Subject	PSYCHOLOGY
Paper No and Title	Paper No. 2 Quantitative methods
Module No and Title	Module No. 18 Types of Correlation- I
Module Tag	PSY_P2_M18

Principal Investigator	Co-Principal Investigator	Co- Principal Investigator (Technical)
Prof N.K.Chadha Head and Professor, Department of Psychology, University of Delhi	Dr Jaswinder Singh (Principal) and Dr.H.V.Jhamb, Associate Professor SGTB Khalsa College University of Delhi	Dr Vimal Rarh Deputy Director, Centre for e learning Assistant Professor, Department of Chemistry, SGTB Khalsa College, University of Delhi *Specialised in : e-Learning and Educational Technologies*
Paper Coordinator	**Author**	**Reviewer**
Dr. Monika Rikhi Associate Professor Sri aurobindo college(eve) University of Delhi	Dr. Monika Rikhi Associate Professor Sri aurobindo college(eve) University of Delhi	Dr. Rajinder K Sokhi Additional Director, RAC/DRDO, Delhi.
Anchor Institute : SGTB Khalsa College, University of Delhi		

TABLE OF CONTENTS

4. Biserial correlation

4.1 The concept of biserial correlation

4.1.1 Bi serial correlation coefficient is useful in:

4.1.2 Bi serial correlation coefficient is NOT useful in:

4.1.3 Calculation formula for biserial correlation coefficient

4.1.4 Calculation formula for biserial correlation coefficient

4.2 Numerical Application of bi- serial correlation

4.2 Numerical Application of bi- serial correlation

4.3 Significance of biserial r

5. Curvilinear relationships

6. IBM SPSS 20 application to correlation

7. Summary

1. Learning Outcomes

After studying this module, you shall be able to

- Know about the other important correlational methods
- Learn the concept and applications of partial correlation
- Learn the concept and applications of canonical correlation

2. Introduction

The concept of correlation

A mother wants to know if providing positive feedback on a dance performance is going to improve the child's performance in it. The three possible results are:

- Positive feedback will improve performance: indicative of a positive relation
- The performance would remain same and show no change: no relation
- The performance would decline: indicative of a negative relation

The concept of correlation coefficient is invaluable to quantitative research. The Pearson's product moment correlation coefficient is a linear relationship among two continuous variables and it has been explained in the module 22. Many a times the research data does not conform to the assumptions of linearity and continuity and hence, not suitable for Pearson's "r". So, other measures of correlation have been also been devised by statisticians.

The present module discusses the statistical relationship among the variables. We shall attempt to understand the different methods in the current and next modules. The data may be in the form of descriptive categories, non- linear and multi- variate. The modules 23 and 24 discuss some of these correlation methods and module 25 focuses on the non- parametric alternatives.

Module 23: Biserial, Point – Biserial, Curvilinear

Module 24: Partial, Multiple

Module 25: Rank Order, Phi Coefficient, Tetrachoric

3. Point- Biserial Correlation

3.1 The concept of point- biserial correlation

The point biserial correlation coefficient is an estimate of product moment correlation under special circumstances when certain assumptions have been met. A research data may comprise of continuous as well as categorically measured variables. If the categorical variable is dichotomous then the relationship between the two variables can be established using point- biserial correlation.

The term dichotomous literally means cut into parts. So, if a variable can be measured in just two categories it can be called dichotomous. An example of a dichotomous variable is gender; "male" or "female" there is no underlying continuum between the two categories.

In point –biserial correlation coefficient (r_{pb}) one of the variables is categorical and dichotomous into mutually exclusive categories and the other is continuous. So, if a study involves exploring the relationship between gender and anxiety point biserial correlation coefficient is an appropriate measure. The anxiety is a continuous variable whereas; gender is a categorical dichotomous variable.

3.1.1 Point bi serial correlation coefficient is useful in:

- When the data comprises of a continuous and other dichotomous variable
- The data is not suitable for pearson's product moment correlation coefficient
- Analysis of items of a test i.e. in item test correlation

3.1.2 Calculation formula for point biserial correlation coefficient

$$r_{pb} = \frac{M_p - M_q}{\sigma} * \sqrt{pq} \quad \text{---------eqn 1}$$

where, M_p and M_q are means of the two categories

p is the proportion of the sample in the first group

q is the proportion of the sample in the second group

σ is the standard deviation of the entire group

3.1.3 The calculation of point biserial correlation coefficient using product moment formula

The point biserial correlation coefficient is a product moment correlation and can also be calculated using the formula:

$$r_{pb} = \frac{(N\Sigma XY) - (\Sigma X . \Sigma Y)}{\sqrt{\{[(N\Sigma Y^2) - (\Sigma Y)^2] [(N\Sigma X^2) - (\Sigma X)^2]\}}} \quad \text{----------eqn. (2)}$$

3.2 Numerical Application of point bi- serial correlation

The researcher wishes to explore if anxiety scores obtained during a task performance were related to the gender of the group being tested. The scores on anxiety test as obtained by the males and females are presented under table I below:

Table I: the anxiety scores obtained by males (1) and females (0)

S.No.	Anxiety (X)	Gender(Y) Male (1) Female (0)	X^2	Y^2	XY
1	25	1	625	1	25
2	23	1	529	1	23
3	18	0	324	0	
4	24	0	576	0	
5	23	1	529	1	23
6	20	0	400	0	
7	19	0	361	0	
8	22	1	484	1	22
9	21	1	441	1	21
10	23	1	529	1	23
11	21	0	441	0	
12	20	0	400	0	
13	21	1	441	1	21
14	21	1	441	1	21
15	22	1	484	1	22
N_{male} = 9 N_{female}= 6	Total= 323		ΣX^2= 7005	ΣY^2= 9	ΣXY= 201

M_{males}=201/9= 22.33 **p= 9/15= 0.6**

$M_{females}$= 122/6= 20.33 **q= 6/15 = 0.4**

M_{total} = 323/ 15 = 21.53 **σ_{total}**

Using equation 1 the point biserial correlation coefficient can be calculated as:

$r_{pb} = (22.33 - 20.33)/1.82\sqrt{(0.6 * 0.4)}$

$= 0.54$

The same value of point biserial correlation coefficient can also be obtained using product moment correlation formula given in equation (2).

$$\mathbf{r_{pb}} = ((\mathbf{15} * \mathbf{201}) - (\mathbf{323} * \mathbf{9}))/\sqrt{(\{[(\mathbf{15} * \mathbf{9} - (\mathbf{9})]\ [(\mathbf{15} * \mathbf{7005}) - (\mathbf{323})^\wedge\mathbf{2}\ ]\})}$$

$$= .54$$

3.3 Significance of point biserial r

The obtained value of point biserial r can put to test against the null hypothesis. The degree of freedom (n- 2) can be used to find the critical value of r from the table. If the r_{calc} is more than r_{crit} it can be taken as significant and reject the null hypothesis.

4. Biserial correlation

4.1 The concept of biserial correlation

The concept of biserial correlation coefficient is quite similar to the point biserial correlation. The difference between the two pertains to the measurement of the categorical variable. The dichotomy in the variable is not true but has an underlying continuity in it. In other words, one may say that the variable in which dichotomy is assumed may be found to be continuous and normally distributed had more information been taken into account. Like, dividing the students into those who are graduates and those who are not, but one ignores small section of those who may have attempted graduation studies but unable to complete or are in the process of accomplishing it. So, the categorical variable cannot be said to be split in two categories of graduates and non- graduates rather an arbitrary split point is created to facilitate the grouping.

Whenever, the research requires establishing relationship between two variables one of which is continuous and the other is dichotomized by placing an arbitrary division, the biserial correlation is a suitable choice.

4.1.1 Bi serial correlation coefficient is useful in:

- When the data comprises of a continuous and other dichotomous variable (with arbitrary division imposed)
- The data is does not meet all the assumptions of pearson's product moment correlation coefficient
- Analysis of items of a test i.e. in item test correlation

4.1.2 Bi serial correlation coefficient is NOT useful in:

- The biserial r is not a very popular method as it is difficult to calculate when the data is not normal. (in such situations "phi" can be used, which is discussed in subsequent module)

- Biserial r cannot be used in a regression equation
- Biserial r cannot be used for comparing with other correlation coefficients as it goes beyond the range of + 1.0 and -1.0.

4.1.3 Calculation formula for biserial correlation coefficient

The biserial correlation coefficient can be calculated using the following formula:

$$r_{bis} = \frac{M_p - M_q}{\sigma} * \frac{pq}{u} \quad \text{-----------eqn(3}$$

Where,

M_p is the mean of the group in category 1

M_q is the mean of the group in category 2

σ is standard deviation of the group

p is proportion of group in category 1

q is proportion of group in category 2

u is the height of the normal curve ordinate dividing the two parts p and q.

4.1.4 calculation formula for biserial correlation coefficient

An alternative formula to calculate biserial r is also available, this is slightly convenient than the earlier one mentioned in equation 3.

$$r_{bis} = \frac{M_p - M_t}{\sigma} * \frac{p}{u} \quad \text{---------eqn(4)}$$

4.2 Numerical Application of bi- serial correlation

A school counselor wants to check out if a special training program she devised is related to performance in numerical ability test for students. Among the 145 students 21 were randomly chosen to undergo the training program. The two groups were of 21 students with training and 124 without training. The scores obtained are shown in table II.

Table II: Scores obtained on numerical ability test by trained and non- trained students.

Scores	Trained group (I) f	Non trained group (II) f	Total f
95- 99	5	6	11
90- 94	2	16	18
85- 89	6	19	25
80- 84	6	27	33
75- 79	1	19	20
70- 74	0	21	21
65- 69	1	16	17
	N= 21	N= 124	N= 145

$M_t = 81.35$ $M_p = 87.0$ $M_q = 80.39$

$\sigma = 8.8$ p = .145(21/145) q = .855(124/145) u = .228

the value of u is obtained from the table that gives the heights of the ordinates in a normal distribution of unit areas, with N= 1, M= 0 and SD= 1. As evident and explained by fig I one can find u= .228.

$$\mathbf{r}_{\mathbf{bis}} = \frac{\mathbf{87-80.39}}{\mathbf{8.8}} * \frac{(\mathbf{0.145*0.855})}{\mathbf{0.228}}$$

$$= .41$$

The same value of biserial correlation can be obtained using equation 4 also.

4.3 Significance of biserial r

The obtained value of point biserial r can put to test against the null hypothesis. The degree of freedom (n- 2) can be used to find the critical value of r from the table. If the r_{calc} is more than r_{crit} it can be taken as significant and reject the null hypothesis.

5. Curvilinear relationship

Let us recall what we learnt in module 22, the pearson's product moment correlation is best suited to data that is linear in nature, otherwise error may occur. The research data however is not linearly related always. When the straight line of fit is not apt enough to describe a set of data it is said to be curvilinear or more simply a non- linear relationship. In such situations r is not an adequate measure of correlation. In such data a curve rather than a line of best fit is used to describe the degree of relationship that the variables share. Fitting a curve instead of straight line would give a better and more accurate measure of correlation among the variables.

One often encounters non- linearly related variables in data measured in ratio scale, psycho-physics etc.

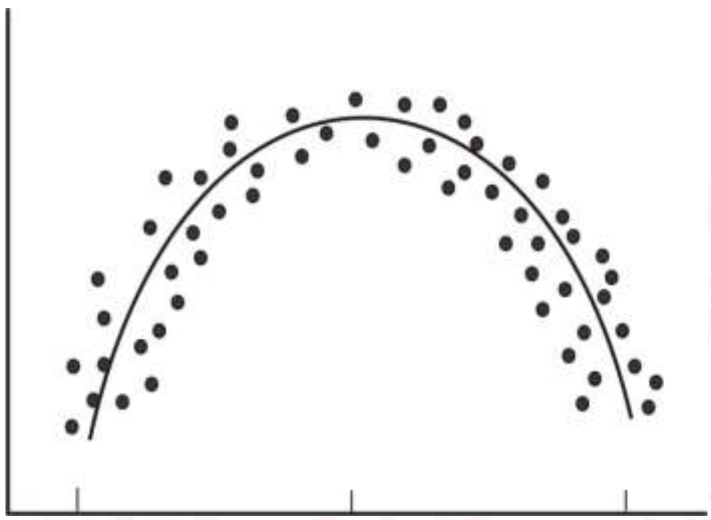

FIG II: curvilinear relationships

6. IBM SPSS 20 Application to correlation

The IBM SPSS 20 cannot be used directly to calculate biserial correlation, point biserial correlation coefficient is calculated and then the equation is adjusted to calculated biserial correlation.

Let us take the numerical example used in section 3.2 and apply the IBM SPSS20 commands to it.

The researcher wishes to explore if anxiety scores obtained during a task performance were related to the gender of the group being tested. The scores on anxiety test are continuous as obtained by the males and females; the dichotomous variable of gender is coded as males (1) and females (0).

- Analyze > Descriptive statistics > frequencies Show further steps

7. Summary

- Many a times the research data does not conform to the assumptions of linearity and continuity and hence, not suitable for Pearson's "r". So, other measures of correlation have been also been devised by statisticians
- The point biserial correlation coefficient is an estimate of product moment correlation under special circumstances when certain assumptions have been met.
- A research data may comprise of continuous as well as categorically measured variables. If the categorical variable is dichotomous then the relationship between the two variables can be established using point- biserial correlation.
- In point –biserial correlation coefficient (r_{pb}) one of the variables is categorical and dichotomous into mutually exclusive categories and the other is continuous.
- Calculation formula for point biserial correlation coefficient and biserial correlation are also discussed in the text.
- The concept of biserial correlation coefficient is quite similar to the point biserial correlation. The difference between the two pertains to the measurement of the categorical variable.
- The dichotomy in the variable is not true but has an underlying continuity in it.
- Biserial r cannot be used in a regression equation and cannot be used for comparing with other correlation coefficients as it goes beyond the range of + 1.0 and -1.0.
- When the straight line of fit is not apt enough to describe a set of data it is said to be curvilinear or more simply a non- linear relationship. In such situations r is not an adequate measure of correlation.