ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Czesław Domański**

SOME REMARKS ON EMPIRICAL POWER OF TESTS FOR PAIRS

Abstract. The paper deals with the tests for paired variables called also tests for pairs of variables. The observations are made of pairs of measurements. They can be correlated. It causes the necessity of applying another significance test of differences for example between means than in case of independent samples. We compare the power of nonparametric tests: sign test, Munzel and Wilcoxon tests with the Student's test for pairs.

Key words: observations for pairs, nonparametric tests, power of tests, Monte Carlo simulations.

1. PARAMETRIC TESTS

1.1. Student's Test for Pairs

Let us denote the two-dimensional parent population, which is characterized by the pair of random variables (X, Y) with two-dimensional normal distribution $N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. The random variable (X, Y) has two-dimensional normal distribution if its density is defined by the formula:

$$f(x, y) = k \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_1\sigma_2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\},$$

where:

$$k = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}},$$

* Professor, Chair of Statistical Methods, University of Łódź.

and the remaining symbols denote:

μ_x, μ_y – expected value of variables respectively X and Y,

σ_x, σ_y – standard deviations of these variables,

ρ – coefficient of these variables.

The sample of size of $n \geqslant 2$ was taken from this population. On the ground of this sample the hypothesis $H_0: \overline{Z} = \mu_x - \mu_y = 0$ against $H_1: \overline{Z} \neq 0$ ($\overline{Z} > 0$; $\overline{Z} < 0$) must be verified. The significance test is as follows. For each n independent pairs of results the difference $z_i = x_i - y_i$ was calculated and next the arithmetic mean of differences $\overline{z}$ and the variance s_z^2 were denoted:

$$\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i; \quad s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})^2.$$

The following statistic makes the gauge of this test:

$$t = \frac{|\overline{Z}|}{\mathrm{S}}\sqrt{\mathrm{n}-1}. \tag{1}$$

The statistic (1) has got Student's distribution t with degrees of freedom of $n-1$ if the hypothesis H_0 is true. The hypothesis H_0 is rejected and if the inequality $|t| > t_\alpha$ proceeds we accept the hypothesis $H_1: \overline{Z} \neq 0$. Considering one-sided tests we act similarly to the Student's test for independent samples.

Student's test t for paired variables can be applied if we examine earlier whether variables X and Y are correlated and, what is more, whether the differences between pairs make the sample taken from the population with the normal distribution or whether the variable (X, Y) has got two-dimensional normal distribution. The significance of the correlation coefficient ρ is examined with the Student's test t while the normality of the distribution is examined with Shapiro-Wilk test.

1.2. Test $q_{\overline{x}}$ for means k

Let us consider k-dimensional population, which is characterized by $X_1, X_2, \ldots, X_k$ continuous random variables with k-dimensional normal distribution. The n-element sample, which is recorded in the form of the matrix $\mathbf{X} = [x_{ij}]$, was sampled from this population. On the ground of this sample it is necessary to verify the hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ against the hypothesis $H_1: \mu_i \neq \mu_j$ for certain pairs $(i \neq j)$, while $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$. Testing the hypothesis H_0 is as follows:

1. We calculate the means' results:

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

2. We denote the matrix of deviation from means $\overline{x}_j$:

$$[x_{ij} - \overline{x}_j].$$

3. We find ranges R_i in matrix's $[x_{ij} - \overline{x}_j]$ rows:

$$R_u = (x_{ij} - \overline{x}_j)_{\max} - (x_{ij} - \overline{x}_j)_{\min}.$$

4. We put means $\overline{x}_j$ in order:

$$\overline{x}_{(1)}, \overline{x}_{(2)}, \ldots, \overline{x}_{(k)}.$$

5. We denote the value of the gauge:

$$q_{\bar{x}} = \frac{\overline{\mathbf{X}}_{(k)} - \overline{\mathbf{X}}_{(1)}}{\overline{\mathbf{R}}} c\sqrt{n}, \tag{2}$$

where c is the coefficient with the value depending on k and n, which can be read from the appropriate table (cf. Domański 1990).

6. We reject the hypothesis H_0, if $q_{\bar{x}} \geqslant q_\alpha$, while we read q_α for the assumed α, v^* and k degrees of freedom in the table while v^* we find earlier in the appropriate table (cf. Domański (1990)).

Test $q_{\bar{x}}$ for k paired samples can be sequentially applied and finding the diversification in the sample for the sake of k variables we can divide the sample and use the presented verification procedure for the results obtained in subgroups of the set of k variables.

2. NONPARAMETRIC TESTS

2.1. Sign Test

Two parent populations with continuous distribution function $F(x)$ and $F_2(y)$ are given. The equal number of n elements corresponding to each other in pairs was sampled. It means that the value of samples $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ is mutually paired because they refer to elements combined in pairs. On the ground of samples' results we can examine the hypothesis $H_0: F_1(x) = F_2(y)$ or $H_0: p = P(X - Y > 0) = P(X - Y < 0) = \frac{1}{2}$ against the alternative hypotheses $H_1: p < \frac{1}{2}$; $H_2: p > \frac{1}{2}$; $H_3: p \neq \frac{1}{2} (H_3: F_1(x) \neq F_2(y))$.

The number of signs makes the gauge of this test

$$r = \min(r^+, r^-), \tag{3}$$

where r^+ and r^- denote, respectively, the number of signs of positive and negative differences $(x_i - y_i)$ of investigated results' pairs in both samples for $i = 1, ..., n$.

Statistics r has got the binomial distribution if the hypothesis H_0 is true:

$$P(R = r) = \binom{n}{r}\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{n-r}. \tag{4}$$

This distribution was tabulated but tables give the number of signs r that $\mathrm{P}(r \leqslant r_\alpha) = \alpha$. In this paper the critical image has got the left-sided construction.

We reject the hypothesis H_0 for the hypothesis H_1, if $r \leqslant r_\alpha$, for H_2, if $r \geqslant n - r_\alpha$, for H_3, if $r \leqslant r_{\alpha/2}$.

2.2. Test of Ranked Signs

Two parent populations with continuous distribution functions $F_1(y)$ and $F_2(y)$ are given. The equal number of n elements for both samples, whose results correspond to each other in pairs, was taken from these populations. On the ground of these results of samples we can examine the hypothesis $H_0: F_1(x) = F_2(y)$. To this end we denote results' differences of both of the samples $(x_i - y_i)$ for all pairs and next we give the ranks to absolute values of these differences. Denoting T^+ and T^-, that is the sum of differences of ranks respectively positive and negative we obtain the gauge of the ranked sign test:

$$T = \min(T^+, T^-). \tag{5}$$

Statistic T has got the known distribution for which Wilcoxon built the table $P(T \leqslant T\alpha)$ if the hypothesis H_0 is true. In this paper the critical region has got the left-sided construction. The hypothesis H_0 is rejected when we obtain the inequality $T \leqslant T_\alpha$.

For $n > 25$ we can use the following gauge:

$$U = \frac{\mathbf{T} - \mathbf{E}(T)}{\mathbf{D}(T)}, \tag{6}$$

where:

$$\mathbf{E}(T) = \frac{1}{4}n(n+1); \quad \mathbf{D}^2(T) = \frac{1}{24}n(n+1)(2n+1).$$

The gauge has got the asymptotically normal distribution $N(0, 1)$ if the hypothesis H_0 is true.

2.3. Fisher's Sign Test

Two parent populations with continuous distribution functions $F_1(x)$ and $F_2(y)$ are given. The equal number of n elements $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ corresponding to each other in pairs was sampled from these populations. The Fisher's sign test will be as follows. We create the sequence of differences $z_i = y_i - x_i$, while:

$$z_i = \theta + e_i \quad (i = 1, \ldots, n), \tag{7}$$

where e_i makes the independent sequence of remainders, that is:

$$P\{e_i < 0\} = \mathbf{P}\{e_i > 0\} = \frac{1}{2} \quad (i = 1, \ldots, n), \tag{8}$$

and θ denotes the parameter we investigate (for example the unknown effect of particular procedure). We are to verify the hypothesis $H_0 : \theta = 0$ against hypotheses $H_2 : 0 > 0$ or $H_3 : \theta \neq 0$.

The following form of the statistic makes the gauge of the Fisher's sign test

$$B = \sum_{i=1}^{n} \psi_i, \tag{9}$$

where:

$$\psi_i = \begin{cases} 1, & \text{when} \quad z_i > 0, \\ 0, & \text{when} \quad z_i < 0. \end{cases} \tag{10}$$

In the tables of distribution function of binomial distribution we can read critical values b_α for the assumed significance level α and $p = \frac{1}{2}$. The hypothesis H_0 is rejected for the hypothesis:

$$H_1 \quad \text{when} \quad B \geqslant b_\alpha$$
$$H_2, \quad \text{when} \quad B \leqslant n - b_\alpha$$
$$H_3, \quad \text{when} \quad B \geqslant b_{\alpha_2} \quad \text{or} \quad B \leqslant n - b_{\alpha_1},$$

where $\alpha = \alpha_1 + \alpha_2$.

In case of big samples we use the following statistic:

$$B^* = \frac{B - E(B)}{D(B)} = \frac{B - \frac{n}{2}}{\sqrt{\frac{1}{4}n}}. \tag{11}$$

This statistic has got the asymptotically normal distribution $N(0, 1)$ if the hypothesis H_0 is true. We use the hypothesis H_0 if $B^* \geqslant u_\alpha(H_1$, when $B^* < u_\alpha H_3$, when $|B^*| > u_\alpha)$ where u_α is the quantile of the distribution $N(0, 1)$ of the rank α.

2.4. Friedman's Test

The k-dimensional population, which is characterized by random variables $X_1, X_2, \ldots, X_k$ with k-dimensional continuous distribution is given. The n-element sample $(n \geqslant 10)$ which can be recorded in the form of the matrix $\mathbf{X} = [x_{ij}]$ $(i = 1, \ldots, n;\ j = 1, \ldots, k)$ was sampled from this population. On the ground of this sample it is necessary to verify the hypothesis $H_0 : F_1(x) = \ldots = F_k(x)$ against the hypothesis $H_1 : F_i(x) \neq F_i(x)$ for certain pairs $(i \neq j)$. We order values of each k of results $(x_1, \ldots, x_k)$ assigning appropriate ranks: 1, 2, ..., k, to the results and next we denote the value of the test's gauge:

$$\chi^2 = \frac{12}{k(k+1)n} \sum_{j=1}^{k} T_j^2 - 3(k+1)n, \tag{12}$$

where $T_j = \sum_{i=1}^{n} r_{ij}$ is the sum of the ranks for the results j.

We reject the hypothesis H_0 if $\chi^2 \geqslant \chi^2_\alpha$.

2.5. Munzel Test

Let $R(X_i)$ and $R(Y_i)$ denote the ranks of X_i and Y_i in the paired sample $X_1, ..., X_n, Y_1, ..., Y_n$. Munzel (1999) has shown that if $H_0: F_1(x) = F_2(y)$ is true then the following statistic

$$M_n^{(1)} = \frac{\overline{R(X)} - \overline{R(Y)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R(X_i) - R(Y_i) - \overline{(R(X)} - \overline{R(Y)})^2}}, \tag{13}$$

where $\overline{R(X)} = \frac{1}{n}\sum_{i=1}^{n} R(X_i)$ and $\overline{R(Y)} = \frac{1}{n}\sum_{i=1}^{n} R(Y_i)$ has the asymptotic standard normal distribution. In a similar way Munzel (1999) proposes another test statistic which is based on the difference between the overall ranks and the internal groups ranks:

$$M_n^{(2)} = \frac{1}{2}\sqrt{n} = \frac{\overline{R(X)} - \overline{R(Y)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}((R(X_i) - R_X(Y_i) - (R(Y_i) - R_Y(Y_i) - \overline{(R(X)} - \overline{R(Y)}))^2}}, \tag{14}$$

where $\overline{R(X)} = \frac{1}{n}\sum_{i=1}^{n} R(X_i)$ and $\overline{R(Y)} = \frac{1}{n}\sum_{i=1}^{n} R(Y_i)$ are the average ranks in the paired sample and $R_X(X_i), R_Y(Y_i)$ define the ranks in each sample separately. The statistic (13) has also the standard normal limiting distribution.

2.6. The Power of Tests

The investigation of the power of tests was made by the assumption of the correlation between variables X and Y. We have investigated the bivariate normal distribution $(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, where $\mu_X = 0$, $\sigma_X = \sigma_Y = 1$. The power of test can be displayed as a function of μ_Y, where $\mu_Y = 0$ corresponds to hypothesis H_0 and $\mu_Y > 0$ to alternative hypothesis H_1.

Note that the joint distribution of X. Y can be shown as follows (see Kraft, Schmidt 2003):

$$H(x, y) = P(X \leqslant x, Y \leqslant y) = C_p(\Phi(x), \Phi(y - \mu_Y)), \tag{14}$$

where C_p is the so-called normal copula in the form of:

$$C_p(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left(\frac{t^2 - 2\rho st + s^2}{1-\rho^2}\right)\right\} dtds.$$

In Monte-Carlo experiment the number of repetitions for $n = 10$, 20, 50 and $\alpha = 0.005$ is $q = 10\ 000$.

Table 1 represents the empirical power of 4 tests for $n = 40$ and $\alpha = 0.05$.

Table 1. Empirical power of tests for pairs for $n = 40$ and $\alpha = 0.05$ (w ‰)

Tests	ρ				
	–0.8	–0.4	0.0	0.4	0.8
t-Student	448	522	621	819	1000
Sign	358	411	502	686	952
Wilcoxon	439	518	615	811	996
Munzel	441	515	616	797	991

3. FINAL REMARKS

The type and the power of correlation influence significantly the power of tests for pairs. The higher correlation coefficient referring to the absolute value the higher power of the considered tests.

Test t, what was easy to predict, is the smallest. The sign test shows the smallest power. Wilcoxon and Muinzel tests for $n = 50$, from the point of view of the power, behave similarly.

The power of the Wilcoxon test for $n = 20$ is higher than for the Munzel test.

For small $n \leqslant 20$ the power of the considered tests is not very high.

REFERENCES

Domański Cz. (1990), *Testy statystyczne*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.

Hollander M., P., Ledger G., Lin P. E. (1974), "Robustness of the Wilcoxon Test to a Certain Dependency Between Variables", *Annals of Statistics*, **2**, 177–181.

Kraft S., Schmidt F. (2003), "Power of Tests for Stochastic Order When Observations Are Paired", *Allgemeines Statistisches Archiv*, **87**, 239–255.

Munzel U. (1999), "Nonparametric Methods for Paired Samples", *Statistica Nurlandica*, **53**, 277–286.

Czesław Domański

UWAGI O EMPIRYCZNEJ MOCY TESTÓW DLA PAR

(Streszczenie)

W artykule prezentowane są testy dla zmiennych połączonych, zwanych także testami dla par zmiennych. Obserwacje składają się z par pomiarów. Mogą być one skorelowane. Sytuacja ta sprawia, że należy zastosować inny test istotności różnic, np. pomiędzy średnimi aniżeli w przypadku prób niezależnych. Porównujemy moc testów nieparametrycznych: znaków, Wilcoxona i Munzela z testem *t*-Studenta dla par.