dichotomous.

**Positive correlation:** The positive correlation denotes that increase in one variable leads to increase in another variable and decrease in one variable leads to decrease in another variable.

**Scatter diagram:** Scatter diagram is drawn by plotting the two variables in same graph, that is variable A on y axis and variable B on x axis.

**Tetrachoric Correlation:** When both the variables are artificially dichotomous, then tetrachoric correlation can be computed to study the relationship between the two variables.

## 6.7 REFERENCES

King, Bruce. M; Minium, Edward. W. (2008). *Statistical Reasoning in the Behavioural Sciences*. Delhi: John Wiley and Sons, Ltd.

Mangal, S. K. (2002). *Statistics in Psychology and Education*. new Delhi: Phi Learning Private Limited.

Minium, E. W., King, B. M., & Bear, G. (2001). *Statistical Reasoning in Psychology and Education*. Singapore: John-Wiley.

Mohanty, B and Misra, S. (2016). *Statistics for Behavioural and Social Sciences*. Delhi: Sage.

Veeraraghavan, V and Shetgovekar, S. (2016). *Textbook of Parametric and Nonparametric Statistics*. Delhi: Sage.

## 6.8 ANSWERS TO CHECK YOUR PROGRESS

**Check Your Progress I**

1) What is Correlation?

It is a measure of association between two or more variables and this relationship is determined not only in terms of direction, whether negative or positive but also in terms of its magnitude, whether high or low.

**Check Your Progress II**

1) List the uses of Correlation

Correlation can be used

- to decide to whether or not to determine prediction:
- to obtain validity and reliability of psychological tests
- for verification of theory
- for putting variables in groups
- for computation of further statistical analysis.

**Check Your Progress III**

Give a brief description and nature of variables 1 and 2 for other methods of correlation

| Method of Correlation | Description | Variable 1 | Variable 2 |
|---|---|---|---|
| **Partial Correlation** | The correlationship between two variables is studied by controlling the influence of a third variable. | Variable is continuous in nature | Variable is continuous in nature |
| **Part Correlation** | The correlationship between two variables is studied, the influence of third variable on one of the variables is controlled | Variable is continuous in nature | Variable is continuous in nature |
| **Multiple Correlation** | One variable is correlated with many other variables (continuous). | Variable is continuous in nature | Variable is continuous in nature |
| **Biserial Correlation** | The correlationship is measured between a continuous variable and an artificially dichotomous variable. | Variable is continuous in nature | Variable is artificially dichotomous |
| **Point-Biserial** | The correlationship is measured between a continuous variable and naturally dichotomous variable | Variable is continuous in nature | Variable is naturally dichotomous |
| **Tetrachoric Correlation** | It is used when both the variables are artificially dichotomous | Variable is artificially dichotomous | Variable is artificially dichotomous |
| **Phi Coefficient** | It is used when both the variables are naturally dichotomous | Variable is naturally dichotomous | Variable is naturally dichotomous |

## 6.9 UNIT END QUESTIONS

1) Explain the concept of correlation with a focus on its direction and magnitude.
2) Discuss linear and nonlinear correlation with the help of diagrams.
3) Describe the properties of correlation.
4) Explain the uses and limitations of correlation.
5) Describe other methods of correlation.

# UNIT 7 COMPUTATION OF COEFFICIENT OF CORRELATION*



## 7.0 OBJECTIVES

After reading this unit, you will be able to :

- learn to compute coefficient of correlation with the help of Pearson's product moment coefficient of correlation and Spearman's rank order correlation.

## 7.1 INTRODUCTION

In the previous unit, we discussed about the basics of correlation. We discussed that correlation indicates relationship between two or more variables. This correlation can be interpreted in terms of direction and magnitude. Thus, a relationship between given two variables can be positive, negative or there could be no correlation. Further, the correlation may range between +1 to -1.

In the present unit, we will learn about the computation of correlation with the help of Pearson's product moment correlation and Spearman's rank order correlation.

One way in which these two methods can be distinguished is that, Pearson's product moment correlation can be categorised under parametric statistics and the Spearman's rank order correlation falls under nonparametric statistics.

To distinguish between parametric and nonparametric statistics, the following table (table 7.1) can be referred to:


* Prof. Suhas Shetgovekar, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi

| Table 7.1: Difference between Parametric and Nonparametric statistics | |
|---|---|
| **Parametric** | **Non-parametric** |
| The assumed distribution is normal. | The assumed distribution may not be normal. It can be any distribution. |
| The variance is homogeneous. | The variance could be heterogeneous or no assumption is made with regard to the variance. |
| The scales of measurement used are interval or ratio. | The scales of measurement used are nominal or ordinal. |
| The relationship between the data needs to be independent. | There is no assumption with regard to the independence of relationship between the data. |
| Mean is the measure of central tendency that is used here. | Median is the measure of central tendency that is used here. |
| It is more complex to compute when compared to the non parametric techniques. | It is simple to calculate. |
| Can get affected by outliers. | Is comparatively less affected by outliers. |

In the next section, we will learn how to compute Pearson's product moment correlation.

## 7.2 PEARSON'S PRODUCT MOMENT CORRELATION

Pearson's product moment correlation is one of the methods to compute coefficient of correlation. This is mainly used when the assumptions of parametric statistics are met. This method is named after Karl Pearson, who invented this method. It is denoted by 'r'.

### 7.2.1 Assumptions of Pearson's Product Moment Correlation

The assumptions of Pearson's product moment correlation are as follows:

1) The variables used to compute 'r' are continuous in nature and the scales of measurement are interval and ratio.

2) The distribution of the variables in this method is unimodal and it is close to symmetrical. The distribution need not be normal.

3) The pairs of scores involved are independent in nature and are in no way connected with other.

4) There is a linear relationship between the two variables. A scatter gram thus drawn with the help of scores in the two variables, will denote a straight line.

5) 'r' is mainly used to ascertain the sign and size of the correlation that can be positive, negative or zero correlation and will range between -1 to +1.

### 7.2.2 Uses of Pearson's Product Moment Correlation

1) It helps in determining the relationship between two variables quantitatively. With quantification, it is possible for us to compare.

2) Based on 'r', regression equation can be computed. Thus, after computing 'r', it is possible to compute regression and determine whether one variable can be predicted based on another variable.

3) 'r' can be used in computation of reliability and validity of psychological tests.

4) It will also assist in computation of factor analysis.

### 7.2.3 Computation of Pearson's Product Moment Correlation

There are two main methods that we will discuss for computing Pearson's product moment correlation. They are discussed as follows:

**Method 1:** The formula for the first method is give below,

$$\mathbf{r_{xy} = \Sigma xy / N\, \sigma_x \sigma_y}$$

Where,

r = Correlation

x = Deviation of any score of X from the mean of X

y = Deviation of any score of Y from the mean of Y

$\Sigma xy$= Indicates the sum of all the products of deviation (that is, each x deviation is multiplied by its corresponding y deviation)

$\sigma_x =$ Standard deviation of scores in X

$\sigma_y =$ Standard deviation of scores in Y

N = Total number of participants (frequencies)

The formula can be simplified as follows

$\sigma_x = \sqrt{\Sigma x^2 / N}$

$\sigma_y = \sqrt{\Sigma y^2 / N}$

Thus, by substituting the values for $\sigma_x$ and $\sigma_y$, the following is obtained :

$$r = \Sigma xy / N \sqrt{\Sigma x^2 / N} \sqrt{\Sigma y^2 / N}$$

$$= \Sigma xy / \sqrt{\Sigma x^2\, \Sigma y^2}$$

Let us understand this method and steps involved in it, with the help of an example,

A researcher wanted to study the relationship between data 1 (X) and data 2 (Y). The data is given below:

| Participants (1) | Data 1 (X) (2) | Data 2 (Y) (3) | x (4) | $y$ (5) | $xy$ (6) | $x^2$ (7) | $y^2$ (8) |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 4 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 3 | -1 | -1 | 1 | 1 | 1 |
| 3 | 4 | 5 | 1 | 1 | 1 | 1 | 1 |
| 4 | 4 | 5 | 1 | 1 | 1 | 1 | 1 |
| 5 | 3 | 5 | 0 | 1 | 0 | 0 | 1 |
| 6 | 2 | 3 | -1 | -1 | 1 | 1 | 1 |
| 7 | 2 | 3 | -1 | -1 | 1 | 1 | 1 |
| 8 | 3 | 4 | 0 | 0 | 0 | 0 | 0 |
| 9 | 5 | 5 | 2 | 1 | 2 | 4 | 1 |
| 10 | 2 | 3 | -1 | -1 | 1 | 1 | 1 |
| | $\Sigma X$= 30 | $\Sigma Y$= 40 | | | $\Sigma xy = 8$ | $\Sigma x^2 = 10$ | $\Sigma y^2 = 8$ |

**Step 1:** First the scores under X and and Y are totalled separately. Thus, ΣX and ΣY is obtained as can be seen above in the second and third column. N is also noted and in this case it is 10.

**Step 2:** Mean is now computed for the data 1 (X) and 2 (Y) as follows:

Mean for scores on X = 30 (30/10) = 3

Mean for scores on Y = 40 (40/10) = 4

**Step 3:** In the third step, deviation is computed for each score of X from its mean, that is, 3 in the case of this example. In a similar manner deviation is computed for each score of Y from its mean, that is, 4. These are entered in the column four and five above under headings 'x' and 'y' respectively.

**Step 4:** The values thus entered under 'x' and 'y' are multiplied and entered in column six and then they are also squared and entered under column seven and eight with headings $x^2$ and $y^2$. Further, the scores under each of these columns are totalled to obtain Σxy, $\Sigma x^2$ and $\Sigma y^2$.

**Step 5:** Use the formula to compute 'r'.

$$r = \Sigma xy / \sqrt{\Sigma x^2 \, \Sigma y^2}$$

$$= 8/ \sqrt{10 \times 8}$$

$$= 8/ \sqrt{80}$$

$$= 8/8.94$$

$$= 0.89$$

Thus, the coefficient of correlation obtained for the above data is 0.89, denoting that there is a positive and high relationship between the two data sets X and Y.

**Method 2:** The formula for the second method is give below,

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

Where,

X and Y= the raw scores for X and Y

ΣXY = The total of the products of each X score multiplied with its corresponding Y score

N= Total number of scores.

In this method, the deviations from the mean are not computed, instead raw scores are used to compute 'r'.

Let us understand this method and steps involved in it with the help of an example used earlier for calculating 'r'.

| Participants (1) | Data 1 (X) (2) | Data 2 (Y) (3) | XY (4) | $X^2$ (5) | $Y^2$ (6) |
|---|---|---|---|---|---|
| 1 | 3 | 4 | 12 | 9 | 16 |
| 2 | 2 | 3 | 6 | 4 | 9 |
| 3 | 4 | 5 | 20 | 16 | 25 |
| 4 | 4 | 5 | 20 | 16 | 25 |
| 5 | 3 | 5 | 15 | 9 | 25 |
| 6 | 2 | 3 | 6 | 4 | 9 |
| 7 | 2 | 3 | 6 | 4 | 9 |
| 8 | 3 | 4 | 12 | 9 | 16 |
| 9 | 5 | 5 | 25 | 25 | 25 |
| 10 | 2 | 3 | 6 | 4 | 9 |
| | **ΣX= 30** | **ΣY= 40** | **ΣXY= 128** | **$\Sigma X^2 = 100$** | **$\Sigma Y^2 = 168$** |

**Step 1:** Total scores for X and Y in column two and three are computed and denoted as ΣX and ΣY. In the case of present example, they are obtained as 30 and 40.

**Step 2:** In column four, XY is computed where the paired values under X and Y are multiplied. Thus, for participant 1, X value 3 and Y value 4 are multiplied and 12 is obtained. Similarly, XY is computed for all the participants and then ΣXY is also computed.

**Step 3:** In column five and six, $X^2$and$Y^2$ are computed. These are squared values of X and Y respectively. Further, $\Sigma X^2$and $\Sigma Y^2$ are computed that are summations of $X^2$and$Y^2$respectively.

**Step 4:** Use the formula to compute 'r'.

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

$$= 10 \times 128 - (30 \times 40) / \sqrt{[(10 \times 100 - (30)^2] [(10 \times 168 - (40)^2]}$$

$$= 1280 - (1200) / \sqrt{[1000 - 900] [1680 - 1600]}$$

$$= 80 / \sqrt{100 \times 80}$$

$$= 80 / \sqrt{8000}$$

$$= 80 / 89.44$$

$$= 0.89$$

Thus, 'r' obtained is 0.89 denoting a positive and high correlationship.

**Check Your Progress I**

1) Pearson Product Moment Correlation is denoted as .............................. .

2) The variables used to compute 'r' are continuous in nature and the scales of measurement are .............................. and .............................. .

3) The formula for the first method of computing Pearson's product moment correlation is .............................. .

## 7.3 SPEARMAN'S RANK ORDER CORRELATION

Another method to compute coefficient of correlation is Spearman's rank order correlation. This method is used when the assumptions of parametric statistics are not met. The method is named after Charles Spearman, who is known for his significant work on factor analysis and theory of intelligence besides Spearman's rank order correlation.

### 7.3.1 Assumptions for Spearman's Rank Order Correlation

The assumptions of Spearman's rank order correlation are as follows:

1) The variables are measured in terms of ordinal scale.

2) The relationship between the two variables is linear in nature.

3) The observations are independent in nature, thus denoting that the sample needs to be randomly selected.

4) The pairs of scores are independent in nature and are in no way connected with other pairs.

### 7.3.2 Uses of Spearman's Rank Order Correlation

1) It is used when the data is measured with the help of ordinal scale.

2) It is especially useful when the sample size is small, that is, less than 25-30 (Mohanty and Misra, 2016).

3) Many a times it is not possible to measure traits directly. Thus, they are measured in terms of ranks. Spearman's rank order correlation involves separately ranking the scores in the two data, followed by computation of correlationship between them.

4) It can be used to study the degree of relationship between two variables that are monotonic. A relationship is termed as monotonic when the variables display consistent but one directional relationship.

### 7.3.3 Computation of Spearman's Rank Correlation

There are two main methods that we will discuss for computing Spearman's rank order correlation, one without tied ranks and one with tied ranks. There are discussed as follows:

**Method 1 (without tied ranks):** The formula for the first method is give below,

$$p = 1 - [(6\Sigma d^2) / [N (N^2 - 1)]$$

where,

$\Sigma d^2$= Sum of the difference squared

N = Total number of participants

Let us understand this method and steps involved in it, with the help of an example,

A researcher wanted to study the relationship between data 1 (X) and data 2 (Y). The data is given below:

| **Participants (c)** | **Data 1(X) (2)** | **Data 2 (Y) (3)** | **Rank for Data 1 ($R_1$) (4)** | **Rank for Data 2 ($R_2$) (5)** | **Difference in Ranks ($R_1 - R_2 = \|d\|$) (6)** | **Difference Squared ($d^2$) (7)** |
|---|---|---|---|---|---|---|
| 1 | 45 | 40 | 9 | 9 | 0 | 0 |
| 2 | 34 | 33 | 8 | 8 | 0 | 0 |
| 3 | 23 | 25 | 3 | 3 | 0 | 0 |
| 4 | 22 | 21 | 2 | 2 | 0 | 0 |
| 5 | 65 | 60 | 10 | 10 | 0 | 0 |
| 6 | 33 | 30 | 7 | 5 | 2 | 4 |
| 7 | 30 | 31 | 5 | 6 | 1 | 1 |
| 8 | 25 | 32 | 4 | 7 | 3 | 9 |
| 9 | 32 | 29 | 6 | 4 | 2 | 4 |
| 10 | 21 | 20 | 1 | 1 | 0 | 0 |
| N= 10 | | | | | | **$\sum d^2$= 18** |

**Step 1:** Ranks are assigned separately to scores under data 1 and those under data 2. These ranks are mentioned in column four and five respectively. Ranks can either be assigned in descending or ascending order. For instance in the present example, rank 1 is assigned to the lowest value and rank 10 to the highest value and this is followed in same way for both the data.

**Step 2:** Difference in Ranks are calculation and these are irrespective of their signs($R_1 - R_2 = |d|$). These are then mentioned in column six. In the last column, that is, column seven, difference squared($d^2$) is computed and the total of this is mentioned as $\sum d^2$. In the case of present example, $\sum d^2$ is 18.

**Step 3:**The formula used to compute Rho is

$$\boldsymbol{p = 1- [(6\Sigma d^2) / [N (N^2 - 1)]}$$

$$= 1- (6 \times 18) / 10 (10^2-1)$$

$$= 1- (108)/ 10 (100-1)$$

$$= 1- (108/ 10 \times 99)$$

$$= 1- (108/ 990)$$

$$= 1- 0.11$$

$$= 0.89$$

Thus, the correlation of coefficient (Rho) obtained for the above data is 0.89, thus denoting a significant and positive correlation between data 1 and data 2.

**Method 2 (with tied ranks):** The formula used for computing rho with tied ranks is the same. Only we need to understand how ranks are assigned when there are two or more similar scores in a given data.

Let us understand this method and steps involved in it, with the help of an example,

A researcher wanted to study the relationship between data 1 (X) and data 2 (Y). The data obtained is given below:

| **Participants (c)** | **Data 1 (X) (2)** | **Data 2 (Y) (3)** | **Rank for Data 1 ($R_1$) (4)** | **Rank for Data 2 ($R_2$) (5)** | **Difference in Ranks ($R_1$- $R_2$= \|d\|) (6)** | **Difference Squared ($d^2$) (7)** |
|---|---|---|---|---|---|---|
| 1 | 45 | 40 | 8 | 7 | 1 | 1 |
| 2 | 23@ | 33 | 2.5 | 6 | 3.5 | 12.25 |
| 3 | 23@ | 25 | 2.5 | 3 | 0.5 | 0.25 |
| 4 | 64 | 21 | 9 | 2 | 7 | 49 |
| 5 | 65 | 60# | 10 | 9 | 1 | 1 |
| 6 | 33 | 60# | 7 | 9 | 2 | 4 |
| 7 | 25# | 31 | 4.5 | 5 | 0.5 | 0.25 |
| 8 | 25# | 60# | 4.5 | 9 | 4.5 | 20.25 |
| 9 | 32 | 29 | 6 | 4 | 2 | 4 |
| 10 | 21 | 20 | 1 | 1 | 0 | 0 |
| N= 10 | | | | | | **$\sum d^2$= 92** |

As can be seen in the above table, there are same values under data 1, that is 23, obtained by participant 2 and 3 and score of 25 obtained by participants 7 and 8. Similarly, in data 2, participants 5, 6 and 8 have obtained 60 score. In such a case ranks are assigned in a bit different manner.

As can be seen in above table, 21 is assigned with rank 1 and then there are two '23' scores that need to be equally assigned ranks 2 and 3. Thus 2 +3 = 5/ 2= 2.5. The rank 2.5 is then allotted to both these score. The next score is then allotted rank 4. But in present example, rank 4 and 5 are shared equally by score 25 obtained by 7th and 8th participants. Thus 4+5= 9/ 2 (because there are two same scores) = 4.5. Thus, 4.5 is allotted to these two scores and the next score, that is, 32 is assigned rank 6.

In data 2, the score 60 equally shares ranks 8, 9 and 10. Thus 8 + 9 +10 = 27/ 3 (because there are three same scores) = 9. Thus the score 60 is assigned rank 9.

Using the same formula Rho is computed as follows

$$p = 1- [(6\Sigma d^2) / [N (N^2 - 1)]$$

$$= 1- (6 \times 92) / 10 (10^2-1)$$

$$= 1- (552)/ 10 (100-1)$$

$$= 1- (552/ 10 \times 99)$$

$$= 1- (552/ 990)$$

$$= 1- 0.56$$

$$= 0.44$$

Thus, the correlation of coefficient (Rho) obtained for the above data is 0.44, thus denoting a positive and high correlationship between the two data sets.

**Check Your Progress II**

1) The variables in Spearman's Rho are measured in terms of ................. scale.

2) A relationship is termed as monotonic when

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

3) The formula for the first method of computing Spearman's rho is

## 7.4 LET US SUM UP

In the present unit, we mainly discussed about the two methods of computing coefficient of correlation. The first method is Pearson's product moment correlation and the other is Spearman's rank order correlation. Pearson's product moment correlation is one of the methods to compute coefficient of correlation. This is mainly used when the assumptions of parametric statistics are met. This method is named after Karl Pearson, who invented this method. It is denoted by 'r'. Spearman's rank order correlation is used when the assumptions of parametric statistics are not met. The method is named after Charles Spearman, who is known for his significant work on factor analysis and theory of intelligence. The assumptions and uses of these method were also discussed. The formula and computation for the two methods were discussed with the help of examples.

## 7.5 REFERENCES

Mangal, S. K. (2002). *Statistics in Psychology and Education*. New Delhi: Phi Learning Private Limited.

Mohanty, B and Misra, S. (2016). *Statistics for Behavioural and Social Sciences*. Delhi: Sage.

Veeraraghavan, V and Shetgovekar, S. (2016). *Textbook of Parametric and Nonparametric Statistics*. Delhi: Sage.

## 7.6 ANSWERS TO CHECK YOUR PROGRESS

**Check Your Progress I**

1) Pearson Product Moment Correlation is denoted as r.

2) The variables used to copute r are continuous in nature and the scales of measurement are interval and ratio.

3) The formula for the first method of computing Pearson's Product Moment Correlation is $\mathbf{r_{xy} = \Sigma xy / N\, \sigma_x \sigma_y}$

**Check Your Progress II**

1) The variables in Spearman's Rho are measured in terms of Ordinal scale.

2) A relationship is termed as monotonic when the variables display consistent but one directional relationship.

3) The formula for the first method of computing Spearman's rho is $p = 1 - [(6\Sigma d^2) / [N (N^2 - 1)]$

## 7.7 UNIT END QUESTIONS

1) Differentiate between Parametric and Nonparameric Statistics.

2) Discuss the assumptions of Pearson's product moment correlation

3) Describe the uses of Pearson's product moment correlation.

4) Discuss the assumptions of Spearman's rank order correlation.

5) Describe the steps involved in computation of Spearman's rho with the help of an example.

146 blank