UDC: 519.235/.237
Professional paper

DIFFERENCES BETWEEN THE CORRELATION COEFFICIENTS PEARSON, KENDALL AND SPEARMAN

Mirlinda SHAQIRI[1], Teuta ILJAZI[2], Lazim KAMBERI[1], Rushadije RAMANI - HALILI[1]

[1]Faculty of Natural Sciences and Mathematics, University of Tetovo, Ilinden n.n., 1200 Tetovo, Republic of North Macedonia,
[2]Faculty of Pedagogy, University of Tetovo
**Corresponding author e-mail: mirlinda.selami@unite.edu.mk*

Abstract

This study aims to compare the values of correlation coefficients such as the Pearson coefficient, and the Kendall and Spearman coefficient. Pearson's coefficient values are taken from only studies conducted on the correlation of variables to be found in this study, while Kendall and Spearman's coefficient values account for the same variables and specifications to carry out this study. The earned values are compared to see which coefficients have the closest values. After comparing the coefficient values it is concluded that the Pearson and Spearman coefficient have more close values than Kendall's coefficient.

Keywords: Correlation, nonparametric coefficients, Pearson coefficient, Kendall coefficient, Spearman coefficient, variable

Introduction

Statistics is a branch of mathematics, which deals with the collection, analysis, interpretation and presentation of data [1]. In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables. Although in the broadest sense, "correlation" can indicate any kind of association, in statistics it usually refers to the extent to which a pair of variables are linearly related. In essence, correlation is the measure of how two or more variables are related to each other [2]. There are several correlation coefficients, which measure the degree of correlations. The most common of all correlation is the Pearson correlation coefficient, which is only sensitive to a linear relationship between two variables (which may be present even when one variable is a nonlinear function of the other). Other correlation coefficients — like Kendall and Spearman's correlation — developed to be more powerful than Pearson's, thus more sensitive to nonlinear relationship[9]. As for the strength of the relationship, the correlation coefficient value varies between +1 and -1. A value of ±1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. The direction of the relationship is indicated by the sign of the coefficient: a "+" sign indicates a positive relationship and a "–" sign indicates a negative relationship[2].

Research Methodology

The idea of the research is to compare the correlation coefficient values of Pearson, Kendall and Spearman for the same variables. Research methodology is a comparative research model, where data on the Pearson coefficient is obtained from the work performed:

1. T. Iljazi (2021)The impact of professor's personality on teaching mathematics, *Journal of Educational Research*, https://eprints.unite.edu.mk/850/
2. T. Iljazi (2022)Teachers opinion on mathematics school books, *Knowledge-International Journal*, http://ikm.mk/ojs/index.php/kij/issue/view/154
3. T. Iljazi (2021)Students opinion about online teaching during the covid-19 pandemic, *Centrum*, https://www.ceeol.com/search/article-detail?id=1022976

against the values of Spearman and Kendall's correlation coefficient, calculated specifically for the same variables. The analysis of the results was performed via a table on which the values of the Pearson, Kendall and Spearman coefficients are presented.

Data Analysis

One of the most commonly used correlation statistics is the Pearson correlation coefficient (r) also called the correlation coefficient. The Pearson correlation describes the strength and direction of the linear relationship between two quantitative variables. It takes values between -1 and +1, indicating a perfect linear relationship between two variables, while the value 0 indicates no linear relationship[8]. This statistic is calculated using an equation that relates two sets of scores for two different measures. The equation results in a single number called the correlation coefficient and designated by the letter "r" [3]. The formula of the Pearson correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right]}} \quad (1)$$

In the (1) formula, x is the independent variable, y is the dependent variable, n is the sample size, and $\bar{x}, \bar{y}$ are the average of x, y respectively [8].

A lesser used coefficient is the Kendall coefficient or Kendall's τ coefficient, Kendall's tau, introduced by Kendall (1938), is a correlation coefficient that can be used as an alternative to Spearman's rho for data in the form of ranks. It is a simple function of the minimum number of neighbor swaps needed to produce one order from another[5]. The formula to calculate Kendall's Tau, often abbreviated τ, is as follows:

$$\tau = \frac{C-D}{\frac{1}{2}n(n-1)} \quad (2)$$

where C is the sum of the number of the largest values y and D is the sum of the number of the smallest values of y. The sums C and D represent the numbers of the pairs that are both incompatible and irreconcilable[8].

Spearman's rank coefficient is a nonparametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function .The formula by which Spearman's coefficient is recognized is

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(R_i-S_i)^2}{n(n^2-1)} \quad (3)$$

where R_i represents the order of the values of x_i, while S_i represents the order of the values of y_i from the smallest to the largest [9].

Below we have the tables given for Kendall's and Spearman's coefficient.

Table 1. Kendall's and Spearman's coefficient for variables that influence mathematics teaching

			MOTIVATION	Our opinions and suggestions on teaching issues	Creative and always use different approaches to teaching	Clearly/ Unclear
Kendall's tau_b	MOTIVATION	Correlation Coefficient	1	.609**	.506**	.514**
		Sig. (2-tailed)	.	0	0	0
		N	220	220	220	220
Spearman's rho	MOTIVATION	Correlation Coefficient	1	.683**	.584**	.576**
		Sig. (2-tailed)	.	0	0	0
		N	220	220	220	220

**. Correlation is significant at the 0.01 level (2-tailed).

From the table we notice that we have a medium correlation between motivation and opinions and suggestions on learning issues, creativity and use in different teaching approaches as well as clear/unclear learning.

Table 2. Kendall and Spearman's coefficient for variables regarding book accuracy

			Accuracy of the book	Objective purpose	The language of the book	The order of the book	Attractive illustrations
Kendall's tau_b	Accuracy of the book	Correlation Coefficient	1	.435**	.398**	.549**	.449**
		Sig. (2-tailed)	.	0.005	0.282	0.034	0.079
		N	52	52	52	52	52
Spearman's rho	Accuracy of the book	Correlation Coefficient	1	.476**	.443**	.594**	.495**
		Sig. (2-tailed)	.	0.005	0.283	0.034	0.087
		N	52	52	52	52	52

***. Correlation is significant at the 0.01 level (2-tailed).*

From the table we notice that we have a weak relationship between the accuracy of the book and the objective purpose, the language of the book and the attractive illustrations, and a medium relationship between the accuracy of the book and the order of the book.

Table 3. Kendall and Spearman's coefficient for variables for online course learning during Covid-19

			Online course	Verbal lectures	Sufficient means	Motivation
Kendall's tau_b	Online course	Correlation Coefficient	1	.355*	0.237	-0.222
		Sig. (2-tailed)	.	0.022	0.115	0.142
		N	37	37	37	37
Spearman's rho	Online course	Correlation Coefficient	1	.381*	0.263	-0.245
		Sig. (2-tailed)	.	0.02	0.116	0.145
		N	37	37	37	37

***. Correlation is significant at the 0.01 level (2-tailed).*

From the table we notice that the correlation between the online course and the verbal lecture is weak, the correlation between the online course and sufficient means is very weak. Whereas the correlation between the online course and motivation is a weak negative relationship, which means that as one variable increases, the other decreases and vice versa.

Table 4. Comparison of Pearson's, Kendall's and Spearman's coefficients for the same variables

	Variable Correlation	Pearson Coefficient	Kendall's Coefficient	Spearman's Coefficient	Difference between Pearson and Kendall's coefficient	Difference between Pearson and Spearman's coefficient
The impact of professor's teaching personality on mathematics	Motivation and our opinions and suggestions on teaching issues	0,652 [5]	0,609	0,683	0,043	-0,031
	Motivation and creativity and always using different approaches to teaching	0.605 [5]	0,506	0,584	0,099	0,021
	Motivation and Clearly/ Unclear	0,572 [5]	0,514	0,576	0,058	-0.004
Teachers opinion on	Accuracy of the book and	0.539 [6]	0,435	0,476	0,104	0,063

	objective purpose					
	Accuracy of the book and the language of the book	0,456 [6]	0,398	0,443	0,058	0,013
	The language of the book and the order of the book	0,621 [6]	0,549	0,594	0,072	0,027
	The order of the book and attractive illustrations	0,499 [6]	0,449	0,495	0,05	0.004
Students opinion about online teaching during the Covid-19 pandemic	Online course and verbal lectures	0,400 [7]	0,355	0,381	0,045	0,019
	Online course and Sufficient means	0,264 [7]	0.237	0,263	0,027	0,001
	Online course and motivation	-0.221 [7]	-0.222	-0.245	0,001	0,024

From the table above, we notice that the difference between Pearson's and Spearman's coefficients is smaller, unlike the difference between Pearson's and Kendall's coefficients.

Conclusion

Pearson's correlation coefficient is used to measure the degree of direct relationship between two continuous variables. In the case when the distribution of variables is normal or close to normal, the Pearson coefficient is used, while in the case when the distribution of variables is far from normal, the ordinal coefficient of Spearman is used. From Table 1, Table 2, and Table 3 we notice that if increases one variable, the other variable increases and vice versa, with the reduction of one variable decrease and the other variable, except in the case of the correlation between an online course and motivation which is a weak negative link, i.e. that’s as one variable increases, the other decreases and vice versa. Also, from Table 4, when comparing the correlation coefficients of Pearson, Kendall, and Spearman, we notice that the coefficient of Pearson and that of Spearman have more approximate values, compared to the coefficient of Kendall. Spearman's coefficient is a direct non-parametric counterpart of Pearson's correlation coefficient, which is why their values are closer to each other.

References

[1]. Al-Oklah H., Titi S., Alodat T., (2014) Introduction to statistics made easy, second edition
[2]. Albayrak A.S., Eroğlu A., Kalayci Ş., Küçüksille E., Ak B., Karaatli M., Keskin H.Ü., Çiçek E.U., Kayiş A., Antalyali Ö.L., Uçar N., Demirgil H., Işler D.B., Sungur O., (2017) Teknikat statistikore me shumë ndryshore me aplikim në SPSS
[3]. Emerson R.W., (2015) Causation and Pearson's Correlation Coefficient, Journal of Visual Impairment & Blindness, Volume 109, Issue 3
[4]. Hauke J., Kossowski T.,(2011) Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data, Quaestiones Geographicae, Bogucki Wydawnictwo Naukowe, Poznań-Poland, Volume 30, Issue 2
[5]. Iljazi T., (2021) The impact of professor's personality on teaching mathematics, Journal of Educational Research
[6]. Iljazi T., (2022) Teachers opinion on mathematics school books, Knowledge-International Journal,
[7]. Iljazi T., (2021) Students opinion about online teaching during the Covid-19 pandemic, Centrum,
[8]. Kamberi L., Statistika e zbatuar, Tetovë
[9]. Puka Ll.,(2010) Probabilitet dhe statistike e zbatuar, Tiranë