
UNIT 3 PARTIAL AND MULTIPLE CORRELATIONS

Structure

- 3.0 Introduction
 - 3.1 Objectives
 - 3.2 Partial Correlation (r_p)
 - 3.2.1 Formula and Example
 - 3.2.2 Alternative Use of Partial Correlation
 - 3.3 Linear Regression
 - 3.4 Part Correlation (Semipartial correlation) r_{sp}
 - 3.4.1 Semipartial Correlation: Alternative Understanding
 - 3.5 Multiple Correlation Coefficient (R)
 - 3.6 Let Us Sum Up
 - 3.7 Unit End Questions
 - 3.8 Suggested Readings
-

3.0 INTRODUCTION

While learning about correlation, we understood that it indicates relationship between two variables. Indeed, there are correlation coefficients that involve more than two variables. It sounds unusual and you would wonder how to do it? Under what circumstance it can be done? Let me give you two examples. The first is about the correlation between cholesterol level and bank balance for adults. Let us say that we find a positive correlation between these two factors. That is, as the bank balance increases, cholesterol level also increases. But this is not a correct relationship as Cholesterol level can also increase as age increases. Also as age increases, the bank balance may also increase because a person can save from his salary over the years. Thus there is age factor which influences both cholesterol level and bank balance. Suppose we want to know only the correlation between cholesterol and bank balance without the age influence, we could take persons from the same age group and thus control age, but if this is not possible we can statistically control the age factor and thus remove its influence on both cholesterol and bank balance. This if done is called partial correlation. That is, we can use partial and part correlation for doing the same. Sometimes in psychology we have certain factors which are influenced by large number of variables. For instance academic achievement will be affected by intelligence, work habit, extra coaching, socio economic status, etc. To find out the correlation between academic achievement with various other factors ad mentioned above can be done by Multiple Correlation. In this unit we will be learning about partial, part and multiple correlation.

3.1 OBJECTIVES

After completing this unit, you will be able to:

- Describe and explain concept of partial correlation;

- Explain, the difference between partial and semipartial correlation;
- Describe and explain concept of multiple correlation;
- Compute and interpret partial and semipartial correlations;
- Test the significance and apply the correlation to the real data;
- Compute and interpret multiple correlation; and
- Apply the correlation techniques to the real data.

3.2 PARTIAL CORRELATION (r_p)

Two variables, A and B, are closely related. The correlation between them is partialled out, or controlled for the influence of one or more variables is called as partial correlation. So when it is assumed that some other variable is influencing the correlation between A and B, then the influence of this variable(s) is partialled out for both A and B. Hence it can be considered as a correlation between two sets of residuals. Here we discuss a simple case of correlation between A and B is partialled out for C. This can be represented as $r_{AB,C}$ which is read as correlation between A and B partialled out for C. the correlation between A and B can be partialled out for more variables as well.

3.2.1 Formula and Example

For example, the researcher is interested in computing the correlation between anxiety and academic achievement controlled from intelligence. Then correlation between academic achievement (A) and anxiety (B) will be controlled for Intelligence (C).

This can be represented as: $r_{\text{Academic Achievement(A) Anxiety (B). Intelligence (C)}}$. To calculate the partial correlation (r_p) we will need a data on all three variables. The computational formula is as follows:

$$r_p = r_{AB,C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1 - r_{AC}^2)(1 - r_{BC}^2)}} \quad (\text{eq. 3.1})$$

Look at the data of academic achievement, anxiety and intelligence. Here, the academic achievement test, the anxiety scale and intelligence test is administered on ten students. The data for ten students is provided for the three variables in the table below.

Table 3.1: Data of academic achievement, anxiety and intelligence for 10 subjects

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

In order to compute the partial correlation between the academic achievement and anxiety partialled out for Intelligence, we first need to compute the Pearson's Product moment correlation coefficient between all three variables. We have already learned to compute it in the first Unit of this Block. So I do not again explain it here.

The correlation between anxiety (B) and academic achievement (A) is -0.369.

The correlation between intelligence (C) and academic achievement (A) is 0.918.

The correlation between anxiety (B) and intelligence (C) is -0.245.

Given the correlations, we can now calculate the partial correlation .

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{-0.369 - (0.918 \times -0.245)}{\sqrt{(1-0.918^2)(1-(-0.245)^2)}} = \frac{-0.369 - (-0.222)}{\sqrt{0.081 \times 0.939}} = \frac{-0.1441}{0.499} = -0.375 \quad (\text{eq.3.2})$$

The partial correlation between the two variables, academic achievement and anxiety controlled for intelligence, is -0.375. You will realise that the correlation between academic achievement and anxiety is -0.369. Whereas, after partialling out for the effect of intelligence, the correlation between them has almost remained unchanged. While computing this correlation, the effect of intelligence on both the variables, academic achievement and anxiety, was removed.

The following figure explains the relationship between them.

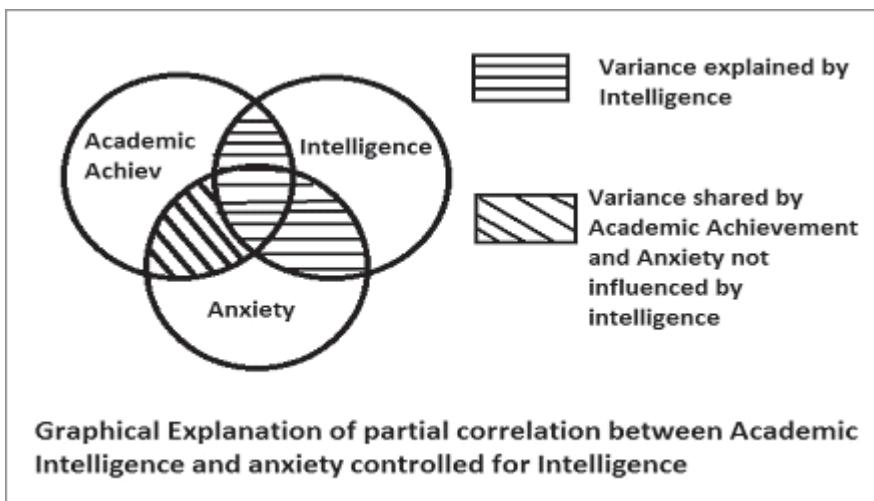


Fig. 3.1: Venn diagram explaining the partial correlation

Significance testing of the partial correlation

We can test the significance of the partial correlation for the null hypothesis

$$H_0 : \tilde{r}_p = 0$$

and the alternative hypothesis

$$H_a : \tilde{r}_p \neq 0$$

Where, the \tilde{r}_p denote the population partial correlation coefficient. The t -distribution is used for this purpose. Following formula is used to calculate the t -value.

$$t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}} \quad (\text{eq. 3.3})$$

Where,

r_p = partial correlation computed on sample, $r_{AB.C}$

n = sample size,

v = total number of variables employed in the analysis.

The significance of the r_p is tested at the $df = n - v$.

In the present example, we can employ significance testing as follows:

$$t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}} = \frac{-0.375 \sqrt{10-3}}{\sqrt{1-(-0.375)^2}} = \frac{-0.992}{0.927} = 1.69$$

We test the significance of this value at the $df = 7$ in the table for t-distribution in the appendix. You will realise that at the $df = 7$, the table provides the critical value of 2.36 at 0.05 level of significance. The obtained value of 1.69 is smaller than this value. So we accept the null hypothesis stating that $H_0 : \tilde{n}_p = 0$.

Large sample example:

Now we take a relatively large sample example. A counseling psychologist is interested in understanding the relationship between practice of study skills and marks obtained. But she is skeptical about the effectiveness of the study skills. She believes that they can be effective because they are good cognitive techniques or they can be effective simply because the subjects believe that the study skills are going to help them. The first is attribute of the skills while second is placebo effect. She wanted to test this hypothesis. So, along with measuring the hours spent in practicing the study skills and marks obtained, she also took measures on belief that study skill training is useful. She collected the data on 100 students. The obtained correlations are as follows.

The correlation between practice of study skills (A) and unit test marks (B) is 0.69

The correlation between practice of study skills (A) and belief about usefulness of study skills (C) is 0.46

The correlation between marks in unit test (B) and belief about usefulness of study skills (C) is 0.39

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{0.69 - (0.46 \times 0.39)}{\sqrt{(1-0.46^2)(1-0.39^2)}} = \frac{.51}{0.82} = 0.625$$

The partial correlation between practice of study skills (A) and unit test marks (B) is 0.625. Let's test the null hypothesis about the partial correlation for a null hypothesis which states that $H_0 : \tilde{n}_p = 0$.

$$t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}} = \frac{0.625 \sqrt{100-3}}{\sqrt{1-0.625^2}} = \frac{1.65}{0.781} = 2.12$$

The t value is significant at 0.05 level. So we reject the null hypothesis and accept that there is a partial correlation between A and B. This means that the partial correlation between practice of study skills (A) and unit test marks (B) is non-zero at population. We can conclude that the correlation between practice of study skills (A) and unit test marks (B) still exists even after controlled for the belief in the usefulness of the study skills. So the skepticism of our researcher is unwarranted.

3.2.2 Alternative Use of Partial Correlation

Suppose you have one variable which is dichotomous. These variables take two values. Some examples are, male and female, experimental and control group, patients and normal, Indians and Americans, etc. Now these two groups were measured on two variables, X and Y. You want to correlate these two variables. But you are also interested in testing whether these groups influence the correlation between the two variables. This can be done by using partial correlations. Look at the following data. This data is for male and female subjects on two variables, neuroticism and intolerance to ambiguity.

Table 3.2: Table showing gender wise data for IOA and N.

Male		Female	
IOA	N	IOA	N
12	22	27	20
17	28	25	15
7	24	20	18
12	32	19	12
14	30	26	18
11	27	23	13
13	29	24	20
10	17	22	9
21	34	21	19

If you compute the correlation between Intolerance of Ambiguity and neuroticism for the entire sample of male and female for 20 subjects. It is -0.462 . This is against the expectation.

This is a surprising finding which states that as the neuroticism increases the intolerance to ambiguous situations decreases. What might be the reason for such correlation? If we examine the mean of these two variables across gender, then you will realise that the trend of mean is reversed.

If you calculate the Pearson's correlations separately for each gender, then they are well in the expected line (0.64 for males and 0.41 for females).

The partial correlations can help us in order to solve this problem. Here, we calculate the Pearson's product moment correlation between IOA and N partialled out for sex. This will be the correlation between neuroticism and intolerance of ambiguity from which the influence of sex is removed.

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{-0.462 - (0.837 \times -0.782)}{\sqrt{(1-0.837^2)(1-(-0.782)^2)}} = \frac{.193}{0.341} = 0.566$$

The correlation partialled out for sex is 0.57. Let's test the significance of this correlation.

$$t = \frac{r_p \sqrt{n-v}}{\sqrt{1-r_p^2}} = \frac{.566 \sqrt{18-3}}{\sqrt{1-.566^2}} = \frac{2.194}{0.824} = 2.66$$

The tabled value form the appendix at $df=15$ for 0.05 level is 2.13 and for 0.01 level is 2.95. The obtained t-value is significant at 0.05 level. So we reject the null

hypothesis which stated that population partial correlation, between IOA and N partialled out for sex is zero.

Partial correlation as Pearson's Correlation between Errors

Partial Correlation can also be understood as a Pearson's correlation between two errors.

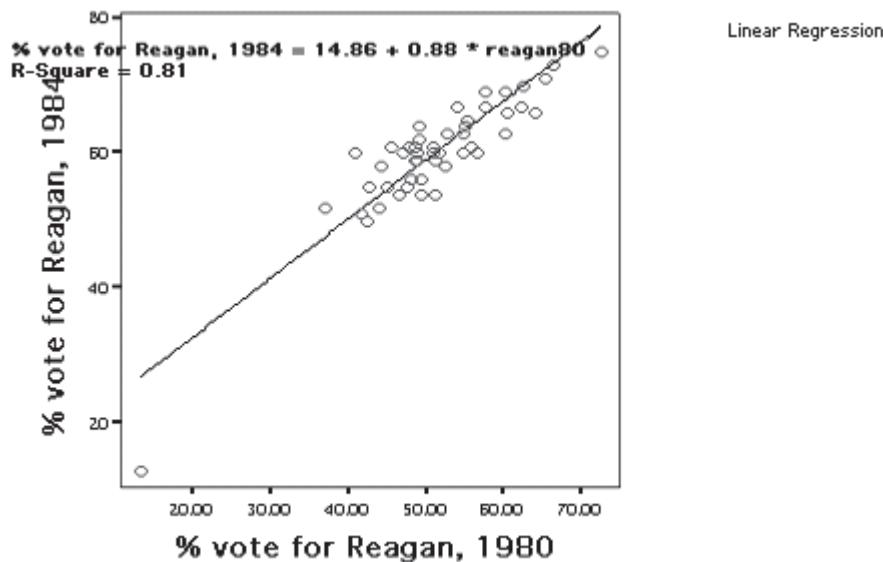
Before you proceed you need to know what is regression equation

3.3 LINEAR REGRESSION

Regression goes one step beyond correlation in identifying the relationship between two variables. It creates an equation so that values can be predicted within the range framed by the data. That is if you know X you can predict Y and if you know Y you can predict X. This is done by an equation called regression equation.

When we have a scatter plot you have learnt that the correlation between X and Y are scattered in the graph and we can draw a straight line covering the entire data. This line is called the regression line.

Here is the line and the regression equation superimposed on the scatterplot:



Source: <http://janda.org/c10/Lectures/topic04/L25-Modeling.htm>

From this line, you can predict X from Y that is % votes in 1984 if known, you can find out the % of votes in 1980. Similarly if you know % of votes in 1980 you can know % of votes in 1984.

The regression line seen in the above diagram is close to the scatterplots. That is the predicted values need to be as close as possible to the data. Such a line is called the best fitting line or Regression line. There are certain guidelines for regression lines:

- 1) Use regression lines when there is a significant correlation to predict values.
- 2) Do not use if there is not a significant correlation.
- 3) Stay within the range of the data. For example, if the data is from 10 to 60, do not predict a value for 400.

- 4) Do not make predictions for a population based on another population's regression line.

The y variable is often termed the **criterion variable** and the x variable the **predictor variable**. The slope is often called the **regression coefficient** and the intercept the **regression constant**. The slope can also be expressed compactly as $\beta_1 = r \times s_y/s_x$.

Normally we then predict values for y based on values of x . This still does not mean that y is caused by x . It is still imperative for the researcher to understand the variables under study and the context they operate under before making such an interpretation. Of course, simple algebra also allows one to calculate x values for a given value of y .

To obtain regression equation we use the following equation:

$$\beta = \{N * \sum xy\} - \{\sum y^2 * \sum y\} / \{(N * \sum x^2) - (\sum y^2)\}$$

Regression equation can also be written including error component ' \hat{a} '

The regression equation can be written as

$$Y = \alpha + \beta X + \varepsilon \quad (\text{eq. 4.8})$$

Where,

Y = dependent variable or criterion variable

α = the population parameter for the y -intercept of the regression line, or regression coefficient ($r = \bar{y}/\bar{x}$)

β = population slope of the regression line or regression coefficient ($r * \bar{x}/\bar{y}$)

ε = the error in the equation or residual

The value of α and β are not known, since they are values at the level of population. The population level value is called the parameter. It is virtually impossible to calculate parameter. So we have to estimate it. The two parameters estimated are α and β . The estimator of the α is ' a ' and the estimator for β is ' b '. So at the sample level equation can be written as

$$Y = a + bX + e \quad (\text{eq. 4.9})$$

Where,

Y = the scores on Y variable

X = scores on X variable

a = the Y -intercept of the regression line for the sample or regression constant in sample

b = the slope of the regression line or regression coefficient in sample

e = error in prediction of the scores on Y variable, or residual

Let us take an example and demonstrate

Example: Write the regression line for the following points:

x	y
1	4
3	2
4	1
5	0
8	0

Solution 1: $\sum x = 21$; $\sum y = 7$; $\sum x^2 = 115$; $\sum y^2 = 21$; $\sum xy = 14$

Thus $\beta_0 = [7*115 - 21*14] \div [5 * 115 - 21^2] = 511 \div 134 = 3.81$ and $\beta_1 = [5*14 - 21*7] \div [5 * 115 - 21^2] = -77 \div 134 = -0.575$.

Thus the regression equation for this example is $y = -0.575x + 3.81$.

Thus if you have x , then you can find or predict y .

If you have y you can predict x .

Let's continue with the first example.

It was relationship between anxiety and academic achievement. This relationship was controlled for (partialled out for) intelligence.

In this case we can write two linear regression equations and solve them by using ordinary least-squares (OLS). They are as follows:

Academic Achievement = $a_1 + b_1 \times \text{Intelligence} + e_1$

Where, ' a_1 ' is a y intercept of the regression line;

' b_1 ' is the slope of the line;

' e_1 ' is the error in the prediction of academic achievement using intelligence.

Anxiety = $a_2 + b_2 \times \text{Intelligence} + e_2$

Where, ' a_2 ' is a y intercept of the regression line;

' b_2 ' is the slope of the line;

' e_2 ' is the error in the prediction of academic achievement using intelligence.

Now we have e_1 and e_2 . They are residuals of each of the variables after intelligence explain variation in them. Meaning, e_1 is the remaining variance in academic achievement once the variance accounted for intelligence is removed. Similarly, e_2 is the variance left in the anxiety once the variance accounted for the intelligence is removed.

Now, the partial correlation can be defined as the Pearson's correlation between e_1 and e_2 .

$$r_p = e_1 e_2 \quad (\text{eq. 3.4})$$

You will realise that this correlation is the correlation of academic achievement and anxiety, from which a linear influence of intelligence has been removed. That is called as partial correlation.

3.4 PART CORRELATION (SEMIPARTIAL CORRELATION) r_{sp}

The Part correlation is also known as semi-partial correlation (r_{sp}). Semipartial correlation or part correlation are correlation between two variables, one of which is partialled for a third variable.

In partial correlations ($r_p = r_{AB,C}$) the effect of the third variable (C) is partialled out from BOTH the variables (A and B).

In semipartial correlations ($r_{sp} = r_{A(B,C)}$), as the name suggests, the effect of third variable (C) was partialled out from only one variable (B) and NOT from both the variables.

Let's continue with the earlier example. The example was about the correlation between anxiety (A) and academic achievement (B).

In the earlier example of partial correlation, we have partialled the effect of intelligence (C) from both academic achievement and anxiety.

One may argue that the academic achievement is the only variable that relates to intelligence.

So we need to partial out the effect of the intelligence only from academic achievement and not from anxiety.

Now, we correlate anxiety (A) as one variable and academic achievement partialled for intelligence (B.C) as another variable.

If we correlate these two then, the correlation of anxiety (A) with academic achievement partialled for intelligence (B.C) is called as semipartial correlation ($r_{A(B,C)}$).

In fact, if there are three variables, then total six semipartial correlations can be computed. They are $r_{A(B,C)}$, $r_{A(C,B)}$, $r_{B(A,C)}$, $r_{B(C,A)}$, $r_{C(A,B)}$, and $r_{C(B,A)}$.

Formula:

In order to compute the semipartial correlation coefficient, following formula can be used.

$$r_{sp} = r_{A(B,C)} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{1 - r_{BC}^2}} \quad (\text{eq. 3.5})$$

Where,

$r_{A(B,C)}$ is a semipartial correlation of A with the B after linear relationship that C has with B is removed

r_{AB} Pearson's product moment correlation between A and B

r_{AC} Pearson's product moment correlation between A and C

r_{BC} Pearson's product moment correlation between B and C

Example:

Let's take the data from the earlier example of academic achievement, anxiety and intelligence. The data table 3.1 is as follows.

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

The correlation between anxiety (A) and academic achievement (B) is -0.369 .

The correlation between intelligence (C) and academic achievement (B) is 0.918 .

The correlation between anxiety (A) and intelligence (C) is -0.245 .

Given the correlations, we can now calculate the semipartial correlation (r_{SP}) as follows. We are not computing the correlation coefficients, simply because you have already learned to compute the correlations earlier. The formula for semipartial correlation is as follows:

$$r_{SP} = r_{A(B.C)} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{1 - r_{BC}^2}} \quad (\text{eq. 3.6})$$

Now we can calculate semipartial correlation by using this formula.

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{1 - r_{BC}^2}} = \frac{-0.369 - (-0.245 \times 0.918)}{\sqrt{1 - (0.918^2)}} = \frac{-0.1441}{0.3966} = -0.363$$

The semipartial correlation between anxiety and academic achievement after the linear relationship between the academic achievement and intelligence is removed is -0.363 .

The significance of the semipartial correlation can be tested by using t-distribution. The null hypothesis and the alternate hypothesis are as follows.

$$H_0: \tilde{r}_{SP} = 0$$

$$H_A: \tilde{r}_{SP} \neq 0$$

Where, the \tilde{r}_{SP} is the semipartial correlation in the population. We test the null hypothesis whether the semipartial correlation in the population is zero. This can be done by using following formula

$$t = \frac{r_{SP}\sqrt{n-\nu}}{\sqrt{1 - r_{SP}^2}} \quad (\text{eq. 3.7})$$

Where,

t = students t -value

r_{SP} = semipartial correlation computed on sample,

n = sample size,

v = number of variables used in the analysis

The significance of this t-value is tested at the $df = n - v$. when three variables are involved then the df is $n - 3$.

For our example, the t-values can be computed as follows:

$$t = \frac{-0.363\sqrt{10-3}}{\sqrt{1-(-0.363^2)}} = -1.032$$

The obtained t-value is tested at $df = n - v = 10 - 3 = 7$.

The t-value at .05 level is 2.364. The obtained t-value is smaller than that. So we accept the null hypothesis that the population semipartial correlation is zero.

It has an interesting implication for our data. The correlation between anxiety and academic achievement is zero in the population if the linear relationship between academic achievement and intelligence is removed.

3.4.1 Semipartial Correlation: Alternative Understanding

Partial Correlation can also be understood as a Pearson's correlation between a variable and error (residual).

Let us continue with the first example. It was relationship between anxiety and academic achievement. This relationship was controlled for (partialled out for) intelligence and academic achievement. (So far you have not learned regression and you may not follow some of the points and equations. So you can revisit this discussion after learning regression.)

In this case we can write a linear regression equation and solve them by using ordinary least-squares (OLS). They are as follows:

$$\text{Academic Achievement} = a_1 + b_1 \times \text{Intelligence} + e_1$$

Where, ' a_1 ' is a y intercept of the regression line;

' b_1 ' is the slope of the line;

' e_1 ' is the error in the prediction of academic achievement using intelligence.

Now we have e_1 . It is a residuals of academic achievement after intelligence explain variation in academic achievement.

That is, e_1 is the remaining variance in academic achievement once the variance accounted for intelligence is removed.

Now, the semipartial correlation can be defined as the Pearson's correlation between anxiety and e_1 .

You will realise that this correlation is the correlation of academic achievement and anxiety, from which a linear influence of intelligence on academic achievement has been removed.

That is called the semipartial correlation.

(Since you have not learned the regression equation you may not be able to understand this point. So revisit this point after learning regression.)

3.5 MULTIPLE CORRELATION COEFFICIENT (R)

The multiple correlation coefficient denoting a correlation of one variable with multiple other variables. The multiple correlation coefficient is denoted as $R_{A,BCD\dots k}$ which denotes that A is correlated with B, C, D, up to k variables.

For example, we want to compute multiple correlation between A with B and C then it is expressed as $R_{A,BC}$. In this case we create a linear combination of the B and C which is correlated with A.

We continue with the same example which we have discussed for partial and semipartial correlations. This example has academic achievement, anxiety and intelligence as three variables. The correlation between academic achievement with the linear combination of anxiety and intelligence is multiple correlation.

This denotes the proportion of variance in academic achievement explained by intelligence and anxiety. We denote this as

$R_{(Academic\ Achievement, Intelligence, Anxiety)}$, which is a multiple correlation.

Often, it is used in the context of regression, where academic achievement is a criterion variable and intelligence and anxiety are called as predictors.

We are not using regression equation since you have not learned it. The Multiple R can be calculated for two predictor variable as follows.

$$R_{A,BC} = \sqrt{\frac{r_{AB}^2 + r_{AC}^2 - 2r_{AB}r_{AC}r_{BC}}{1 - r_{BC}^2}} \quad (\text{eq. 3.7})$$

Where,

$R_{A,BC}$ = is multiple correlation between A and linear combination of B and C.

r_{AB} = is correlation between A and B

r_{AC} = is correlation between A and C

r_{BC} = is correlation between B and C

Example

We shall continue with the earlier data.

The data table 3.1 is as follows.

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

The correlation between anxiety (A) and academic achievement (B) is -0.369.

The correlation between intelligence (C) and academic achievement (B) is 0.918.

The correlation between anxiety (A) and intelligence (C) is -0.245.

The multiple correlation can be calculated as follows.

$$\begin{aligned}
 R_{A,BC} &= \sqrt{\frac{r_{AB}^2 + r_{AC}^2 - 2r_{AB}r_{AC}r_{BC}}{1 - r_{BC}^2}} \\
 &= \sqrt{\frac{-0.369^2 + 0.918^2 - 2 \times -0.369 \times 0.918 \times -0.245}{1 - (-0.245)^2}} \\
 &= \sqrt{\frac{0.813}{0.94}} \\
 &= 0.929
 \end{aligned}$$

This means that the multiple correlation between academic achievement and the linear combination of intelligence and anxiety is 0.929 or 0.93. We have earlier learned that the square of the correlation coefficient can be understood as percentage of variance explained.

The R^2 is then percentage of variance in academic achievement explained by the linear combination of intelligence and anxiety. In this example the R^2 is 0.929² which is 0.865. The linear combination of intelligence and anxiety explain 86.5 percent variance in the academic achievement.

We have already converted the R into the R^2 value. The R^2 is the value obtained on a sample. The population value of the R^2 is denoted as P^2 . The R^2 is an estimator of the P^2 .

But there is a problem in estimating the P^2 value from the R^2 value.

The R^2 is not an unbiased estimator of the P^2 .

So we need to adjust the value of the R^2 in order to make it unbiased estimator.

Following formula is used for this purpose.

Let \tilde{R}^2 denote an adjusted R^2 . Then \tilde{R}^2 can be computed as follows:

$$\tilde{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \quad (\text{eq. 3.8})$$

Where,

\tilde{R}^2 = adjusted value of R^2

k = number of predicted variables (or the variable for which a linear combination is created)

n = sample size

For our example the \tilde{R}^2 value need to be computed.

$$\tilde{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

$$\tilde{R}^2 = 1 - \frac{(1-0.865)(10-1)}{10-2-1}$$

$$\tilde{R}^2 = 1 - \frac{1.217}{7} = 0.826$$

So the unbiased estimator of the R^2 the adjusted value, \tilde{R}^2 , is 0.826 which is smaller than the value of R^2 . It is usual to get a smaller adjusted value.

The significance testing of the R^2 :

This can be used for the purpose of the significance testing. The null hypothesis and the alternative hypothesis employed for this purpose are

$$H_0 : P^2 = 0$$

$$H_A : P^2 \neq 0$$

The null hypothesis denotes that the population R^2 is zero whereas the alternative hypothesis denotes that the population R^2 is not zero.

The F -distribution is used for calculating the significance of the R^2 as follows:

$$F = \frac{(n-k-1)R^2}{k(1-R^2)} \quad (\text{eq. 3.9})$$

The value of R^2 can be adjusted R^2 ($\tilde{R}^2 = .825$) value of the R^2 value (.865).

When the sample size is small, it is recommended that \tilde{R}^2 value be used. As the sample size increase the difference between the resulting F values reduce considerably. Since our sample is obviously small, we will use unbiased estimator.

$$F = \frac{(n-k-1)R^2}{k(1-R^2)}$$

$$F = \frac{(10-2-1) \times 0.826}{2 \times (1-0.826)} = \frac{5.783}{0.348} = 16.635$$

The degrees of freedom employed in the significance testing of this F value are $df_{\text{num}} = k$ and $df_{\text{denominator}} = n - k - 1$.

For our example the degrees of freedom are as follows:

$$df_{\text{num}} = k = 2$$

$$df_{\text{denominator}} = n - k - 1 = 10 - 2 - 1 = 7$$

The tabled value for the $F_{(2,7)} = 4.737$ at 0.05 level of significance and the $F_{(2,7)} = 9.547$ at 0.01 level of significance. The calculated value of the F 16.635 is greater than the critical value of F . so we reject the null hypothesis and accept the alternative hypothesis which stated that the P^2 is not zero at less than 0.01 level.

You could have computed the F value using the R^2 instead of adjusted R^2 (\tilde{R}^2).

Often the statistical packages report the significance of R^2 and not significance of adjusted R^2 (\tilde{R}^2).

It is the judgment of the researcher to use either of them. In the same example if R^2 value is substituted for the adjusted R^2 (\tilde{R}^2) value then the F is 22.387 that is significant at .01 level.

3.6 LET US SUM UP

In this unit we have learned about the interesting procedures of computing the correlations. Especially, when we are interested in controlling for one or more variable. The multiple correlations provide us with an opportunity to calculate correlations between a variable and a linear combination of other variable. You practice them by solving some of the example given below, and you will understand the use of them.

3.7 UNIT END QUESTIONS

Problems

- 1) A clinical psychologist was interested in testing the relationship between health and stress. But she realised that coping skills will have an influence on this relationship so she administered General Health Questionnaire, Stress Scale and a coping scale. The data was collected on 15 individuals. Calculate the multiple correlation for this problem and test the significance. The data is as follows:

Health	Stress	Coping
9	5	18
10	5	21
8	7	17
7	4	16
8	7	22
9	6	19
12	8	25
11	3	17
10	5	20
14	6	22
7	7	18
9	9	22
6	7	17
16	3	20
14	8	26

In addition, answer the following questions:

Which correlation coefficient she should compute, if she wants to control the relationship between stress and health for the effect of coping?

Write a null and alternative hypothesis for partial correlation.

Correlation and Regression Calculate the partial correlation between stress and health controlled for the effect of coping.

Test the significance of the relationship.

Which correlation coefficient she should compute, if she wants to control the relationship between stress and health for the effect of coping only on stress?

Write a null and alternative hypothesis for part correlation.

Calculate the part correlation between stress and health controlled for the effect of coping on stress and test the significance.

Which correlation coefficient she should compute, if she wants to know the relationship between health (Y) and linear combination of stress (X_1) and coping (X_2)?

Answer:

a) Partial correlation; b) $H_0: r_p = 0$ and $H_A: r_p \neq 0$; c) -0.77 ; d) significant at 0.01 level; e) part (semipartial); f) $H_0: r_{sp} = 0$ and $H_A: r_{sp} \neq 0$; g) -0.62 , $p < .05$. h) Multiple correlation; i) $R = 0.86$, $p < .01$.

2) A social psychologist was interested in testing the relationship between attitude towards women (ATW) and openness to values (OV). But she realised education (EDU) will influence this relationship so she administered attitude to women scale, openness to values scale and also recorded the years spent in formal education. The data was collected on 10 individuals.

Calculate the multiple correlation for this problem and test the significance.

The data is as follows:

ATW	OV	Edu
2	7	14
4	10	13
8	14	11
7	13	9
8	9	5
9	10	14
1	6	5
0	9	6
6	12	11
5	10	12

In addition, answer the following questions:

- 1) Which correlation coefficient she should compute, if she wants to control the relationship between ATW and OV for the effect of EDU?
- 2) Write a null and alternative hypothesis for partial correlation.
- 3) Calculate the partial correlation between ATW and OV controlled for the effect of EDU.
- 4) Test the significance of the relationship.

- 5) Which correlation coefficient she should compute, if she wants to control the relationship between ATW and OV for the effect of EDU only on OV?
- 6) Write a null and alternative hypothesis for part correlation.
- 7) Calculate the part correlation between ATW and OV controlled for the effect of EDU on OV and test the significance.
- 8) Which correlation coefficient she should compute, if she wants to know the relationship between ATW (Y) and linear combination of OV (X_1) and EDU (X_2)?
- 9) Write a null and alternative hypothesis for multiple correlation.

Answer:

- a) Partial correlation; b) $H_0: r_p = 0$ and $H_A: r_p \neq 0$; c) .64; d) insignificant (the n is small); e) part (semipartial); f) $H_0: r_{sp} = 0$ and $H_A: r_{sp} \neq 0$; g) 0.61, $p > .05$.
h) Multiple correlation; i) $R = 0.67$, $p > .05$.

3.8 SUGGESTED READINGS

Garrett, H.E. (19). *Statistics In Psychology And Education*. Goyal Publishing House, New Delhi.

Guilford, J.P.(1956). *Fundamental Statistics in Psychology and Education*. McGraw Hill Book company Inc. New York.