academicresearchJournals

Vol. 8(4), pp. 99-106, July 2020
DOI: 10.14662/IJELC2020.085

ISSN: 2360-7831
http://www.academicresearchjournals.org/IJELC/Index.htm

International Journal of English Literature and Culture

Review

REVIEW ON CORRELATION RESEARCH

Wubante Mekonnen

Debre Markos University, Ethiopia, P.O.B 269

Accepted 6 July 2020

Correlational research is a type of non-experimental research method, in which a researcher measures two variables, understands and assesses the statistical relationship between them with no influence from any extraneous variable. It is designed to test research hypotheses in cases where it is not possible or desirable to experimentally manipulate the independent variable of interest. It is also desirable because it allows the investigation of behavior in naturally occurring situations. This article reviews the purpose, design, methods, threats and other aspects like pros and cons of correlational research. It also considers the possible topics that can be investigated in correlational research.

Key Words: Correlation, Design, Method

Cite This Article As: Wubante M (2020). REVIEW ON CORRELATION RESEARCH. Inter. J. Eng. Lit. Cult. 8(4): 99-106

INTRODUCTION

What is Correlation Research?

- ❖ According to Creswell, correlational research designs are used by investigators to describe and measure the degree of relationship between two or more variables or sets of scores.
- ❖ It is a procedure in which subjects' scores on two variables are simply measured, without manipulation of any variables, to determine whether there is a relationship.
- ❖ Correlational research examines the relationship between two or more non-manipulated variables.
- ❖ Correlation is simply a statistical measurement of the relationship between two variables.
- ❖ Correlation refers to how much two things relate to each other.
- ❖ A correlation is a special kind of association; there is a linear relation between the values of the variables.
- ❖ Correlational research is sometimes called associational research.
- ❖ Sometimes correlational research is referred as a form of descriptive research.
- ❖ Describes the degree to which two or more quantitative variables are related.

Purposes of Correlation Research

• Two basic purposes:

Help explain important human behaviors (Explanatory Studies)

Explanatory Study: An explanatory correlational design explains or clarifies the degree of association among two or more variables at one point in time. Researchers are interested in whether two variables co-vary, in which a change in one variable is reflected in changes in the other. An example is whether motivation is associated with academic performance.

Predict Likely Outcomes (Prediction Studies)

In a prediction design, the investigator identifies variables that will positively predict an outcome or criterion. In this form of research, the researcher uses one or more predictor variables and a criterion (or outcome) variable. A prediction permits us to forecast future performance, such as whether a student's GPA in college can be predicted from his or her high school performance.

Correlational designs provide an opportunity for us to predict scores and explain the relationship among variables. In correlational research designs, investigators use the correlation statistical test to describe and measure the degree of association (or relationship) between two or more variables or sets of scores. In this design, the researchers do not attempt to control or manipulate the variables as in an experiment; instead, they relate, using the correlation statistic, two or more scores for each person (e.g., a student motivation and a student achievement score for each individual).

Correlation Research Design

- ❖ Typically oriented by research questions or hypotheses
- ❖ A relatively straightforward design:
- ❖ Identify variables for inclusion
- ❖ Formulate questions or hypotheses
- ❖ Select a random sample (preferably with $n > 30$)
- ❖ Obtain data for each member of the sample on each variable being investigated
- ❖ Compute correlations in order to determine degree of relationship

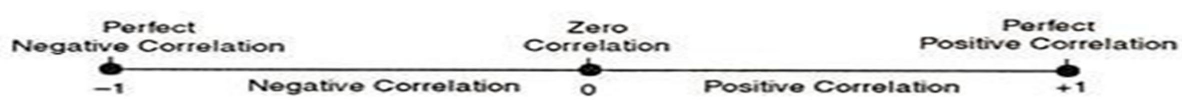
Types of Correlational Research

Essentially there are **three** types of correlational research that have been identified:

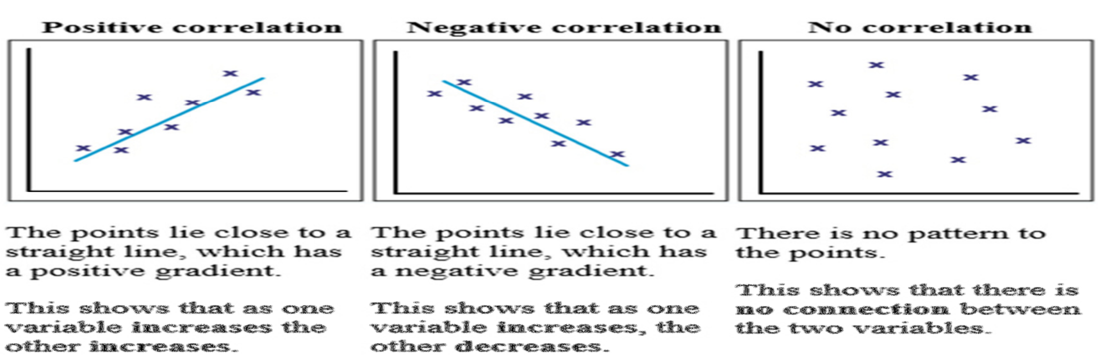
1. Positive Correlation: A positive correlation between two variables is when an increase in one variable leads to an increase in the other variable and a decrease in one variable will see a decrease in the other variable. For example, the amount of money a person has might positively correlate with the number of cars he has.

2. Negative Correlation: A negative correlation is quite literally the opposite of positive correlation. This means, if there is an increase in one variable, the second variable will show a decrease and vice versa. For example, the level of being educated might negatively correlate with the crime rate when an increase in one variable leads to a decrease in another and vice versa. This means if in some ways the level of education in a country is improved, it can lead to lowering the crime rates.

3. No Correlation: In this third type, two variables are not correlated. This means a change in one variable may not necessarily see a change in the other variable. For example, being a millionaire and happiness are not correlated. This means an increase in money doesn't lead to happiness.



The types of correlational relationships are summarized using diagrams as follows:



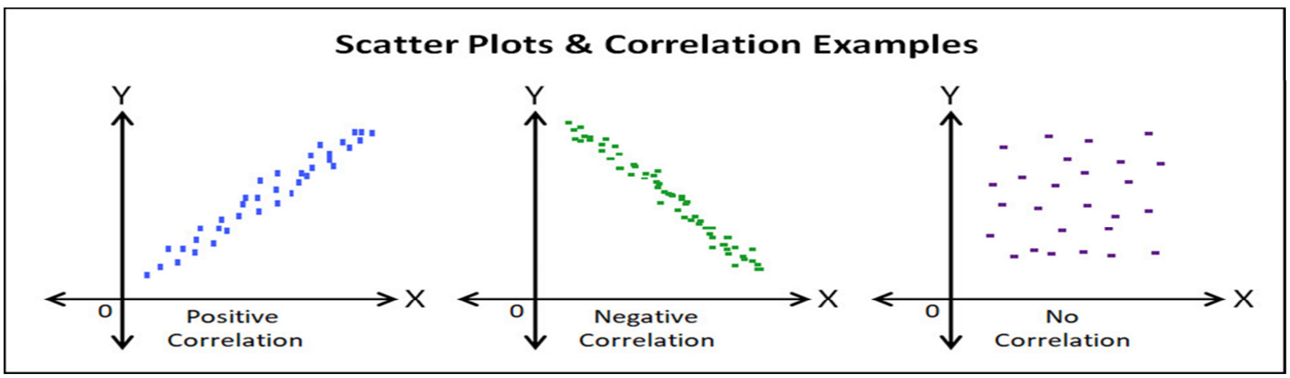
Methods of Computing Coefficient of Correlation

In ease of ungrouped data of bivariate distribution, the following three methods are used to compute the value of co-efficient of correlation:
1. Scatter Diagram Method
2. Pearson’s Product Moment Coefficient of Correlation
3. Spearman’s Rank Order Coefficient of Correlation

1. Scatter Diagram Method

Scatter diagram or dot diagram is a graphic device for drawing certain conclusions about the correlation between two variables.

In preparing a scatter diagram, the observed pairs of observations are plotted by dots on a graph paper in a two dimensional space by taking the measurements on variable X along the horizontal axis and that on variable Y along the vertical axis. The placement of these dots on the graph reveals the change in the variable as to whether they change in the same or in the opposite directions. It is a very easy, simple but rough method of computing correlation. The frequencies or points are plotted on a graph by taking convenient scales for the two series. The plotted points will tend to concentrate in a band of greater or smaller width according to its degree. ‘The line of best fit’ is drawn with a free hand and its direction indicates the nature of correlation. Scatter diagrams, as an example, showing various degrees of correlation are shown in the figure below.



2. Pearson’s Product Moment Coefficient of Correlation

- The Pearson correlation coefficient is just one of many types of coefficients in the field of statistics. The following lesson provides the formula, examples of when the coefficient is used, and its significance.
- The Karl Pearson’s product-moment correlation coefficient (or simply, the Pearson’s correlation coefficient) is a measure of the strength of a linear association between two variables and is denoted by **r or r_{xy}** (x and y being the two variables involved).
- This method of correlation attempts to draw a line of best fit through the data of two variables, and the value of the Pearson correlation coefficient, **r**, indicates how far away all these data points are to this line of best fit.
- The Pearson correlation coefficient is a very helpful statistical formula that measures the strength between variables and relationships. In the field of statistics, this formula is often referred to as the Pearson R test. When conducting a statistical test between two variables, it is a good idea to conduct a Pearson correlation coefficient value to determine just how strong that relationship is between those two variables.

Formula

In order to determine how strong the relationship is between two variables, a formula must be followed to produce what is referred to as the coefficient value. The coefficient value can range between -1.00 and 1.00. If the coefficient value is in the negative range, then that means the relationship between the variables is negatively correlated, or as one value increases, the other decreases. If the value is in the positive range, then that means the relationship between the variables is positively correlated, or both values increase or decrease together. Let's look at the formula for conducting the Pearson correlation coefficient value.

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

Where:

N = number of pairs of scores
Σxy = sum of the products of paired scores
Σx = sum of x scores
Σy = sum of y scores
Σx^2 = sum of squared x scores
Σy^2 = sum of squared y scores

Method: Calculating Correlation Coefficient

Step 1: Make a chart. Use the given data, and add three more columns: xy, x^2, and y^2.
Step 2: Multiply x and y together to fill the xy column.
Step 3: Take the square of the numbers in the x column, and put the result in the x^2 column
Step 4: Take the square of the numbers in the y column, and put the result in the y^2 column.
Step 5: Add up all of the numbers in the columns and put the result at the bottom of the column. The Greek letter sigma (Σ) is a short way of saying "sum of."
Step 6: calculate the correlation coefficient using the formula.
Step 7: Interpret the results

Example 1: Correlation between Two Variables

No.	**X**	**Y**	**X^2**	**Y^2**	**XY**
1	10	80	100	6400	800
2	30	60	900	3600	1800
3	50	140	2500	19600	7000
4	60	160	3600	25600	9600
N=4	∑X=150	∑Y=440	$\sum X^2_{=7100}$	$\sum Y^2_{=55200}$	$\sum XY_{=19200}$

The correlation of coefficient between X' and Y' will be:

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

Putting values from table,

$$r = \frac{4 \times 19200 - (150)(440)}{\sqrt{[4 \times 7100 - (150)^2][4 \times 55200 - (440)^2]}}$$

$$= \frac{10800}{\sqrt{5900 \times 27200}} = \frac{108}{126.68} = +0.85$$

The value of coefficient of correlation is **+ .85**. This shows a *high degree of relationship* between the two variables.

Example 2

Calculate the correlation of coefficient between the number of cigarettes smoked and the longevity of a test subject.

Solution

Let us first assign random variables to our data in the following way:
X – The number of cigarettes smoked
Y – Years lived
Let us now construct a table to compute all the values we are going to use in our correlation formula. Note that **N** here = 9

X	X^2	Y	Y^2	XY
25	625	63	3669	1575
35	1225	68	4624	2380
10	100	72	5184	720
40	1600	62	3844	2480
85	7225	65	4225	5525
75	5625	46	2116	3450
60	3600	51	2601	3060
45	2025	60	3600	2700
50	2500	55	3136	2750
ΣX = 425	**ΣX^2 = 24525**	**ΣY = 542**	**ΣY^2 = 33188**	**ΣXY = 24640**

$(\Sigma X)^2 = 425^2 = 180625$ **$(\Sigma Y)^2 = 542^2 = 293764$**

Using the values in the formula, we get**: r_{xy}=–0.61**

This implies a *negative correlation* between the considered variables i.e. the higher the number of cigarettes smoked per week in last 5 years, the lesser the number of years lived. Note that it DOES NOT mean that smoking cigarettes decreases the life span. Because there might be other factors responsible for one's death. Still, it is an important conclusion nevertheless.

3. Spearman's Rank Order Coefficient of Correlation

- When data are measured on, at least, an ordinal scale, the ordered categories can be replaced by their ranks and Pearson's correlation coefficient calculated on these ranks. This is called Spearman's rank correlation coefficient (r) and provides a measure of how closely two sets of rankings agree with each other.
- A correlation can easily be drawn as a scatter graph, but the most precise way to compare several pairs of data is to use a statistical test - this establishes whether the correlation is really significant or if it could have been the result of chance alone.
- Spearman's Rank correlation coefficient is a technique which can be used to summarize the strength and direction (negative or positive) of a relationship between two variables. The result will always be between 1 and minus 1.
- This non-parametric analysis tool provides a way to compare two sets of ordinal data (data that can be rank ordered in a meaningful manner). The result, r, is a measure of the association between two datasets.
- You may want to know if two reviewers have similar ratings for movies, or if two assessment techniques provide similar results. If **r** is 1 it means when one series increases the other does also. If **r** is -1, there is a negative relationship, meaning as one series increases the other decreases. At zero there is no relationship between the two series.
- This does not work for sets with a non-linear relationship; say a parabolic function for example.

Method: Calculating the Coefficient

Step 1: Collect the data, and create a table from your data.
Step 2: Rank the data: Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
Step 3: Calculate the difference in ranks **(D)**: This is the difference between the ranks of the two values on each row of the table.
Step 4: Square the differences **(D^2)** to remove negative values and then sum them (**D^2**).
Step 5: Sum column **D^2**
Step 6: Calculate **r**
Step 7: Interpret the results

Calculating Correlation from Ranked Data

Rank Correlation

❖ The formula of Spearman's Rank Correlation Coefficient is given as:

$$\mathbf{r} = \frac{1-6x\sum D^2}{N\,(N^2-1)}$$

r= the coefficients of rank correlation
$D^{2=}$ the difference in rank between paired values of X and Y
N= sample size

Example 1: In a speech contest two professors judged 10 students. Their judgments were in ranks, which are presented below. Determine the extent to which their judgments were in agreement.

Students	Professor K's Ranks (R-1)	Professor M's Ranks (R-2)	Difference (D= R1-R2)	D^2
A	1	1	0	0
B	3	2	1	1
C	4	5	-1	1
D	7	9	-2	4
E	6	6	0	0
F	9	8	1	1
G	8	10	-2	4
H	10	7	3	9
I	2	4	-2	4
J	5	3	2	4
N=10			**∑D=0**	**$\sum D^2_{=28}$**

$$\mathbf{r} = \frac{1-6x\sum D^2}{N\,(N^2-1)} = \frac{1-6x28}{10(10^2-1)} = 1-\frac{168}{990} = 1-.17 = \mathbf{+.83}$$

The value of coefficient of correlation between scores in Mathematics and General Science is *positive* and *high*.

Calculating r from Test Scores

Example 2: The following data give the scores of 5 students in Mathematics and General Science respectively: Compute the correlation between the two series of test scores by Rank Difference Method.

Students	Score in Math	Score in Gen. Sci.	Rank in Test 1 (R1)	Rank in Test 2 (R2)	Diff. in Ranks D (R1-R2)	D^2
A	8	10	2	1	1	1
B	7	8	3	2	1	1
C	9	7	1	3	-2	4
D	5	4	4	5	-1	1
E	1	5	5	4	1	1
N=5					**∑D=0**	**∑D2=8**

$$r = \frac{1-6x\sum D^2}{N\,(N^2-1)} = \frac{1-6x8}{5(5^2-1)} = 1-\frac{48}{120} = 1-.40 = \mathbf{+.60}$$

The value of coefficient of correlation between scores in Mathematics and General Science is *positive* and *moderate.*

Interpretation of the Coefficient of Correlation

- Correlation coefficient is used to test the strength of relationship between two variables.
- A coefficient can range from r = **+1.00 to -1.00**.
- Merely computation of correlation does not have any significance until and unless we determine how large must the coefficient be in order to be significant, and
- What does correlation tell us about the data?
- What do we mean by the obtained value of coefficient of correlation?

Size of Correlation	Interpretation
± 1	Perfect Positive/Negative Correlation
± .90 to ± .99	Very High Positive/Negative Correlation
± .70 to ± .90	High Positive/Negative Correlation
± .50 to ± .70	Moderate Positive/Negative Correlation
± .30 to ± .50	Low Positive/Negative Correlation
± .10 to ± .30	Very Low Positive/Negative Correlation
± .00 to ± .10	Markedly Low and Negligible Positive/ Negative Correlation

(Cohen, Manion & Morrison, 2000)

Threats to Internal Validity

• There are some threats identified in conducting correlational research:

1. Subject Characteristics

- Individuals or groups have two or more characteristics; might be a cause of variation in the other two variables.

2. Location

- Location is different for different subject.
- One location may be more comfortable compared to others.

3. Instrumentation

- Instrument decay; care must be taken to ensure the observers don't become tired, bored or inattentive.
- Data collector characteristics; different gender, age or ethnicity may affect specific response.

4. Testing: Experience of responding to the first instrument may influence subject responses to the second instrument.

5. Mortality: Loss of subjects may make a relationship more (or less) likely in the remaining data.

Strengths and Limitations of Correlational Research

Strengths	Limitations
1. Can collect much information from many subjects at one time. 2. Can study a wide range of variables and their interrelations. 3. Allow researchers an opportunity to investigate variables that would be impossible or unethical to manipulate. 4. Correlational studies tend to have high external validity.	1. Correlation does not indicate causation (cause and effect). 2. Problems with self-report method. 3. Individuals are not studied in-depth. 4. Correlational studies tend to have low internal validity.

Correlation Research Questions/ Topics: Examples

1. What is the relationship between gender, academic performance and university drug use?
2. Is there an association between personality type and seniority in companies?
3. What is the relationship between obesity and heart disease?
4. Is there a relationship between family income and grade point average?
5. What is the relationship between education and income?
6. Is there a relationship between intelligence and self-esteem?
7. What relationship exists between anxiety and achievement?
8. Use of aptitude test to predict success

CONCLUSION

Correlational research is designed to test research hypotheses in cases where it is not possible or desirable to experimentally manipulate the independent variable of interest. It is also desirable because it allows the investigation of behavior in naturally occurring situations.

REFERENCES

Cohen, L., Manion, L. & Morrison, K. (2000). Research Methods in Education (5th ed.). London: New York.

Crawford, M. (2014). Strengths and Limitations of Correlational Design. Walden University.

Creswell, J. (2012). Educational Research (4th ed.). Boston: Pearson Education Inc.

Fitzgerald, S., Rumrill, P., & Schenker, J. (2004). Correlational Designs in Rehabilitation Research. Journal of Vocational Rehabilitation, 20 (2), 143-150.

Gallitano, G. (2002). Identifying Predictive and Correlational Variables in Male Recidivistic Criminal Offenders (Doctoral Dissertation). Retrieved from Pro-Quest Information and Learning Company.

Kumar, R. (2005). Research Methodology. SAGE Publications.

Mujis, D. (2004). Doing Quantitative Research in Education. SAGE Publications.

Stangor, C. (2011). Research Methods for the Behavioral Sciences. Boston: Houghton Mifflin Company.