International Journal of Sociology and Anthropology Research
Vol.1, No.3, pp.22-28, November 2015
Published by European Centre for Research Training and Development UK (www.eajournals.org)

THE RELEVANCE AND SIGNIFICANCE OF CORRELATION IN SOCIAL SCIENCE RESEARCH

Maiwada Samuel and Lawrence Ethelbert Okey

Department of Sociology, University of Jos,
PMB 2084, Jos, Nigeria

ABSTRACT: *As important as statistics in the social sciences are, their application to real life situation has been minimal. Many scientific discoveries of great importance would have been impossible if scientists had only conceived of the world in terms of certainty. In many situations studied by scientists, and most certainly in all situations studied in social sciences, researchers can at best identify and measure imperfect associations between variables. Drawing largely from secondary sources, this paper examined the relevance and significance of correlation in social science. Findings showed that correlation is indispensable especially in studies that require the understanding of certainty and the degree to which variables show a mutual association.*

KEYWORDS: Research, Statistics, Correlation, Variable,

INTRODUCTION

Statistics is among the most important tool used by social scientists in doing their research (Fisher, 1929). From market forecasting in economics to general social behaviour in sociology, statistics has become a veritable instrument that most social researchers cannot do without. Correlation, a test of relationships and associations between variables is one of most important statistical technique employed in social scientific studies. This paper argues for the relevance and significance of the Pearson's correlation coefficient (r) and coefficient of determination (r^2) in social and behavioural science research. The author, rather deliberately, did not discuss other aspects of correlation such as the Spearman's correlation and focussed only on Pearson's correlation coefficient while leaving the former as a topic for another paper.

When analyzing vast amounts of data, simple statistics can reveal a great deal of information. However, it is often more important to examine relationships within the data, especially in the social sciences. Through correlation measures, these relationships can be studied in-depth, limited only by the data available to the researcher. This paper will attempt to explain these powerful tools and techniques with a statistical background and concise examples.

In the next part of the paper, the concept of correlation, and particularly, as it concerns Pearson's correlation coefficient, is defined and explained and the important elements under the concept where clarified. An example was worked concisely, in such a way that even those in first contact with the topic will find it attractive. The coefficient of determination was also briefed. Some relevance and significance of correlation in social science research were discussed in bullet points and the paper was concluded with a summary of the issues discussed and a re-emphasis on the significance of the correlation in social science research.

ISSN 2059-1209, ISSN 2059-1217

THE CONCEPT: CORRELATION

Correlation in social science research talks about relationships and association between variables. According to Ibanga, the two terms (i.e "relationships" and "association") mentioned above "are often used interchangeably; and they refer to the extent to which one variable changes (in quantity or quality) in response to change in another variable" (1992:137).

According to Coven (2003) there are different types of correlation depending on the number of variables involved. They include the simple, partial and multiple correlations.

Simple, Partial and Multiple Correlations: In simple correlation, relationships between two variables are studied. In partial correlation more than two variables are studied, but the effect on one variable is kept constant and the relationship between the other two variables is studied. Three or more variables are simultaneously studied in multiple correlations.

While simple correlation is good in understanding simple relationship between variables, it would appear that multiple correlation yields better results in the social sciences. This is partly because social phenomena are increasingly being understood from different perspectives and therefore, require more sophisticated methods to analyse.

Linear and non-linear correlation: Correlation depends upon the constancy of the ratio of change between the variables. In linear correlation, the percentage change in one variable will be equal to the percentage change in another variable. It is not so in non-linear correlation.

Measurement: Usually, correlation is described in terms of its direction and strength;

a. **Strength:** In describing the strength of a relationship, it could be strong, moderate or weak. The extremes of strength will be a perfect relationship which is 1, or a 0 (zero) relationship which is also known as spurious or no relationship.

b. **Direction:** in terms of direction, a relationship between one or more variable can either be positive or negative. These relationships, whether positive or negative, can also be perfect, strong, moderate, weak or spurious.

Correlation is a statistical measurement of the relationship between two variables. Possible correlations range from +1 to –1. A zero correlation indicates that there is no relationship between the variables. A correlation of –1 indicates a perfect negative correlation, meaning that as one variable goes up, the other goes down. A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together.

Correlation Co-efficient Definition:
A measure of the strength of linear association between two variables. Correlation will always fall between -1.0 and +1.0. If the correlation is positive, we have a positive relationship. If it is negative, the relationship is negative.

Correlation Co-efficient:

The quantity *r*, called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the *Pearson product moment correlation coefficient* in honour of its developer Karl Pearson.

ISSN 2059-1209, ISSN 2059-1217

The value of *r* is such that -1 ≤ *r* ≤ +1. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.

Positive Correlation: If *x* and *y* have a strong positive linear correlation, *r* is close to +1. An *r* value of exactly +1 indicates a perfect positive fit. Positive values indicate a relationship between *x* and *y* variables such that as values for *x* increase, values for *y* also increase.

Negative Correlation: If *x* and *y* have a strong negative linear correlation, *r* is close to -1. An *r* value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between *x* and *y* such that as values for *x* increase, values for *y* decrease.

Spurious Correlation: If there is no linear correlation or a weak linear correlation, *r* is close to 0. A value near zero means that there is a random, non-linear relationship between the two variables

Note that *r* is a dimensionless quantity. That is; it does not depend on the units employed. A Perfect Correlation of ± 1 occurs only when the data points all lie exactly on a straight line. If *r* = +1, the slope of this line is positive. If *r* = -1, the slope of this line is negative. A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined. A study utilizing 'core' scientific data may require a stronger correlation than a study using social science data. A dependent variable's values are continuously being changed by its relationship with the independent variable. However, an independent variable's values are not changed by its relationship with another variable. All variables have values, but not all are necessarily related.

The mathematical formula for computing Pearson's *r* is:

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

From the formula above, the numerator shows the extent to which the independent variable (x) and the dependent (y) correlate or move together while the denomenator shows the extent to which both variables co-vary.

The symbols in the formula are interpreted thus:

n = Number of values or observations
x = Independent variable (First Scores)
y = Dependent variable (Second Scores)
Σxy = Sum of the product of first and Second Scores
Σx= Sum of First Scores
Σy = Sum of Second Scores
Σx^2 = Sum of square First Scores
Σy^2 = Sum of square Second Scores

ISSN 2059-1209, ISSN 2059-1217

Correlation Co-efficient Example: To find the Correlation of

X Values	Y Values
60	3.1
61	3.6
62	3.8
63	4
65	4.1

Step 1: Count the number of values.
N = 5

Step 2: Find XY, X^2, Y^2
See the below table

X Value	Y Value	X*Y	X*X	Y*Y
60	3.1	60 * 3.1 = 186	60 * 60 = 3600	3.1 * 3.1 = 9.61
61	3.6	61 * 3.6 = 219.6	61 * 61 = 3721	3.6 * 3.6 = 12.96
62	3.8	62 * 3.8 = 235.6	62 * 62 = 3844	3.8 * 3.8 = 14.44
63	4	63 * 4 = 252	63 * 63 = 3969	4 * 4 = 16
65	4.1	65 * 4.1 = 266.5	65 * 65 = 4225	4.1 * 4.1 = 16.81

Step 3: Find ΣX, ΣY, ΣXY, ΣX^2, ΣY^2.
$\Sigma X = 311$
$\Sigma Y = 18.6$
$\Sigma XY = 1159.7$
$\Sigma X^2 = 19359$
$\Sigma Y^2 = 69.82$

Step 4: Now, Substitute in the above formula given.
Correlation(r) =[$N\Sigma XY - (\Sigma X)(\Sigma Y)$ / Sqrt($[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]$)]
= ((5)*(1159.7)-(311)*(18.6))/sqrt([(5)*(19359)-$(311)^2$]*[(5)*(69.82)-$(18.6)^2$])
= (5798.5 - 5784.6)/sqrt([96795 - 96721]*[349.1 - 345.96])
= 13.9/sqrt(74*3.14)
= 13.9/sqrt(232.36)
= 13.9/15.24336
= 0.9119

This example is a guide to find the relationship between two variables by calculating the Correlation Co-efficient from the above steps.

ISSN 2059-1209, ISSN 2059-1217

It is pertinent, however, to note that Correlation is *not* causality (Afonja, 1982; Kenny, 1987; Nunes & Bryant, 2011; Sotos et al and Yount, 2006). People commonly confuse correlation with causation. Correlational data do not indicate cause-and-effect relationships. When a correlation exists, (as mentioned earlier in this paper), changes in the value of one variable reflect changes in the value of the other. The correlation does not imply that one variable causes the other, only that both variables somehow relate to one another.

Coefficient of Determination (r^2)

After calculating the strength of the relationship using Pearson's the coefficient correlation, we can go a step further to calculate the coefficient of determination (r^2) to find out the amount of variation in the dependent variable which explains its relationship with the independent variable. The coefficient of determination (r^2) shows, in percentage terms, the amount of variation in the independent variable. It also helps one to understand or speculate about the unexplained variation. In other words, we can, using coefficient of determination (r^2), to try to explain in percentage terms, the amount of variation in y that is explained by its relationship with $X_1 X_2 X_3$...

RELEVANCE AND SIGNIFICANCE

The usefulness of correlation in social science research cannot be overemphasised. Establishing relationships and associations between variables, as ordinary as it may seem, does a lot to the social science researcher. Briefly discussed below are some of the relevance and significance of correlation in social science research.

- Correlation matrices (generally Pearson) are among the most widely used techniques for studying the construct validity of data in factor analysis, whether exploratory or confirmatory, and this method is used to obtain factor solutions (Holgado –Tello P. et al 2011).

- Correlation provides the platform for regression to predict the values of the dependent variable based on the known relationship that exist between the independent variable and the dependent variable.

- Correlational research can also play an important role in the development and testing of theoretical models. Once the nature of bivariate relations has been determined, this information can then be used to develop theoretical models. The idea here is to attempt to explain the nature of the bivariate correlations rather than to simply report them. At this point, methods such as factor analysis, path analysis and structural equation modelling can come into play (Duncan, 1966).

- Correlational research has had and will continue to have an important role in quantitative research in terms of exploring the nature of the relations among a collection of variables. In part, unrelated variables can be eliminated from further consideration, thereby allowing the researcher to give more serious consideration to related variables.

- More sophisticated multivariate extensions enable researchers to examine multiple variables simultaneously (Stockwell, 2010).

ISSN 2059-1209, ISSN 2059-1217

- Once descriptive research has helped to identify the important variables, correlational research can then be used to examine the relations among those important variables. For example, researchers may be interested in determining which variables are most highly related to a particular outcome, such as student achievement. This can then lead into experimental research in which the causal relations among those key variables can be examined under more tightly controlled conditions. Here one independent variable can be manipulated by the researcher (e.g., method of instruction), with other related variables being controlled in some fashion (e.g., grade, level of school funding). This then leads to a determination of the impact of the independent variable on the outcome variable, allowing a test of strong causal inference.

The relevance and significance of correlation in social science research brought out in this paper as are by no means exhaustive but are considered some of the most important.

CONCLUSION

The paper tried to argue for the relevance and significance of Pearson's correlation in social science research by first, clarifying the key concepts in the topic of discussion and then enumerated some of the usefulness of correlation to the social scientist. We argued that Pearson's correlation is a *sine qua non,* particularly in social science research.

Pearson's correlation is an excellent analytical tool when implemented correctly; and even more so, when used with its 'first cousin', regression. The former establishes a relationship between variables showing the degree of the strength of such relationship as well as the direction, while the latter, based on the already established existing relationship, predict outcomes about social phenomena. Despite scathing criticisms from some authors (Schmitt, 1996; Mahoney, 2001; Zimmerman, Zumbo, & Williams, 2003) which is a normal practice in the academia and particularly in the social science intellectual circles, correlation still remain a very important tool in social science research and its techniques almost inevitable especially in quantitative studies involving variables.

REFERENCES

Afonja, B. (1982). *Introductory Statistics: A Learner's Motivated Approach*. Ibadan: Evans Brothers (Nigeria Publishers) Limited

Best, E.G. (2012). *Research Methods II* (Unpublished Lecture notes). Department of Sociology, University of Jos.

Coven, V. (2003). A History of Statistics in Social Science. *Gateway: An Academic Journal on the Web*,

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology, 72, 1-16.*

Fisher, I. (1929). The application of mathematics to the social sciences. 7th *Josiah Willard Gibbs Lecture,* read at Des Moines, December 31, before a joint session of the American Mathematical Society and the American Association for the Advancement of Science.

Holgado–Tello P. et al (2011). Polychoric versus Pearson Correlations In Exploratory and Confirmatory Factor Analysis of Ordinal Variables. *Quantity and Quality, 44,*

ISSN 2059-1209, ISSN 2059-1217

(1**)**, 153-166,. Retrieved from http://www.springerlink.com/content/

Ibanga, U. A. (1992). *Statistics for Social Sciences*. Jos: Centre for Development Studies, University of Jos.

Kenny, D. A. (1987) *Statistics for the Social and Behavioural Sciences*. Toronto: Little Brown and Co. Ltd.

Mahoney, J. (2001). Beyond Correlational Analysis: Recent Innovations in Theory and Method. *Sociological Forum, 16, .(3),* 576-593.

Nunes, T. & Bryant, P. (2011). Understanding risk and uncertainty: The importance of correlations. *Web Journal of Mathematics and Technology Education, 2* (2), 2-30.

Schmitt, N. (1996).Uses and abuses of coefficient alpha. *Psychological Assessment*. Vol. 6, No 4, 350-353.

Stockwell, I. (2010). *The Quest for Causation: An Introduction to Correlation and Regression Analysis*. Baltimore: The Hilltop Institute (UMBC).

Yount, R. (2006). *Research Design and Statistical Analysis in Christian Ministry (4th ed)*. Ft. Worth, TX: Department of Foundations of Education, School of Religious Education, Southwestern Theological Seminary.

Zimmerman, D. W., Zumbo, B. D. & Williams, R. H. (2003). Bias in Estimation and Hypothesis Testing of Correlation. *Psicológica, 24,* 133-158.

ISSN 2059-1209, ISSN 2059-1217