Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 390 · 2019

Informatyka i Ekonometria 15

Wojciech Gamrot

University of Economics in Katowice
College of Management
Department of Statistics, Econometrics and Mathematics
wojciech.gamrot@ue.katowice.pl

ON USING THE *T*-TEST FOR ASSESSING ASSOCIATION BETWEEN LIKERT-SCALE VARIABLES

Summary: Survey objectives often include assessment of associations among study variables. These variables are sometimes discrete. In particular they may be expressed using the well known Likert scale, where the respondents choose one of several mutually exclusive, predefined responses. In such a case some authors advocate the use of correlation tests dedicated to testing hypotheses about correlation between continuous variables for discrete ones. In this paper, statistical consequences of such an approach are investigated, and resulting problems are illustrated for the well-known *t*-test based on assumption of bivariate normality.

Keywords: Likert scale, correlation, hypothesis testing, normality.

JEL Classification: C12, C18.

Introduction

In statistical surveys closed questions are frequently used. The respondent chooses one of several mutually exclusive response options. A typical case is the use of the Likert [1932] scale variables. Such an approach has many practical advantages, and in some situations it is unavoidable. At the same time research objectives frequently concern investigation of complex issues such as interdependence among two or more variables. A tendency may be observed to employ in such circumstances statistical methods dedicated to continuous – rather than discrete – variables. In particular, the well-known *t*-test, which is based on multivariate normality assumption, is employed to test for correlation [Norman, 2010;

Murray, 2013; Sullivan, Artino, 2013; Willits, Theodori, Luloff, 2016]. In this paper statistical effects associated with such an approach are investigated. Serious problems with test properties are illustrated.

1. Testing for correlation

Let us begin with the introduction of the Pearson correlation coefficient which characterizes the dependency between two random variables. Let *X* and *Y* denote two random variables. The coefficient is defined as [Bartoszyński, Niewiadomska-Bugaj, 2008, p. 243]:

$$\rho = \frac{E(XY) - E(X)E(Y)}{D(X)D(Y)} \tag{1}$$

and dates back to the works of Galton [1886; cf. Stigler, 1989] and Pearson [1895]. It takes values from the [–1,1] interval. The value of zero is interpreted as indication of no linear correlation. Non-zero values are interpreted as instances of linear correlation. To test the null hypothesis H_0: $\rho = 0$ it is quite common to employ the statistic [Zeliaś, 2000, p. 277; Jóźwiak, Podgórski, 1995, p. 246]:

$$t = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n-2} \tag{2}$$

where r_{xy} stands for the Pearson correlation coefficient calculated from the sample. When the sample is i.i.d. and both variables (*X*,*Y*) obey a bivariate normal distribution it follows the well-known Student's *t* distribution with *n*-2 degrees of freedom, and may be employed to verify if H_0 is true. Values of *t* which are sufficiently distant from zero are treated as evidence that variables *X* and *Y* are linearly correlated.

The application of this test for discrete variables measured using Likert scale, by treating numeric identifier of response variants as values of a numeric variable is problematic and controversial. Firstly, it is highly questionable whether the categories of Likert scale may indeed be treated as equally distant. Secondly, the correlation is defined in terms of averages, and it is not at all obvious what meaning should be assigned to averages of Likert-scale variables. For example, one might ask what is the average of categories such as "always", "often" and "rarely"? Thirdly, discrete variables by definition violate the bivariate normality assumption, so anyone who tries to apply the test to Likert variables should justify such an approach. One eventual possibility of justification could be to invoke limit theorems. It could be argued that sample moments have asymptotic normal distribution, and so for large samples the distribution of the statistic (2) approximates the Student's distribution. In the following section we will investigate whether this might be true in practical situations.

2. A counterexample distribution

If all the variants of two variables measured in a 5-point Likert scale are denoted by numbers 1,…,5, then their joint distribution may be expressed by the following matrix:

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} \quad (3)$$

The matrix *P* contains 25 numbers which represent joint probabilities of events:

$$p_{ij} = Pr\{X = i, Y = j\} \quad (4)$$

and add up to unity, so in fact the distribution is characterized by 24 free parameters whose sum does not exceed one. If the *t*-test is to be trusted, then it should hold desired properties for any valid values of these 24 parameters. In particular, the probability of rejecting the true null hypothesis H_0 while test's critical values are obtained from the Student's distribution for the 0.05 significance level should indeed be equal to 0.05 or at least not higher. This probability depends on distribution parameters and we will shortly call it the *test power function* (TPF). When the null hypothesis is true, it is a probability of making a type-I error when discreteness is disregarded and normality is assumed. It is not necessarily equal to the test size (which constitutes its lowest upper bound over the set of all parameter sets that satisfy the null hypothesis). Let us consider the following special case of the joint discrete distribution:

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & 0 & 0 & 0 & p_{15} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ p_{51} & 0 & 0 & 0 & p_{55} \end{bmatrix} \quad (5)$$

It reflects a situation where both variables can only take values at the opposite ends of the range: 1 and 5. This reflects an extreme polarization of respondent views – a situation that might occur in some political disputes. The four parameters of such a distribution: p_{11}, p_{15}, p_{51} and p_{55} must add up to unity, so in fact the distribution may be fully characterized by specifying only three of them. If in addition, let us assume that the variables X and Y are independent (which implies that the null hypothesis holds) and take maximum value of 5 respectively with probabilities $p_1 = \Pr\{X=5\}$ and $p_2 = \Pr\{Y=5\}$. This lets us write:

$$p_{11} = (1-p_1)(1-p_2) \tag{6}$$

$$p_{15} = (1-p_1)p_2 \tag{7}$$

$$p_{51} = p_1(1-p_2) \tag{8}$$

$$p_{55} = p_1 p_2 \tag{9}$$

Hence we obtain a family of discrete distributions for which the null hypothesis is satisfied, which is characterized by only two parameters: p_1 and p_2, both taking values in the [0,1] interval. If the t-test is to be trusted, then the true rejection probability should be equal or at least not exceed 0.05 for any of these distributions, and hence for any pair $(p_1,p_2)\in(0.1)^2$.

3. The distribution of the test statistic

Let us consider an i.i.d. sample consisting of n observations $(X_1,Y_1),\ldots,(X_n,Y_n)$ of the bivariate variable (X,Y) following the distribution introduced in the previous section. An empirical distribution of the sample may be expressed in the form:

$$N = \begin{bmatrix} n_{11} & 0 & 0 & 0 & n_{15} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ n_{51} & 0 & 0 & 0 & n_{55} \end{bmatrix} \tag{10}$$

where the four counts satisfy the condition $n_{11} + n_{15} + n_{51} + n_{55} = n$. It is important to note, that any of these counts may turn out to be equal to zero, which may hinder the calculation of the test statistic. In particular, it may happen that:

$$n_{15} = n_{51} = 0 \tag{11}$$

while the other two counts remain positive. This results in the Pearson coefficent $r_{xy} = 1$. Consequently the test statistic takes the value:

$$t = \frac{1}{\sqrt{1-1^2}}\sqrt{n-2} = \frac{1}{0} \tag{12}$$

Similarly, it may happen that:

$$n_{11} = n_{55} = 0 \tag{13}$$

while the other two counts remain positive. Consequently the Pearson coefficent takes the value $r_{xy} = -1$. In this case, the attempt at calculating the test statistic yields:

$$t = \frac{-1}{\sqrt{1-1^2}}\sqrt{n-2} = \frac{-1}{0} \tag{14}$$

Both results may be rather safely ignored in the case of continuous variables where the probability that they occur is equal to zero, but they cannot be ruled out in the discrete case. Both are problematic and appear with a strictly positive probability. One might wonder whether they should be interpreted as inability to test the null or rather – as a case of perfect correlation – constitute a basis for rejecting H_0. In this paper, the latter option is chosen and H_0 is rejected when they appear.

Another problem manifests itself when sample counts satisfy any of the following four conditions:

$$n_{11} = n_{15} = 0 \tag{15}$$

$$n_{51} = n_{55} = 0 \tag{16}$$

$$n_{11} = n_{51} = 0 \tag{17}$$

$$n_{15} = n_{55} = 0 \tag{18}$$

In each of these cases at least one of sample standard deviations of X and Y is equal to zero, while the covariance is also equal to zero. Hence we obtain $r_{xy} = 0/0$. This situation again requires careful interpretation. In this paper, we will treat such an outcome as a reason to not reject H_0. This is because any of the above-mentioned conditions – if satisfied – strongly suggests that at least one of the variables has its standard deviation equal to zero an hence is a constant. And if so, then there can be no linear correlation between this variable and any other one.

To illustrate the distribution of possible counts let us consider the sample of size $n = 3$. All the possible values of these counts are shown in the table 1. Corresponding values of the correlation coefficient, and the t-statistic are also provided (the asterisk * represents the inability to calculate t). When discreteness of the variable is disregarded and the alternative hypothesis H_1: $\rho \neq 0$ is considered, the two-sided critical region is limited by quantiles of the order 0.05 and 0.95 of the Student distribution with one degree of freedom. Hence the null hypothesis is rejected when $t \notin (-12.70, 12.70)$. The decision taken for each possible sample outcome is also shown in the table 1. One interesting fact is that for $n = 3$ the only way to reject the null hypothesis is to obtain $r_{xy} = 1$ or $r_{xy} = –1$ in the sample. When n grows it becomes possible to reject H_0 for a finite value of the test statistic. To calculate the probability of rejection one has to know the probabilities of obtaining every possible sample. The vector $[n_{11},n_{15},n_{51},n_{55}]$ has a multinomial distribution with parameters p_{11}, p_{15}, p_{51}, p_{55} and n. The first four of these parameters depend on p_1 and p_2. The last one may be set arbitrarily. This enables exact calculation of rejection probabilities for any p_1, p_2 and n.

Table 1. Possible configurations of sample counts and testing results for $n = 3$

n_{11}	n_{15}	n_{51}	n_{55}	r_{xy}	t	Reject
0	0	0	3	0/0	*	
0	0	1	2	0/0	*	
0	0	2	1	0/0	*	
0	0	3	0	0/0	*	
0	1	0	2	0/0	*	
0	1	1	1	–0.5	–0.58	
0	1	2	0	–1	–∞	reject
0	2	0	1	0/0	*	
0	2	1	0	–1	–∞	reject
0	3	0	0	0/0	*	
1	0	0	2	1	∞	reject
1	0	1	1	0.5	0.58	
1	0	2	0	0/0	*	
1	1	0	1	0.5	0.58	
1	1	1	0	–0.5	–0.58	
1	2	0	0	0/0	*	
2	0	0	1	1	∞	reject
2	0	1	0	0/0	*	
2	1	0	0	0/0	*	
3	0	0	0	0/0	*	

Source: Own elaboration.

4. Rejection probabilities

The TPF that reflects the probability of rejecting the null hypothesis H_0, using the Student's critical values, has been calculated numerically for p_1, $p_2 \in \{0.01, 0.02,…,0.99\}$ and for $n \in \{10,20,…,170\}$. Hence for any sample size n a total of 9801 differing (p_1,p_2) pairs was analysed. The TPF for $n = 50, 110, 170$ is shown in figures 1, 2, and 3, denoted by the symbol Z. It varies with p_1 and p_2. When these parameters take values close to 0.5, TPF oscillates around 0.05. However, when both p_1 and p_2 take values much higher or lower than 0.5, large deviations from this level appear. These deviations are negative when only one parameter is far from 0.5, and positive or negative when both of them are. The pattern of deviations slightly changes with sample size – locations of maximum and minimum deviations seem to shift when it grows.

The statistician usually does not have any control over true parameters of the studied distribution. They remain unknown. Hence from his point of view, the most interesting are the pessimistic cases: maximum and minimum distortions from the desired significance level. In figure 4 the maximum and minimum values of rejection probability recorded for $n = 10,20,\ldots,170$ are presented. Both exhibit a tendency to grow when the sample size increases. The growth is steady for the minimum, and quite irregular for the maximum. Most importantly the maximum value does not appear to tend to the desired level of 0.05. For the highest investigated sample size $n = 170$ it still exceeds 0.07 (so the true rejection probability is over 40% higher than expected). Meanwhile, the minimum does not exceed 0.01 so it is five times lower than expected.

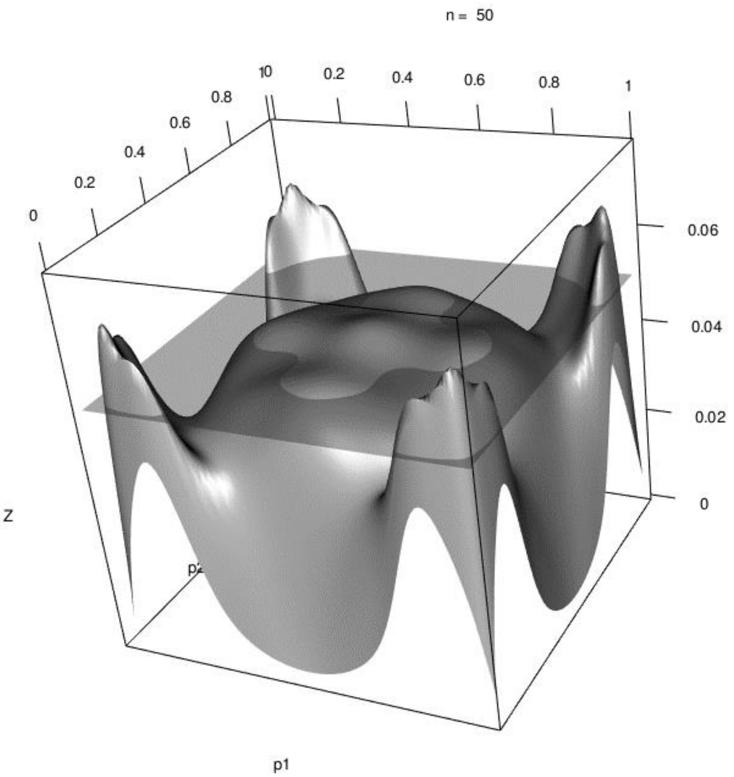

Fig. 1. Test power function for varying values of (p_1,p_2) and $n = 50$

Source: Own elaboration.

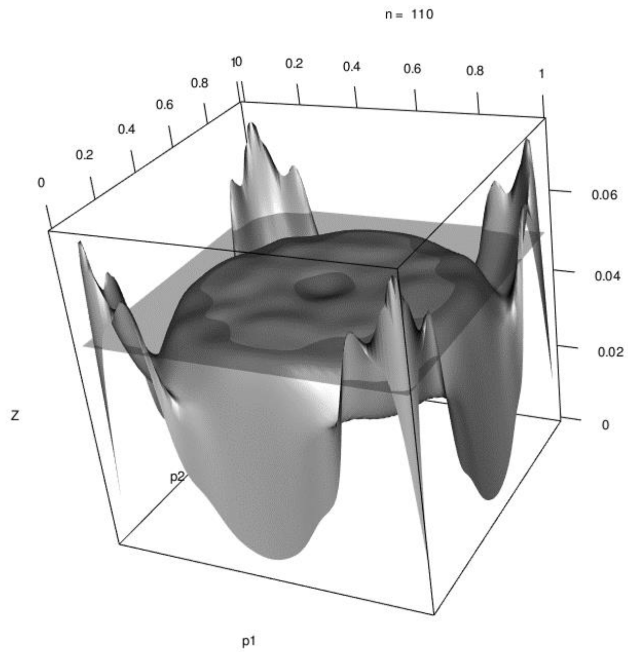

Fig. 2. Test power function for varying values of (p_1,p_2) and n = 110

Source: Own elaboration.

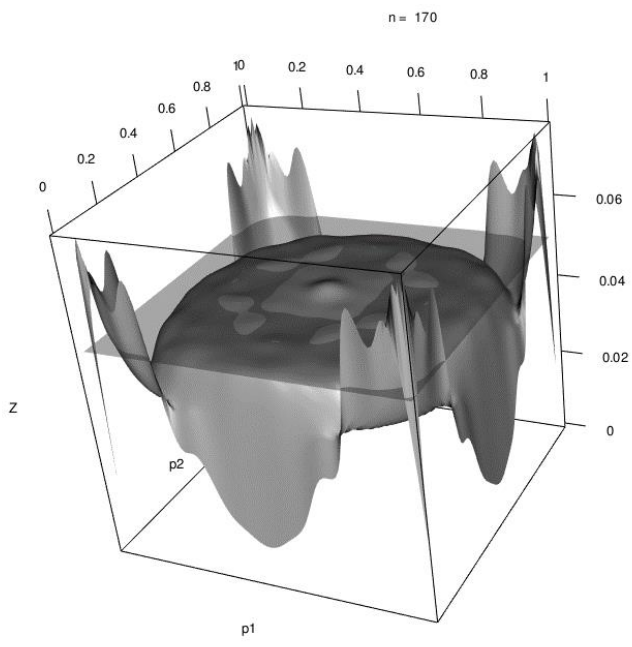

Fig. 3. Test power function for varying values of (p_1,p_2) and n = 170

Source: Own elaboration.

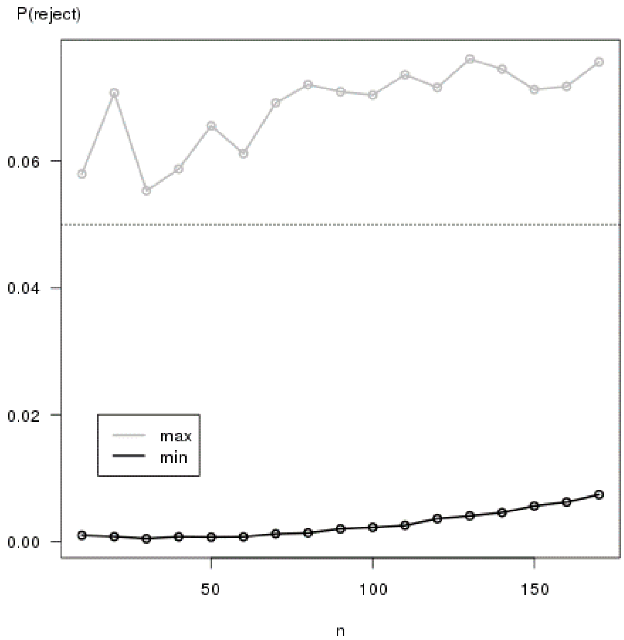

Fig. 4. Minimum and maximum probability of rejecting true null hypothesis (TPF) for Student's approximation observed on a grid, as a function of sample size $n = 10,\ldots,170$

Source: Own elaboration.

Conclusions

Presented results appear sufficient to question the use of Student's *t*-test for testing correlation of Likert-scale variables with sample sizes up to 170 units. For higher sample size no calculations were performed due to difficulties in numerical representation of large numbers. However, observed tendencies suggest that for higher sample sizes there may exist such distributions which result in even more pronounced positive deviations from the desired significance level. It should also be emphasized that in this study only some subset of possible distribution parameters (p_1,p_2) was investigated. Hence recorded minimum and maximum rejection probabilities should respectively be treated as upper and lower bounds for true minimum and maximum. The distortions may be more pronounced even for the studied sample sizes. This is especially true for parameter values very close to zero and one: the values which are higher than 0.99 or lower than 0.01 were not investigated. As a final recommendation, let us suggest avoiding the use of Pearson coefficient for ordinal variables. A safer choice would be to use measures of association which are dedicated to ordinal variables such as Somer's delta, Kendall's tau or Goodman and Kruskal's gamma with associated significance tests [Goodman, Kruskal, 1979; Agresti, 2019]. For situations where the existence of a continuous latent variable is postulated the polychoric correlation coefficient [Drasgow, 1986] may be considered.

References

Agresti A. (2019), *An Introduction to Categorical Data Analysis*, Wiley, New York.

Bartoszyński R., Niewiadomska-Bugaj M. (2008), *Probability and Statistical Inference*, Wiley, Hoboken.

Drasgow N. (1986), *Polychoric and Polyserial Correlations* [in:] S. Kotz, N.L. Johnson (eds.), *Encyclopedia of Statistical Sciences*, Vol. 7, Wiley, New York, p. 68-74.

Galton F. (1886), *Regression Towards Mediocricity in Hereditary Stature*, "Journal of the Anthropological Institute of Great Britain and Ireland", Vol. 15, p. 246-263.

Goodman L.A., Kruskal W.H. (1979), *Measures of Association for Cross Classifications*, Springer, New York.

Jóźwiak J., Podgórski J. (1995), *Statystyka od podstaw*, PWE, Warszawa.

Likert R. (1932), *A Technique for Measurement of Attitudes*, "Archives of Psychology", Vol. 22, No. 140, p. 5-55.

Murray J. (2013), *Likert Data: What to Use, Parametric or Non-Parametric?*, "International Journal of Business and Social Science", Vol. 4, No. 11, p. 258-264.

Norman G. (2010), *Likert Scales, Levels of Measurement and the "Laws" of Statistics*, "Advances in Health Sciences Education", Vol. 15, No. 5, p. 625-632.

Pearson K. (1895), *Notes on Regression and Inheritance in the Case of Two Parents*, "Proceedings of the Royal Society of London", Vol. 58, p. 240-242.

Stigler S.M. (1989), *Francis Galton's Account of the Invention of Correlation*, "Statistical Science", Vol. 4(2), p. 73-79.

Sullivan G.M., Artino A.R. (2013), *Analysing and Interpreting Data From Likert-Type Scales*, "Journal of Graduate Medical Education", Vol. 5, No. 4, p. 541-542.

Willits F.K., Theodori G.L., Luloff A.E. (2016), *Another Look at Likert Scales*, "Journal of Rural Social Sciences", Vol. 31, No. 3, p. 126-139.

Zeliaś A. (2000), *Metody statystyczne*, PWE, Warszawa.

O WYKORZYSTANIU TESTU *T* DO BADANIA WSPÓLZALEŻNOŚCI ZMIENNYCH MIERZONYCH NA SKALI LIKERTA

Streszczenie: W badaniach statystycznych często wykorzystywane są pytania zamknięte, na które respondent odpowiada, wybierając jedną z kilku wzajemnie wykluczających się opcji odpowiedzi. Typowym przypadkiem jest wykorzystanie do wyrażenia wartości zmiennych skali Likerta. Równocześnie dość często cel badania stanowi wykrycie współzależności pomiędzy zmiennymi. W niniejszej pracy rozważono konsekwencje kontrowersyjnej praktyki rozpatrywanej w literaturze, polegającej na stosowaniu testu *t*-Studenta przeznaczonego dla rozkładów ciągłych do badania współzależności pomiędzy zmiennymi dyskretnymi.

Słowa kluczowe: skala Likerta, korelacja, weryfikacja hipotez, normalność.