Economic Horizons, May - August 2014, Volume 16, Number 2, 161 - 163
UDC: 33 eISSN 2217-9232

VIEWPOINTS

Slavica Jovetic*'s comment on Correlation analysis of indicators of regional competitiveness: The case of the Republic of Serbia (2013)

doi: 10.5937/ekonhor1402165J

This letter is to bring to attention some inaccurate information provided in the article entitled Correlation analysis of the indicators of regional competitiveness: The case of the Republic of Serbia, published in *Economic Horizons* Vol. 15, No 3, as an original research paper. The research paper applied/used a statistical methodology for data analysis to which I add the following remarks:

Hypothesis formulation (p. 198) – Concerning the hypotheses testing thecausal relationship based on the simple linear correlation coefficient, the null hypothesis (H_0) assumes: there is no quantitative agreement between the occurrences, i.e. the simple linear correlation coefficient in the population equals zero, while the alternative hypothesis is a rival hypothesis stating quite the opposite from the zero hypothesis. In the relevant literature, which concerns hypotheses testing, if hypotheses are related to a statistical methodology, the null and alternative hypotheses are always given.

The correlation analysis does not examine dependence, but rather a quantitative agreement – please note that, on p. 198, in the second paragraph, the author(s) state that the paper does not address the issues of indicator values... but rather their correlational dependence.

The correlation analysis does not examine the frequency of relationships, but rather a quantitative agreement between the occurrences; to this end, please note that, on p. 201, the second paragraph states the following: „The correlation analysis..., but only on the existence and frequency of these relationships". I cannot conclude what the term frequency implies in the mentioned paper; however, a correlation analysis is a static analysis and it can also be a dynamic one, if a sample is selected at certain times where for each of the samples (time series t_1, t_2,...) simple linear correlation coefficients (R_1, R_2, ..., etc.) are determined for two random variables, which is normally used in determining a lag length when choosing lagged variables in a regression analysis.

Please note that, on p. 201, the second paragraph states that acorrelation analysis is the most complex analysis. On the contrary, a correlation analysis is not 100% reliable and is only used with some other analyses, i.e.:

- Regression analysis – Firstly, concerning the selection of independent variables that will be used in a regression analysis model, a simple linear correlation coefficient can be used. In that case, one should be careful because all variables in a regression model, which do not have aneffect on a dependent variable, must be eliminated from the model (significance $p > \alpha$). Furthermore, it can also be used in calculating a coefficient of determination (a coefficient of determination is a ratio of the explained variation to the total variation, where as a simple linear correlation coefficient is the positive square root of the R – squared, i.e., a coefficient of determination), which holds an important place/plays an important role in aregression analysis. It shows how much % of the variability of the dependent variable is explained by variations of independent variables which remained in the selected regression model.

**Correspodence to:* S. Jovetic, Faculty of Economics, University of Kragujevac; D. Pucara 3, 34000 Kragujevac, Serbia;
e-mail: sjovetic@kg.ac.rs

- Factor Analysis - One of the conditions of a factor analysis requires that there should be a statistically significant correlation between the independent variables in the model. The foregoing requirement at the beginning of a factor analysisis first checked by using three methods: a correlation coefficient and its statistical significance, Bartlett's test and the KMO (Kaiser-Mayer-Olkin) measure of sampling adequacy and their statistical significance. The condition that must be fulfilled is that all the three tests show the same level of statistical significance.

The reader is informed that the SPSS software package used for the statistical analysis was used for the purposes of the research in the paper; however, the exact version of the mentioned software package is not provided, regardless the fact that this is a mandatory requirement for all scientific papers.

Concerning the entire text of the paper, whenever Spearman's correlation coefficient is mentioned, the word „rank" must be mentioned, i.e. the correct wording is: Spearman's rank correlation coefficient.

It is indicative that the formula for calculating Spearman's rank correlation coefficientis given in the paper although this coefficient is not calculated in the paper, while the formula for Pearson's coefficient, i.e. a simple linear correlation coefficient, is not given in the paper although this coefficient was used in the paper. The formula used for the calculation of Pearson's coefficient is as follows:

$$R = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad \text{or} \quad R = \frac{\operatorname{cov}(x_i y_i)}{s_x s_y} \quad ,$$

$\bar{y}$ - the average of the observed y_i values, $\hat{y}_i$ - the estimated values, $cov(x_i y_i)$ - the covariance of the sample observations $x_i y_i$, $i=1, 2,...n$ i $s_x s_y$ - standard deviationsof the sampe observations $x_i y_i$, $i=1,2,...,n$.

The denotations used in the formula are incorrect. The following denotations are considered to be the standard ones: a correlation coefficient concerning a sample is marked with an R_s / r_s, while when applied to a population it is commonly represented by the Greek letter ρ_s. The Greek letter σ is reserved for apopulation standard deviation. Furthermore, the paper uses the letter n to denote the number of elementary units in the sample, which means that the letters used to denote the sample and the population are not used as prescribed by the standard and this may cause vagueness.

In addition, letters x and y (lowercase) are used to mark variables. Random variables are marked in capital letters (X and Y), while lowercase is reserved for the realizations in the sample (x_i and y_i, i = 1,2, ..., n).

Please, also note that it is stated in the paper that, if a piece of information is given on an ordinal scale, only Spearman's rank correlation coefficient can then be applied. The inaccuracy of this statement is further confirmed by the results of the indicators used in the paper. Some qualitative data were obtained by a survey; such data must be encrypted (e.g. 1 – the lowest value,... 5 – the highest value or *vice versa*) and only then can the simple linear correlation coefficient, i.e. Pearson's coefficient (the paper uses only Pearson's coefficient), be calculated.

Finally, the most significant remark concerning this paper is that the hypothesis on the statistical significance of the simple linear correlation coefficient is not tested by using p-empirical probability. The SPSS statistical software package does this automatically, and gives the following outputs: Pearson's coefficients, Spearman's rank correlation coefficients and statistical significance (p statistics). The results accounted forin the tables indicate that the SPSS software was not used, as the aforementioned outputs would have been included in these tables. Tables 1, 2 and 3 show the results of a hypothetical example contained in the SPSS 15.0 for Windows.

Table 1 Descriptive Statistics

	Mean	Std. Deviation	N
X	62,7000	16,96167	13
Y	27,4308	11,67000	13

Table 2 Correlations

		X	Y
	Pearson Correlation	1	,812(**)
X	Sig. (2-tailed)		,001
	N	13	13
	Pearson Correlation	,812(**)	1
Y	Sig. (2-tailed)	,001	
	N	13	13

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

Table 3 Correlations

			X	Y
Spearman's rho		Correlation Coefficient	1,000	,809(**)
	X	Sig. (2-tailed)	.	,001
		N	13	13
		Correlation Coefficient	,809(**)	1,000
	Y	Sig. (2-tailed)	,001	.
		N	13	13

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

Since $p < \alpha$, including a possible risk of an error of $\alpha = 0.01$ and $\alpha = 0.05$, the alternative hypothesis is confirmed, which means that there is a high statistical significance in terms of a quantitative agreement between the observed variables (Pearson's coefficient) in the population and the high statistical significance of the linear interdependence of the ranks of the observed variables in the population (Spearman's rank correlation coefficient).

Based on the scale given in the paper, valid conclusions on the statistical significance of the coefficients in the population cannot be derived. It is an imperative that a hypothesis for statistical significance should be tested.

The conclusion should not contain the following statement: „Pearson's coefficient shows that these indicators ...do not have any effect on ...". Let me emphasize once more that the simple linear correlation coefficient indicates an agreement/interactive relationship while a regression analysis, which is not used in the paper, makes it possible to measure an impact.

Received on 28th March 2014,
after revision,
accepted for publication on 19th August 2014.