ISSN: 2708-4841(Print), 2708-7042 (Online)

Volume: 7, Issue: 1
Page: 1-23
YEAR: 2026

Commonwealth Journal of Academic Research (CJAR.EU)

Appropriate Use of Correlation Coefficient in Educational Research

AUTHOR(S): ABE Thomas Olabode, GBORE Lawrence Olu, OSAKUADE Joseph Oluwatayo, OWABUMOYE Ogunleye Sunday, AJELE Ayodeji Johnson, OGUNBADEJO Olayiwola,

CJAR
Accepted 15 January 2026
Published 30 January 2026
DOI: 10.5281/zenodo.18722362

Abstract:

Correlation analysis is one of the most frequently applied statistical techniques in educational research because it provides a straightforward way to describe the direction and strength of association between variables such as achievement, attitude, interest, retention, and other learning outcomes. This paper discusses the appropriate use of correlation coefficients in educational research, clarifying the concept of correlation and the meaning of the coefficient of correlation as a unit-free index ranging from −1 to +1. It explains how correlation is used to summarise relationships, support prediction, and guide interpretation of educational data, while emphasising that correlation does not establish cause-and-effect relationships. The paper also outlines key properties of correlation coefficients and differentiates simple (bivariate) correlation from multiple and partial correlations. In addition, it reviews major correlation techniques commonly used in education such as Pearson product moment, Spearman rank-order, Kendall's tau, point-biserial, and related approaches highlighting that method selection should align with measurement scale, distributional assumptions, and the nature of the relationship (linear or non-linear). Finally, the paper notes extended applications of correlation in educational measurement, including hypothesis testing, comparing correlations, and estimating reliability through test–retest and split-half procedures, thereby strengthening statistical decision-making in educational inquiry.

Keywords: Correlation coefficient; Pearson product–moment; Spearman rank-order; Educational measurement; Reliability estimation,

About Author

Author(s): ABE Thomas Olabode
Department of Counselling Psychology,
Bamidele Olumilua University of Education, Science and Technology,
Ikere-Ekiti, Ekiti State.

GBORE Lawrence Olu
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State.

OSAKUADE Joseph Oluwatayo
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State.

OWABUMOYE Ogunleye Sunday
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State.

AJELE Ayodeji Johnson
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State.

OGUNBADEJO Olayiwola
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State.

Introduction

Research is a formal and systematic application of the scientific methods to the study of education (Gay, 1996). It is a process by which knowledge are constantly questioned and by explaining those questions, individuals are able to change and improve understanding. Research in education is therefore, the continued search and research into issues of educational interest and new ways of improving, developing and changing our comprehension concerning educational issues. Academic research entails formal explicit research activities that occurs within academic institutions in the forms of projects, thesis or dissertation (Bright, 1991, Gay, 1996 & Adeyemi, 2007, Alonge, 2010, Abe, 2015, Gbore, 2019).

Abe (2004) and Bandele (2004) argue that educational research is a systematic approach to providing solution to educational problems. It involves the application of the scientific method in finding solutions to educational problems. Educational research is a type of systematic investigation that applies empirical methods to solving challenges in education. It adopts rigorous and well-defined scientific processes in order to gather and analyze data for problem solving and knowledge advancement. The primary purpose of educational research is to expand the existing body of knowledge by providing solutions to different problems in pedagogy while improving teaching and learning practices. Educational researchers also seek answers to questions bothering on learner motivation, development, and classroom management. (Alonge, 2010, Mukaku, 2012) posited that research is a cyclical process of steps that typically begins with identifying a research problem or issue of study. It then involves reviewing the literature, specifying a purpose for the study, collecting and analyzing data, and forming an interpretation of information. This process culminates in a report disseminated to audiences which is evaluated and used in the educational community.

Statistical Techniques in Educational Research

Data collection, analysis, interpretation, and presentation are all considered parts of the mathematical discipline known as statistics. Statistical analysis is utilized extensively in educational research by academic institutions, departments of natural and social sciences, governments, and educational organizations since it is crucial for decision-making (Abe, & Abe, 2002). To analyze unknowable facts, researchers can use inferential statistics. Inferential statistics allows researchers to draw conclusions or make claims about a large population from which samples of known data have been taken. A statistical technique called correlation is used to evaluate a possible linear link between two continuous variables and both the calculation and interpretation are easy (Mukaka, 2012).

Concept of Correlation

Correlation is a statistical tool which studies the relationship between two or more variables. It examines the degree {extent] of relationship existing between two variables. Normally researchers are interested in determining the extent of relationship that exists between two or more variables. For example, if a classroom teacher is interested in studying the relationship between achievement and retention, or achievement and interest, Obodo (2014) opined that the objective is to discover the nature of the relationship that exist between the variables, measure it and makes prediction about the value of one variable from value of the other[s]. When two variables are involved, it is simple [linear] correlation. When more than two variables are involved it is multiple correlations. Correlation analysis involve various method and techniques used for studying and measuring the extent of the relationship between two variables.

The correlation expresses rates between the groups of items but not between the individual items. The relationship between two variables is not functional that is to say; correlation analysis is a statistical procedure by which the degree of associate or relationship between two or more variables can be determined. Correlation analysis in research is a statistical method used to measure the strength of the linear relationship between two variables. That is, correlation analysis calculates the level of change in one variable due to change in other.

Coefficient of Correlation

Correlation can be defined as the tendency towards interrelation variation and the coefficient of correlation is a measure of such a tendency, that is, the degree to which the two variables are interrelated is measured by a coefficient which is called the coefficient of correlation. It gives the degree of correlation.

Therefore, from this it can be stated that the relationship between two or more variables such that a change in one variable results in corresponding greater or smaller change in other variable is known as correlation. The coefficient of correlation between the two variable x,y is generally donated by r or r_{xy} or p[x,y] or p.

Properties of Coefficient of Correlation

i. It is a measure of the closeness of a fit in a relative sense.
ii. Correlation coefficient lies between -1 and +1 i.e. $-1 \leq r \leq 1$.
iii. The correlation is perfect and positive if r = 1 and it is perfect and negative If r = - 1 if r = 0, then there is no correlation or relationship between the two variables and thus the variables are said to be independent.
iv. The correlation coefficient is a pure number and is not affected by a change of origin and scale.
v. It is a relative measure of association between two or more variables.

From the above, correlation coefficient is the unit measurement used to calculate the intensity in the linear relationship between the variables involved in a correlation analysis, this is easily identifiable since it is represented with the symbol r and is usually a value without units which is located between 1 and -1.

Positive correlation: i.e. r =1

This depicts a positive relationship between two variables which means both the variables move in the same direction. An increase in one variable leads to an increase in the other variable and vice versa.

Negative correlation i.e. r = -1

This indicates a negative correlation or relationship between two variables which means both the variables move in opposite directions. An increase in one variable leads to a decrease in their variable vice versa e.g. increasing the speed of vehicle decrease the time you take to reach your destination.

Weak/ Zero correlation: it implies that no relationship or correlation exists, when one variable does not affect the other. E.g, there is no correlation or relationship between the number of years of school a person has attended and the letter in his/her name.

Computation of the correlation coefficient

There are many methods devised for the calculation of correlation coefficients. The use of any particular method may depend on the type and number of variables involved and/or the form of data [whether interval, ordinal, nominal or ratio scale that was used] some of the method are as follows;

i. The Pearson's Product Moment Correlation coefficient [r]

ii. The Spearman's Rank – order coefficient [p]
iii. The Kendall's tall correlation coefficient [4]
iv. The Point- Biserial correlation coefficient [ø][phi]
v. The Fourhold or Phi correlation correlation[ø]
vi. The Tetrachoric Correlation Coefficient [r_t]
vii. The Partial Correlation Coefficient [$r_{xy\,x}$
viii. The Multiple Correlation Coefficient[$R^2_{x1\,x\,2}$]

The four commonly used in educational and social sciences research are:
Pearson Product Moment Correlation [r] method, Spearman Rank Order, Point Biserial Correlation and Kendal's Tau while in vocametrics, Partial and Multiple correlations are prevalently used.

Pearson Product Moment Correlation [Pearson r] method

This technique is based on some assumptions:

i. The samples from the two variables should be randomly drawn from the population.
ii. Both variables should be measured on the interval scale.
iii. The distribution of the two variables should have equal variability.
iv. The distribution of the two variables may need to be normal. They should be Unimodal.
v. A linear [rectilinear] relationship in the distributions of the two variables must be tenable – easy to defend computation of the Pearson r

Computation of Pearson r

There are four method of calculating Pearson r. They are;
[I] The Raw score method
[II] The deviation from the mean method
[iii] The standard deviation method
[iv] The Z-score method.

The Raw Score Method of Computing Pearson r

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

Where N is number of paired variables [subjects]
$\sum X$ = sum of the data in X distribution
$\sum Y$ = sum of the data in Y distribution
$\sum XY$ = sum of the product of X and Y
$\sum X^2$ = sum of the squares of X
$\sum Y^2$ = sum of the squares of Y

Example; [1] find the correlation coefficient between the scores of a student in mathematics [X] and statistics [Y] and comment.

Mathematics [X]	1	2	3	4	5	6	7	8	9	10
Statistical [Y]	3	5	5	2	4	6	7	10	8	12

Table 1: Showing solution to the example 1

Solution X	Y	X^2	Y^2	XY
1	3	1	9	3
2	5	4	25	10
3	5	9	25	15
4	2	16	4	8
5	4	25	16	20
6	6	36	36	36
7	7	49	49	49
8	10	64	100	80
9	8	81	64	72
10	12	100	144	120
54	62	368	472	413

$\Sigma X = 54$, $\Sigma Y = 62$, $\Sigma X^2 = 368$, $\Sigma Y^2 = 472$, $\Sigma XY = 413$

Now $r = \dfrac{N\Sigma XY - [\Sigma X][\Sigma y]}{\sqrt{N\Sigma X^2 - \{[\Sigma X]^2\} N\Sigma Y^2 - [\Sigma Y^2]\}}}$

$= \dfrac{(10)x(413) - (54)(62)}{\sqrt{[(10)(368) - (54)^2][10(472) - (62)^2]}}$

$= \dfrac{4130 - 3348}{\sqrt{(764)(426)}}$

$= \dfrac{782}{818.085}$

$r = 0.96$.-the relationship is very high and positive

Computation using Deviation From Mean Method.

$R = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$ --------- [2]

Where $X = X - \dot{X}$ and $Y = Y - \bar{Y}$

Example 2; calculate the coefficient of linear correlation between Mathematics [X] and physics [Y].

X	23	27	28	29	30	31	31	35	36	38
Y	18	22	23	24	25	26	28	29	32	33

Table 2 Showing the Computation of correlation coefficient by deviation from mean

X	Y	$X = X - \dot{X}$	$Y = Y - \bar{Y}$	$(XY) = (X - \dot{X})(Y - \bar{Y})$	$(X)^2 = (X - \dot{X})^2$	$(Y)^2 = (Y - \bar{Y})^2$
23	18	-8	-8	64	64	64
27	22	-4	-4	16	16	16
28	23	-3	-3	9	9	9
29	24	-2	-2	4	4	4
30	25	-1	-1	1	1	1

31	26	0	0	0	0	0
33	28	2	2	4	4	4
35	29	4	3	12	16	9
36	32	5	6	30	25	36
38	33	7	7	49	49	49
310	260			189	188	192

ΣX=310, $\dot{X} = \frac{\Sigma X}{N} = \frac{310}{10} = 31$, $\Sigma(X - \dot{X})^2 = 188$

ΣY= 260, $\bar{Y} = \frac{\Sigma X}{N} = \frac{260}{10} = 26$, $\Sigma(Y - \bar{Y})^2 = 192$

Now r =

$$\frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

Where X = X - Ẋ and Y = Y - Ȳ

$$r = \frac{189}{\sqrt{(188)(192)}} = 0.995$$

r is very high and positive computation using assumed mean method

Computation Using Assumed Mean Method

$$r = \frac{\Sigma dx\,dy - (\Sigma dx)(\Sigma dy)}{\sqrt{\Sigma dx1^2 - \frac{(\Sigma X)^2}{N}\left[\Sigma dx^2 - \frac{(\Sigma dy)^2}{N}\right]}} \quad \text{.........................(3)}$$

Where dx = x – Ax , y – Ay

Where Ax and Ay are corresponding pairs of assumed mean of x and y.

Example 3. Using the data in example 2 above

Table 3 Showing solution to the example 3

X	Y	Dx	Dy	dxdy	dx^2	dy^2
23	18	-7	-7	49	49	49
27	22	-3	-3	9	9	9
28	23	-2	-2	4	4	4
29	24	-1	-1	1	1	1
30	25	0	0	0	0	0
31	26	1	1	1	1	1
33	28	3	3	9	9	9
35	29	5	4	20	20	16
36	32	6	7	42	42	49
38	33	8	8	64	64	64
		10	10	199	198	202

$$r = \frac{\sum dxdy - \frac{\sum dx \sum dy}{N}}{\sqrt{[\sum dx^2 - (\frac{\sum dx}{N})^2][(\sum dy^2)^2 - (\frac{\sum dy}{N})^2]}}$$

$$= \frac{199 - (\frac{10 \times 10}{10})}{\sqrt{\left[198 - \frac{10^2}{10}\right][202 - (\frac{10}{10})^2]}}$$

$$= \frac{189}{(188)(192)}$$

$$= \frac{189}{189.989} = 0.995$$

Computation of Pearson r using standard score method

R= $\frac{\Sigma ZxZy}{N-1}$ …………… (4)

Where 2x = $\frac{x - Mean}{standarddeviation}$ =

= $\frac{x - Mx}{sdx}$

Where 2y = $\frac{y - Mean}{standarddeviation}$

= $\frac{y - My}{sdy}$

N = is the total number of observation [sample size].

Example 4. Find the correlation coefficient of the score of a student between measurement and evaluation [X] and statistics Y is given as .

X	30	15	5	20	20	25	30	25	20	10
Y	25	10	15	40	10	25	40	25	30	25

Table 4 showing the solution to example 4

X	Y	Z_x	Z_y	$Z_x Z_y$
30	25	1.29	0.65	0.06
15	10	-0.65	-1.43	0.93
5	15	-1.94	-0.94	1.82
20	40	0	1.53	0
20	10	0	-1.43	0
25	25	0.65	0.05	0.03
10	40	1.29		1.97
25	25	0.65	0.53	0.03
20	30	0	0.54	0
10	25	-1.29	0.05	-0.06
Σx=200 Mx=20	Σy = 245 My=24.5			ΣZxZy = 4.74

r= $\frac{\Sigma ZxZy}{N-1}$

r= $\frac{4.78}{10-1}$

r = 0.53

The relationship is moderate and positive coefficient of correlation can also be computed by first determining the coefficient of determination. This is the ratio of the explained variation to the total variation. The total variation of Y is defined as Σ [Y – Y]^2. This is the sum of the squares of the deviation of the values of Y from the mean Y and is given by Σ[Y – Y]^2 = Σ [Y – Yest]^2 + Σ[Yest – Y]^2-----[V]
While Σ[Y – Yest]^2 is called the unexplained variation the second time that is , Σ[Yest – Y]^2 is called the explained variation. If the explained variation is zero the ratio is zero meaning that the total variation is all explained, the ratio is one. This implies that the ratio lies between +1 and -1, since the ratio is always non-negative. It is denoted by r^2. This quantity is the coefficient of determination and is given by

$$r^2 = \frac{\text{explained –Variation}}{\text{Total variation}}$$

$$r = \frac{\Sigma(Yest-\bar{Y})}{\Sigma(Y-\bar{Y})^2}$$

The square root of the coefficient is the coefficient of variation.

Coefficient of determination = $r^2 = \frac{\pm\sqrt{explained\ Variation}}{Total\ Variation}$

$r=\pm\sqrt{eplainedvariation})$ or $\pm\sqrt{\frac{\Sigma(Yest-\bar{Y})}{\Sigma(Y-\bar{Y})^2}}$

Correlation by Ranks or Spearman Rank – Order correlation

Spearman Rank – Order correlation should not be used at all times since it is regarded as a non-parametric test statistic.

When to use Spearman Rank-Order correlation.

i. Use the spearman rank order correlation when the actual value of some observations are not available, known or collected rather such characteristic as good, excellent, poor, fair e.t.c are used in rating variables of individual.
ii. When ranks are assigned to individuals based on their characteristics. When ordinal data are involved. And, when data are of interval form , they can be converted to ordinal data in order to apply spearman rank correlation.
iii. When two set of ranks are compared to determine there degree of equivalence or relatedness.
iv. Spearman rank order correlation is also to be used when one is to compare judgment of two objects or the scores of a group of objects on two measures by a group of judges.
v. To assess inter-judge equivalence of judgment over a set of item or objects a spearman rank order correlation can be used

Also, to compare judgment by two judges of a group of object or items spearman rank order can be use[Uzoagulu, 1998]

Spearman Rank – Order Correlation Formula

The spearman rank-order correlation formula is given as ;

rs (or p) =

$$=1-\frac{6\Sigma d2}{N^3-N} \quad \text{or} \quad 1-\frac{6\Sigma d2}{N(N^2-1)} \quad (6)$$

Where rs = spearman rank-order correlation coefficient.

1 = unity is perfect correlation from which any value in the quality maybe taken to reduce the coefficient.
6 = this is a constant value
Σd^2 = the sum of the difference in ranks squared.
N = number of cases.
Afemileke&Onyemunwa [1997] argued that, in solving problem involving rank correlation, three type of problem may be encountered;
[1] Actual rank are given
[2] Actual rank are not given
[3] Problems of having equal rank.
Abe &Abe [2002] and Obodo [2014] argue that problems that have actual rank are simple and straight forward. Only three steps are involved;
[a] take the difference of the two ranks i.e R1- R2 and depict this as difference by D.
[b] square the difference and obtain the total ΣD^2.
[c] apply the formula rs = $1 - \frac{6\Sigma D^2}{N[N^2-1]}$

Example; 5
A class of 12 students was examined in general mathematics and further mathematics. The table below gives the order of merit of the student [arranged in alphabetical order] in both subjects.

G. Maths	1	2	3	4	5	6	7	8	9	10	11	12
F Maths	12	9	6	10	3	5	4	7	8	2	11	1

Calculate the coefficient of rank correlation what degree of agreement is there between the judges?
Table 5 showing the computation of Rank Correlation Coefficient
Computing Rank Correlation Coefficient

G Maths R1	F Maths R2	R1 – R2 = D	D2
1	12	-11	12 1
2	9	-7	49
3	6	-3	9
4	10	-6	36
5	3	2	4
6	5	1	9
7	4	3	1
8	7	1	1
9	8	1	64
10	2	8	0
11	11	0	121
12	1	11	ΣD = 416

$r = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$

$r = 1 - \frac{(6)(416)}{12(12^2-1)}$

r= 1- $\frac{2496}{1716}$

= -0.454 (Moderate and Negative)
When the actual ranks are not given, in this situation, it is necessary to assign ranks to them. Ranks can be assigned by taking either the higher value as 1. An important remark to be borne in mind is that it the same order should be adoptedin ranking two variables(Afemilehe& 0nyemuwa,(1997)
Example 6
Calculate the rank co efficient if correlation forthe following data

Maths X	92	89	87	86	77	71	63	53	57
Physics Y	86	82	91	77	85	52	82	37	57

Computing Rank correlation co efficient
Table 6 showing Computation of Rank Correlation Coefficient

X	Rx	Y	Ry	Rx-Ry	(Rx-Ry) = D
92	1	86	2	-1	1
89	2	83	4	2	4
87	3	91	1	2	4
86	4	77	6	-2	4
83	5	68	7	-2	4
77	6	85	3	3	9
71	7	52	9	-2	4
63	8	82	5	3	9
53	9	37	10	-1	1
57	10	57	8	2	4
					44 = D

r= 1- $\frac{6\Sigma D^2}{N(N^2 - 1)}$

r= 1- $\frac{(6)(44)}{10(10^2-1)}$

= 1-0.267 = 0.733
Comment The correlation coefficient is high and positive

For variables with Equal Ranks

$$rs = 1 - \frac{6(\Sigma D^2 - \frac{1}{12} - (m^3 - m) + \frac{1}{12}(m^3 - m)}{N(N^2 - 1)} \ldots\ldots\ldots\ldots\ldots (7)$$

Sometimes, it becomes necessary to rank two or more items or scores as equal. In such a case it is customary to give each individual an average rank. For example , if two items are ranked equal at 5th place, they are each given the rank $\frac{5+6}{2}$ = 5.5. Also if three items are rank equal at 5th place, they are given rank $\frac{5+6+7}{3}$ = 6

When two or more items are to be ranked equal, determine the average of the ranks and add half 1/2[m^3 – n] to the value of ΣD^2, where m stands for the number of item whose rank are common. If there are more than one such group of item with common ranks , this value Is added as many time as the number of such groups . The equation 7 above simplify the formula.

Example 7 below are the score obtained by 8 students in statistics and research method.

Statistics [x]	15	20	28	12	40	60	20	80
Research method[y]	40	30	50	30	20	20	30	60

Table 7 showing Computation of rank correlation with ties

Statistics x	R1	Research Y	R2	R1-R2=d	(R1-R2)
15	7	40	3	4	16
20	5.5	30	5	0.5	0.25
28	4	50	2	2	4
12	8	30	5	3	9
40	3	20	7	-4	16
60	2	10	8	-3	36
20	5.5	30	5	0.5	0.25
80	1	60	1	0	0
					81.5

$$rs = 1 - \frac{6\ (\ \sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) +}{N(N^2 - 1)}$$

$\sum D^2 = 81.5$
N = 8
In variable X, m = 2
In variable Y, m = 3
Substituting in equation 7

$$rs = 1 - \frac{(81.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^2 - 3) +}{8(8^2) - 1)}$$

$$= 1 - \frac{6(81.5 + 0.5 + 2)}{504}$$

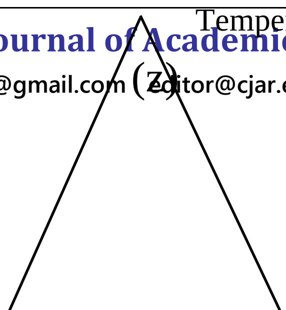
$$= 1 - \frac{504}{504}$$

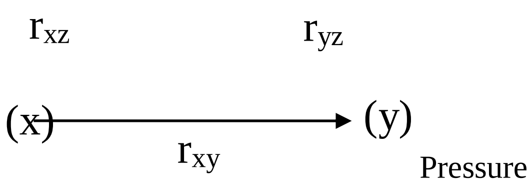
$$= 1 - 1 = 0$$

There is no relationship between the performance of student in statistics and research method.
Note however, this formula or method cannot be used in determining the correlation of external frequency distribution . While the formula is more appropriate in finding the relationship or correlation of data that are qualitative in nature such as honesty, beauty and intelligence [Afemileke & Onyemunwa, 1997] .

Partial Correlation Coefficient

This is referring to the functional relationship which exists between any two variables when all other variables connected with the two are kept constant. For example,





volume
Relationship between x and y when z is held constant. The above shows if the researcher is interested in the correlation or relationship of the volume [x] and pressure of gas. There is need to control the influence of temperature [z].
Hence, the partial correlation coefficient formula is given as

$$r_{(xy).z} = \frac{r_{xy} - (r_{xz} * r_{yz})}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}} \text{..........} \qquad (8)$$

Similarly the formula can be restructured to obtain the partial correlation coefficient of XZ when y is constant and YZ when X is constant as follows

I. $$r_{(xz).y} = \frac{r_{xz} - (r_{xy} * r_{zy})}{\sqrt{(1-r_{xy}^2)(1-r_{zy}^2)}} \text{..........} \qquad (9)$$

$$r_{(AB)C} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{(1-r_{AC})(1-r_{BC})}}$$

$$r_{(AB)C} = \frac{0.78 - (0.59)(0.02)}{\sqrt{(1-0.59^2)(1-0.002^2)}}$$

$$r_{(AB)C} = \frac{0.768}{\sqrt{0.652}}$$

$r_{(AB)C}$ = 0.951 The partial correlation coefficient of (MTH111)A and MTH112 (B) when MTH113 C is held constant is very strong and positive.

II. $$r_{(AB)C} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{(1-r_{AB}^2)(1-r_{AC}^2)}}$$

$$r_{(AB)C} = \frac{0.59 - (0.78)(0.02)}{\sqrt{(1-0.78^2)(1-0.02^2)}}$$

$$r_{(AB)C} = \frac{0.574}{\sqrt{0.391}} = 0.918$$

III. $$r_{(AB)C} = \frac{r_{BC} - (r_{AB})(r_{AC})}{\sqrt{(1-r_{AB}^2)(1-r_{AC}^2)}}$$

$$= \frac{0.02 - (0.78)(0.59)}{(1-0.78^2)(0.059^2)}$$

$$= \frac{-0.440}{\sqrt{o.255}} = 0.871$$

The partial correlation between B and C without the influence of A is negative and very strong.

$$r_{(yz).x} = \frac{r_{xy} - (r_{yz} * r_{zx})}{\sqrt{(1-r_{yx}^2)(1-r_{zx}^2)}} \text{..........} \qquad (10)$$

Where r [XY] Z is partial correlation between X and Y when variable Z is kept constant.
r[XZ]y is partial correlation between X and Z when variable Y is kept constant
r[XZ]x is partial correlation coefficient between Y and Z when variable X is held constant.
rxy is correlation between X and Y

rxz is correlation between X and Y
rzy is correlation between Z and Y.
Example, the correlation in between the variables A B C i.e. the group of health science student in BOUESTI performance in MTH 111, MTH 112 and MTH 113 is given as A,B,C
The below illustrated the correlation matrix between MTH(iii)(A), MTH 112 (B) and MTH 1113(c).

	A	B	C
A	1.00	0.78	0.59
B		1.00	0.02
C			1.00

∫∫alculate (i) $r_{AB.C}$ (ii) $r_{AC.B}$and $r_{BC.A.}$

(i) r(AB) C = rAB – rAC. rBC
∫(1- rAC) (1-rBC)
r(BC). C = 0.78-(o.59) (o.o2)
∫(1-0.59^2) (1-0.02^2)
r(AB). C = 0.768
∫0.652
rAB>C = 0.951.
The partial correlation coefficient of (MTH 111) A and MTH 112 (B) when MTH 113 C is held constant is very strong and position.

(ii) rAC(B) = r(AB) (rAC)
∫(1-r^2AB) (1-r^2AC)
rAC(B) = 0.59 – (0.78) (0.02)
∫(1-0.78^2) (1-0.02^2)
rAC(B) = 0.574
∫0.391 = 0.918

(iii) rBC(A) =rBC – (rAB) (rAC)
∫ (1-r^2AB) (1-r^2AC)

=0.02 – (0.78) (0.59)
∫(1-0.78^2) (0-0.59^2)

= -0.440
∫0.255 = -0.871
The partial correlation between B and C without the influence of A is negative and very strong.

Kendall's Tau τ
Most correlation coefficient employ the product-moment principle of Pearson in one form or the other. Maurice Kendall, an English statistician made effort to compute the measurement of relations between variable using a different principle other than the product-moment principle. Kendall based his calculation of correlation coefficient on the number of pairs that are arranged in an ordered form in the same direction on both variable X and Y. The Kendall Tau method which denoted τ indicates the extent of disagreement in the rankings of X and Y.
Assuming that eight persons are ranked on X and Y thus:

Table 8 Calculating kendall'sTau,r

	(1 is the Highest rank)	
Person	X	Y
A	1	3
C	2	1
B	3	2
H	4	5
E	5	7
F	6	8
D	7	4
G	8	6

Looking closely at the ranks, for example, A is higher than B on variable X, while B is ranked higher than A on variable Y. This relationship is far from a direct relationship. The relationship between A and B on the table illustrates an inverse relationship between X and Y, for A and B in particular. In a situation like this we count the number of agreements and inversions in the ranking. Examining closely the ranks of persons A and H, there seems to be an agreement because A ranks higher than H on X [1 against 4], and on Y [3 against 5]. Persons A and H contribute an agreement, while person A and B an inversion as well as F and D which also contribute another inversion, for F has a higher rank than D on X and a lower rank than D on Y .

Where there are no tied ranks, all n (n-1)/2 pairs of persons will contribute either an agreement or an inversion. On the other hand, when the rankings on both X and Y are identical, there will be n (n-1)/2 agreement and no inversions. Kendall therefore defined his coefficient τ thus:

$$\frac{(total\ number\ of\ agreements) - (total\ number\ of\ inversions)}{n(n-1)/2}$$

He later called the total number of agreements P and the total number of inversions Q, and the formula becomes:

$$r = \frac{P - Q}{n(n-1)/2}$$

Kendall went further to simplify the formula by denoting P-Q with a symbol S, thus making the formula to be written:

$$r = \frac{S}{n(n-1)/2}$$

Another formula used for the calculation of Kendall τ which is exactly equivalent to the above formula is:

$$r = \frac{4p}{n(n-1)} - 1$$

Going back to our ranks in Table 3.5, let us calculate the number of agreement and inversions.

Table 9:Showing comparison of Agreements and Inversions in Kendall Tau, τ

Person	X	Y	Agreements	Inversions
A	1	3	5	2
C	2	1	6	0

B H E F D G	3 4 5 6 7 8	2 5 7 8 4 6	5 3 1 0 1 0	0 1 2 2 0 0
			P = 21	Q = 7

Cursory look at Table 9 above, the n persons are ordered 1-8 on variable X. To calculate the number of agreement, starting with the first person, we count the number of times his rank on Y is smaller in magnitude than the rank below it. For example, the top person has a score of 3 on Y, 3 is lesser than the following scores: 5, 7,8 ,4 and 6 making 5 scores. Therefore we record 5 has the number of agreement for A. Going down to the next person C, C has a score of 1 on Y and I is smaller in magnitude than the ranks below it -2. 5, 7, 8, 4, 6. We therefore record 6 as number of agreements for C. This is done for all the number person.

In order to compute the inversion, we count number of times a score on Y is greater than a score below it on column Y. For instance, for the first person A, his score on Y is 3, and the number is -3, is only greater than 1, and 2 out of all the scores below it. Therefore we record 2 for A under inventions. The next person is C and he scored 1 on column Y. Looking at all the other scores under 1, 1 is not greater than any of the scores, therefore we record Zero for C under inversions. This computation is carried out for all persons (that is in persons).

The total of the "agreements" column (P) and the total of the inversions column (Q) is calculated. Using kendall ….. formula, the correlation coefficient will be :

$$\tau = \frac{P - Q}{n(n-1)/2}$$

$$\tau = \frac{21 - 7}{8(8-1)/2} = \frac{14}{8(7)/2}$$

$\tau = \frac{14}{28} = 0.50$r is positive and moderate

Table 10 below is showing, the raw scores of 10 Senior Secondary School students on tests in English language and English Literature are given below:

Student	English Language	English Literature
A	50	47
B	59	52
C	54	35
D	52	50
E	34	29
F	58	49
G	46	39
H	35	24
I	55	54
J	38	26

In this situation, the first thing to be done is to convert the raw scores into ranks, using variable X and Y to indicate scores in English Language and English Literature thus:

Table 11showing calculation of Kendall Y(1 is the Highest Rank)

Student	X	Y

B	1	2
F	2	4
I	3	1
C	4	7
D	5	3
A	6	5
G	7	6
J	8	9
H	9	10
E	10	8

Suppose we select any pair of student's e.g. B and I, B is ranked higher than 1 on variable X, but I is ranked higher than B on variable Y. This indicates a departure from a direct relationship between X and Y, as B and I are concerned. On the other some students contribute an agreement. For example, B is ranked higher than A on X (1 vs 6) and also ranked higher on Y (2 vs 5). In this case we calculate the number of agreements and inversions. We had earlier indicated that to get the number of agreements, starting with the first student, we count the number of times his rank on Y is smaller in magnitude than the ranks below it. The product is entered under the column agreements. For the inversion, we calculate the number of times a score on Y is greater than a score below it in column Y, and also enter the product under the column inversion. For example, looking at Table 11, the total agreement for student B is 8 because B is ranked 1 on X and his rank on Y is 2, therefore the number 2 is smaller in magnitude than 4,7,3,5,6,9,10 and 8, giving us 8 for agreements. For the inversions, B will have 1, because B is ranked 2 on Y, and number of times is Y score is greater than a Y score below it in the column is 1.2 is only greater than 1 amongst the scores 4,1,7,3,5,6,,9,10,8 in that column. This computation is done for every member of the group, thus:

Table 12 showing computation Computing the Number of Agreements and Inversion for Kendall τ

Person	X	Y	Agreements	Inversions
B	1	2	8	1
F	2	4	6	2
I	3	1	7	0
C	4	7	3	3
D	5	3	5	0
A	6	5	4	0
G	7	6	3	0
J	8	9	1	1
H	9	10	0	0
E	10	8	0	0
			P = 37	Q = 7

Using our formula $\tau = \frac{P-O}{n(n-1)/2}$

Where P = 37

Q = 7

n = 10

$$\tau = \frac{37-7}{10(10-1)/2} = \frac{30}{45}$$

= 0.67 r is positive and high

Interpretation of correlation coefficient

The correlation coefficient obtained from using any of the methods discussed in this chapter has the same meaning. It is a measure of linear relationship that exists between two variables. A correlation coefficient can take any value from -1 to +1, and can be interpreted as follow:

Table 12 showing the interpretation of correlation coefficient values.

Correlation Coefficient Value	Interpretation of Correlation for +ve values	Interpretation of Correlation for –ve values
1.00 0.8 to 1.0 0.6 to 0.8 0.4 to 0.6 0.2 to o.4 0	Perfect positive Very high positive High positive Moderate positive Low positive Very low positive No correlation	Prefect negative Very high negative High negative Moderate negative Low negative Very low negative No correlation

Note The real mid points of the values can be used to separate neighbouring classes when the values are in three decimal places, like 0-0.199,0.2-0.399,0.4-0.599,0.6-0.799,0.8-0.999, and +1 for perfect correlation

Test of significance of a correlation coefficient

The statistical significance of a correlation coefficient can be tested using either of the two formulas discussed below, depending on the size of the sample observed.

For a small sample where N<100, a correlation coefficient can be tested for significance using the formula:

$$t_c = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

r is the correlation coefficient being tested

Example: In a study of the relationship between attitude towards school and achievement, the following results were obtained.

	Boys	Girls(2)
N	25	29
R	0.45	0.39

Are the correlation coefficients statistically different (α=0.5)

Soln = $H_o = P_1 = P_2$

$H_1 - P_1 = P_2$

Given that α = 0.05 and a two tailed test critical region -1.96 ≤Z_o or Z_o≥1.96

Computation Boys (1) N = 25, r1 = 0.45

Z_{r1} = 0.465

From $Z_{r1} = \frac{1}{2} Log_e (1 + 0.45) - \frac{1}{2} Log_e (1 - 0.45) = 0.485$

Girls (2) = N = 29_1 r_2 = 0.39, Z_{r2} = 0.412

From $Z_{r2} = \frac{1}{2} Log_e (1 + 0.39) - (\frac{1}{2} Log_e (1- 0.39)$

Z_{r2} = 0.412

$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{N_{1-2}} + \frac{1}{N_{2-3}}}}$

$= \frac{0.485 - 0.412}{\sqrt{\frac{1}{25-1} + \frac{1}{29-3}}} = 0.252$

$= Z_{tab} = 1$

.. $Z_{cal} < Z_{tab}$

$0.252 < Z_{tab}$

$0.252 < 1.96$ H_o is not statistically different.

N is the number of observations (sample size)

(N-2) is the degree of freedom.

Test of Difference between two correlations

$$Z_r = \frac{1}{2} log\, log\, e\,(1 + r) - Log\, e\,(1 - r)$$

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{N, -3} + \frac{1}{N_2 - 3}}}$$

Where Zr1 is transformed value of r1 and Z_{r1} i transformed value of r1 N1 and N2 are the number of subject/ people in group 1 and 2 respectively. And Zr = $\frac{1}{2}Log_2(1 + r) - \frac{1}{2}Log_2(1 - r)$

Using correlation co efficient as descriptive statistics in Research

As a measure of strength of relationship and direction

What is the strength of relationship among the school Based Assessment scores in Boys, Girls and Co- educational schools

Table 14 Showing the strength of relationship between SBA1, SBA2, SBA3 in Boys, Girls and Co-Educational schools

School Type	SBA1 & SBA2	SBA1 & SBA 3	SBA 2 & SBA 3
Boys	0.33 low & positive	0.25 low & positive	0.43 moderated positive
Girls	0.42 moderate & positive	0.64 high & positive	0.64 high& positive
Co Educational	0.43 moderate &positive	0.40 moderate & positive	0.61 high & positive

Source, Abc (200q)

The table shows that low and positive relationship existed between (SBA1 and (SBA2) and (SBA1 and SBA3) in Boys only school, moderate & positive relationship existed between (SBA2 & SBA3) Boys only School, (SBA1 &SBA 2) Girls only school, and (SBA1 & SBA2) and (SBA 1 & SBA 3) in Co educational schools while High and Moderate and positive relationship existed between (SBA 1 & SBA3) and (SBA2 and SBA3) in Girls only schools and also between SBA2 and SBA3 in college of Educational schools.

Using correlation coefficients as Hypothesis Testing in educational Research

For examples from the above descriptive analysis in the table 4 above: This hypothesis was formulated thus;

There is no significant relationship among the school Based Assessment scores (SBA1, SBA2, & SBA3) in Boys, Girls and Co-Educational schools.

This Table15 showing the correlation analysis of school- based assessment scores (SBA1,SBA2,&SBA3) in school types

School Type	SBA1 &SBA2	SBA1 & SBA3	SBA2 & SBA 3
Boys	0.33	0.25	0.43
Girls	0.42	0.64	0.64
Co- Educational	0.43	0.40	0.61

Critical value of 0.36 at $P < 0.52$
*Significant.

Using Correlation Coefficient as a reliability estimator:

-- The this could be done using reliability estimator such as shown below:
[1]Test – retest method [the measure of stability].
To apply this method takes the following steps:
--- Prepare a test [or other instrument such as questionnaire].
--- Define the group that will be the test takers [i. e population or sample].
--- Administer the test to the group.
--- Allow an intervening time period [e. g two weeks, one month e.t.c.]
--- Administer the same test to the same group who took the test previously.
--- Obtain two set of test scores or those arising from questionnaire, observation or interview e.t.c

From the table 15 above $p < 0.05$ depicts significant relationship existed between (SBA_{2+} and SBA_3) in boys only schools, (SBA_1 &SBA_3) and (SBA_2 & SBA_3) in both girls and co-education schools this led to the upholding of the hypothesis while the hypothesis was upheld between. (SBA_1& SBA_2) and (SBA_2 & SBA_3) in boys schools which implies that there was no significant relationship among the school-based assessment scored in (SBA_1& SBA_2 and SBA_1 & SBA_3) at $P<0.05$ level of tolerable limits of error or significant level.

--- correlate the two set of scores using appropriate statistical tool such as Pearson Product Moment Correlation or Spearman Rank – Order Correlation e.t.c.

From your analysis obtained correlation coefficient which should be the 8, coefficient of reliability or stability. The correlation between the two scores of the two administration of the same test is a measure of the reliability of the instrument.

[ii] **Using correlation coefficient to estimate Reliability by split - halves Method.**

To apply this method takes the following steps:
--- Prepare a test and administer the single test once on sample respondents
--- Split the test into two, the first odd numbered, the other even numbered
--- Be sure the item of the halves are equally matched on content and difficulty index o one single test.
--- Score each half independently of the other
--- Obtain two sets of scores
--- The scores of the two halves should be correlated using any correctional technique such as spearman rank order, Pearson Product Moment Correlation or Kendall rank correlation coefficient.

[iii] The correlation coefficient obtained is the reliability coefficient of the half- tests and is denoted by 8

--- Apply Spearman- Brown step-up [or prophency] formula to estimate the reliability of the whole test. The formula is as follows:

$$r_w = \frac{2r1/2\ 1/2}{1 + r_{\frac{1}{2}1/2}}$$

Whole r_W= reliability coefficient of the whole test.

r = correlation coefficient of the two split halves

This is measure of internal constituency of the test or instrument.

--- use to determine the inter- item correlation coefficient by applying the Cronbach alpha with the formula.

$$\alpha = \frac{k\bar{r}}{[1+\bar{r}(k-1)]}$$

Where k = number of item in the test

$\bar{r}$ = the mean inter-item correlation. [Uzoagulu1998]

[iv] The ȣ can obtained by summing all the coefficient values of the correlation values of the correlation between all the items.

The mean inter – item correlation $\bar{r}$ can be obtained by summing up all these correlations and dividing the sum by 28 (The number of correlations coefficients) the r for the correlation matrix can be computed as follows: 0.1227 + 0.1257 + 0.547 – 0.472 + 0.001 + 0.0084 + 0.881 – 0.096 + 0.258 – 0.414 – 0.589 +m0,305 + 0.348 + 0.578 – 0.546 – 0.377 + 0.213 – 0.013 + 0.347 + 0.511 – 0.505 + 0.218 + 0.359 + 0.587 – 0.181 + 0.242 – 0.011 – 0.591.

$$\bar{r} = \frac{2.664}{28} = 0.095$$

The depicts the mean inter-item correlation for the ITAT which indicates a low and positive inter-item correlation coefficient. The inter-item correlation can be found using the formula.

$$\alpha = \frac{k\bar{r}}{1+\bar{r}(k-1)} \quad \text{where K = 130}$$

$$= \frac{130 \times 0.0951}{1+0.0951\,(139-1)} = 0.91$$

α = 0.91 This is an excellent reliability coefficient.

For example, the sub- test of the Introductory technology achievement test [ITAT] had a correlation matrix as shown in the table below.

Table 16 showing the Correlation Matrix of the ITAT and its sub-tests

Content Items	MW	EE	WW	BD	TD	FT	PR	CR
Mental work MW	1.00							
Electrical Election EE	0.1227	1.00						
Wood work WW	0.1257	0.547	1.000					
Building BD	-0.4712	-0.0009	0.0084	1.000				
Technical Drawing	0.8812	-0.0964	0.2578	-0.4143	1.000			
Food Technolog y FF	-0.5888	0.3049	0.3480	o.5776	-0.5463			
Plastic & Rubber PR	-0.3767	0.2627	-0.0129	0.3473	0.5111	0.5050	1.000	
Ceramic CR	0.2154	0.2585	0.5865	-0.1805	0.2424	0.0110 9	-o.5932	1.000

The mean….

$\bar{r} = \frac{2.664}{28}$ = 0.0951 K = 130

$$\alpha = \frac{K\bar{r}}{1+\bar{r}(K-1)} = \frac{130 * 0.0951}{1+0.0951\,(130-1)}$$

$$\alpha = 0.93$$

From the table 1 above at p< r. 05, * depicts significant relationship existed between: [SBA2 and SBA3]

[SBA1 and SBA2] in boys only schools

[SBA1 and SBA3] and [SBA2 and SBA3] in both Girls and co- education schools this led to the upholding of the hypothesis while the hypothesis was uphold between [SBA1 and SBA2] and [SBA and SBA2] in boys schools which implies that, there was no significant relationship among the school based assessment scores in [SBA1 and SBA2 and SBA1 and SBA3] at $p < 0.05$ level of tolerable limits of error or significant level.

Conclusion

This paper explains and demonstrates theoretically with illustrative examples the concept of correlation coefficient based on various methods of calculating it, in education research, its usage in answering general questions, testing of hypothesis, estimation of reliability co efficient and the use of correlation matrix to determine inter- items correlation for the reliability estimate were extensively trashed in the teaching of correlation for both undergraduate and post graduate students. Though, the use of SPSS has demystified the rigorous computations involves but this paper would assist the beginners researches devoid of good back group in statistics to effectively teach the educational research and statistics with ease.

References

Abe, J. B., & Abe, T. O. (2002). *Statistics made easy* (2nd ed.). Bubean Information Services.

Abe, T. O. (2004). *Critique of what to do and how to do it in project writing*. Government Educational Publisher.

Abe, T. O. (2005). Statistical techniques for estimating sample size in educational research. *Journal of Education, 1*, 68–76.

Abe, T. O. (2009). Correlation of mathematics school-based assessment scores in school type and location. *Journal of Education*.

Adeyemi, T. O. (2007). *Research method and thesis writing in educational studies*. New Neven Publishers.

Afemikhe, O. A., & Onyenunwa, S. G. (1977). *An introduction to statistics*. AMFITOP Books.

Alonge, M. F. (2001). *Essentials of research methods and designs for educators*. Green Line Publishers.

Arora, F. N., Arora, S., Arora, S., & Arora, A. (2012). *Comprehensive statistical methods* (4th ed.). S. Chand & Company.

Bamidele, S. O. (2004). *Educational research in perspectives* (1st ed.). Niyi Commercial & Printing Ventures.

Bright, B. P. (1991). *Introduction to research methods in postgraduate theses and dissertations* (Newland Papers No. 18). School of Education, University of Hull.

Gay, L. G. (1996). *Educational research: Competencies for analysis and applications* (5th ed.). Prentice Hall.

Gbore, L. O. (2019). Beginning researchers and keys to valid and reliable research results. *Journal of Measurement and Evaluation in Education and Humanities, 2*(1), 1–13.

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal, 24*(3), 49–54.

Nworgu, B. G. (1991). *Educational research: Issues and methodology* (1st ed.). University Trust Publishers.

Nworgu, B. G. (2006). *Educational research: Issues and methodology* (2nd ed.). University Trust Publishers.

Obodo, G. C. (2014). *Ordinary and advanced level statistics for undergraduates and postgraduates*. Celex Printers.

Ofo, J. E. (1994). *Research methods and statistics in education and social sciences*. Joja Educational Research.

Olayinka. (2006). *Definition, spectrum and types of research in methodology of basic and applied research* (2nd ed.). Postgraduate School, University of Ibadan.

Uzoagulu, A. E. (1998). *Practical guide to writing research project reports in tertiary institutions*. John Jacobs Classic Publishers.

Cite this article:

Author(s), ABE Thomas Olabode, GBORE Lawrence Olu, OSAKUADE Joseph Oluwatayo, OWABUMOYE Ogunleye Sunday, AJELE Ayodeji Johnson, OGUNBADEJO Olayiwola, (2026). “Appropriate Use of Correlation Coefficient in Educational Research”, *Name of the Journal:* Commonwealth Journal of Academic Research, (CJAR.EU), P, 1 - 23. DOI: http://doi.org/10.5281/zenodo.18722362 , Issue: **1,** Vol.: **7,** Article: **1,** Month: **January,** Year: **2026.**

Retrieved from https://www.cjar.eu/all-issues/

Published by

AND

ThoughtWares Consulting & Multi Services International (TWCMSI)