

Driver salary identification by hypothesis testing

Miloš Poliak¹, Patrícia Šimurková¹, Adela Poliaková², Marek Jaśkiewicz³

¹Department of Road and Urban Transport, University of Žilina, Univerzitná 1, Žilina, Slovakia

²Faculty of Management and Economics, Tomas Bata University in Zlin, Mostni c. 5139 – U2, Zlin, Czech Republic

³Faculty of Mechatronics and Machine Desing, Kielce University of Technology, al. Tysiąclecia Państwa Polskiego 7, Kielce, Poland

Abstract This paper outlines comparing two independent samples, particularly average driver salary in Slovakia and the other EU countries. Significant differences between salaries are pointed out through two examples dealing with hypothesis testing. For Western European Union countries, the level of salaries in road transport is higher in absolute terms compared to the salaries of drivers in the central and eastern parts of the European Union. Different salaries and requirements of wage regulations create a discriminatory environment among entrepreneurs however harmonization of salaries in the field of road transport is specific because carriers offer transport throughout the whole European Union market.

Keywords driver, salary, hypothesis testing, comparing

JEL J31, R40

1. Introduction

In practice, we often need to compare the basic characteristics of a random variable in two basic sets based on the knowledge of the sample data only. For example, we want to compare average salaries of drivers and their variability in individual countries or the annual tax on motor vehicles and the like. If the comparison is based on the results of the sample research, the statistical induction methods must be reused for the interval estimates of the difference or share of the comparison parameters of the basic files or for verifying the hypotheses of their compliance.

2. Comparing Two Independent Samples

In the two-parameter match tests, independent choices is considered in which the collection of statistical units from a single base file does not depend on removing units from the second base file. Based on independent selections, the hypotheses of matching the parameters of two basic sets are verified, using selection characteristics with known probability distributions. These distributions often depend on whether the variance in the basic files is the same or not.

2.1. Testing a Difference of Two Variances

The random variable X is monitored in two basic files. Its probability distribution is normal for both files. It is assumed

that in both the basic files, the random variable X has the same variability, which is expressed by the hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2 \quad (1)$$

The test characteristic is used as the test criterion for its test. If the null hypothesis is true, the test is

$$F = \frac{\bar{s}_1^2}{\bar{s}_2^2} \quad (2)$$

and Fisher's probability distribution $F(n_1 - 1; n_2 - 1)$ which has number of degrees of freedom $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. An overview of critical regions for various alternative hypotheses is given in Table 1 [5], [6].

Table 1. Critical Region for Various Alternative Hypotheses (F-Test)

Alternative hypothesis	Critical Value	Critical Region
$H_1: \sigma_1^2 \neq \sigma_2^2$	$F_{\frac{\alpha}{2}, F_{1-\frac{\alpha}{2}}}$	$v_\alpha = (0, F_{\frac{\alpha}{2}}) \cup (F_{1-\frac{\alpha}{2}}, +\infty)$
$H_1: \sigma_1^2 > \sigma_2^2$	$F_{1-\alpha}$	$v_\alpha = (F_{1-\alpha}, +\infty)$
$H_1: \sigma_1^2 < \sigma_2^2$	F_α	$v_\alpha = (0, F_\alpha)$

2.2. Testing a Difference between Two Means - Normal Distribution

If we know the variances σ_1^2, σ_2^2 of these distributions and test the null hypothesis

$$H_0: \mu_1 = \mu_2 \quad (3)$$

against any alternative, as a test criterion we can use a characteristic assuming the null hypothesis is valid

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (4)$$

and normal probability distribution $N(0; 1)$. Selected level of significance α and critical regions for different types of alternative hypotheses are given in Table 2.

Table 2. Critical Region for Various Alternative Hypotheses (Z-Test)

Alternative hypothesis	Critical Value	Critical Region
$H_1: \mu_1 > \mu_2$	$-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}$	$v_\alpha = (-\infty, -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, +\infty)$
$H_1: \mu_1 > \mu_2$	$z_{1-\alpha}$	$v_\alpha = (z_{1-\alpha}, +\infty)$
$H_1: \mu_1 > \mu_2$	$-z_{1-\alpha}$	$v_\alpha = (-\infty, -z_{1-\alpha})$

If we do not know the variances σ_1^2, σ_2^2 but we can assume its equality $\sigma_1^2 = \sigma_2^2$, the null hypothesis $H_1: \mu_1 = \mu_2$ is tested against any alternative by the test characteristic, which is in the case of null hypothesis

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (5)$$

and distribution $t(n_1 + n_2 - 2)$.

Critical values are determined at the chosen level of significance α as the appropriate fractil of Student's distribution $t(n_1 + n_2 - 2)$ according to Table 3 [5], [6].

Table 3. Critical Region for Various Alternative Hypotheses (T-Test)

Alternative hypothesis	Critical Value	Critical Region
$H_1: \mu_1 \neq \mu_2$	$-t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}$	$v_\alpha = (-\infty, -t_{1-\frac{\alpha}{2}}) \cup (t_{1-\frac{\alpha}{2}}, +\infty)$
$H_1: \mu_1 > \mu_2$	$t_{1-\alpha}$	$v_\alpha = (t_{1-\alpha}, +\infty)$
$H_1: \mu_1 < \mu_2$	$-t_{1-\alpha}$	$v_\alpha = (-\infty, -t_{1-\alpha})$

If we do not know the variances σ_1^2, σ_2^2 and we cannot assume its equality, a suitable test criterion is the characteristic which, in the case of the null hypothesis validity and in the case of validity of the null hypothesis $\mu_1 = \mu_2$ is

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (6)$$

and the division $t(v_r)$.

Critical regions in various alternative hypotheses are given in Table 3 with the difference that we are looking for the fractile of t-distribution at degrees of freedom v_r .

2.3. Testing a Difference between Two Means – Other or Unknown Distribution

The assumptions about the consistency of the mean values of the two base files in which the random variable X does not have normal but any distribution or unknown

probability distribution is tested if the sample files have a large range, $n_1 \rightarrow +\infty, n_2 \rightarrow +\infty$. The formula for the test statistic in this case is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (7)$$

and normal probability distribution $N(0; 1)$.

The normality of its distribution is a consequence of the validity of the central limiting theorem and the use of selective variances instead of the variances of the base sets is due to the consistency of point estimates [9], [10].

2.4. Testing a Difference between Two Means – Alternative Distribution

This is a special case of the average value match test of two arbitrary distributions. If X has a distribution $A_{(\pi_1)}$ in the first one and $A_{(\pi_2)}$ in the second base file, the formula for null hypothesis about the consistency of mean values is

$$H_0: \pi_1 = \pi_2 \quad (8)$$

An appropriate test criterion in its test against any alternative is the characteristic, which is simplified in the case of the validity of the null hypothesis

$$Z = \frac{(P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \quad (9)$$

Assuming $n_1 \rightarrow \infty, n_2 \rightarrow \infty$, the Z -test has normalized normal distribution and the assumption or rejection of the null hypothesis $H_0: \pi_1 = \pi_2$ is decided based on Table 2 [12], [13].

3. Driver Salary Identification

Regarding the remuneration of drivers, we conducted a research of drivers' compensation in selected transport companies. The research was conducted from January to March 2017 and the drivers of transport companies from 10 different EU states participated on it. Overall, drivers from 310 transport companies took part in the research. The countries were chosen so that in the research were the states which uphold the introduction of a minimum salary for the work of international road transport drivers (Germany and Austria), Western European countries which incline to a certain protection of the transport market (Belgium, Denmark, the Netherlands, Luxembourg, Italy) and the countries of the central and eastern part of the European Union which are against the unification of the minimum salary across the EU (Czech Republic, Poland, Slovakia).

In the research, the drivers responded to the form of remuneration and the amount of remuneration which appertain to them. The drivers were divided into groups according to whether they are remunerated with a monthly or hourly rate. On average, the highest hourly salary rate is reached by drivers in Luxembourg, it is 15.875 € / hour (Table 4). By the rates above € 10 / hour are also remunerated drivers in Germany and Austria. Significantly lower rates are in the Czech Republic and Slovakia, where they are below 5 € / hour.

In several countries, drivers are not remunerated with € / hour, but with monthly salaries. The highest average salary for drivers is in Luxembourg, Germany and the Netherlands and it is above 2000 € / month.

Table 4. Remuneration of drivers in international road transport

Country	Wage (€/h)	Wage (€/km)	Wage (€/month)
Germany	12.875	-	2347
Austria	10.33	-	1633
Belgium	-	-	1605
Denmark	-	-	1600
Netherlands	-	-	2100
Luxemburg	15.875	-	2381
Italy	-	0.15	1516
Czech Republic	3.77	0.083	1115
Poland	-	0,08	1400
Slovakia	4.33	0.131	1282

3.1. Example 1 – Salary Comparison of Slovak and Czech Driver

From research we know the average salaries of drivers in two different countries. In the SR we determined the average salaries of drivers with random sample of size 116 (all entries are in euro):

1380; 1500; 1050; 1170; 1250; 1350; 1400; 900; 550; 600; 1600; 2400; 1200; 1500; 1200; 1600; 1800; 850; 1700; 1800; 2000; 1200; 1200; 1700; 1200; 1550; 1300; 1650; 1500; 1400; 1500; 1500; 800; 850; 900; 1200; 1200; 750; 1200; 1600; 1100; 1000; 1350; 1400; 600; 1400; 550; 1500; 1200; 600; 1200; 1600; 1200; 1000; 700; 1000; 1600; 2000; 1500; 500; 1000; 1450; 1500; 600; 650; 1400; 750; 450; 2500; 1200; 2000; 1400; 600; 800; 1400; 550; 2500; 1100; 650; 1100; 1300; 900; 700; 800; 1300; 1500; 1000; 530; 2300; 1000; 1400; 1500; 2200; 1400; 750; 1600; 1200; 2100; 1100; 1600; 1200; 1500; 1300; 1750; 2000; 540; 404; 1400; 900; 1810; 1500; 1200; 500; 1300; 1400; 1500.

In the Czech Republic we determined the average salaries of drivers (all entries are in euro recalculated on the basis of the National Bank of Slovakia exchange rate sheet on 2.7.2017) with random sample of size 31:

1356; 573; 1600; 1146; 620; 1612; 1774; 2063; 1070; 1375; 554; 1135; 688; 1146; 1176; 722; 1031; 1713; 1135; 879; 1146; 825; 840; 722; 928; 1031; 877; 2475; 722; 933; 1719.

At the level of significance $\alpha = 0.05$, we assume that average salaries are the same in both countries. We want to verify the hypothesis $H_0: \mu_1 = \mu_2$ versus the alternative $H_1: \mu_1 \neq \mu_2$. The sample size is large with an unknown variances σ_1^2, σ_2^2 . In order to be able to choose a test criterion correctly, we must first verify the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ against the alternative $H_1: \sigma_1^2 \neq \sigma_2^2$.

Next, we calculate the mean and variance for each sample. Doing so gives us:

$$\bar{x}_1 = 1258.5690$$

$$\tilde{s}_1^2 = 211810.4563$$

$$\tilde{s}_1 = 460.2287$$

$$\bar{x}_2 = 1149.8757$$

$$\tilde{s}_2^2 = 210709.1835$$

$$\tilde{s}_2 = 459.0307$$

Now we can fit into the test criterion and get it:

$$F = \frac{\tilde{s}_1^2}{\tilde{s}_2^2} = \frac{211810.4563}{210709.1835} = 1.0052$$

Then we determine critical values of F-distribution with degrees of freedom $\alpha = 0.05$:

$$F_{\frac{\alpha}{2}} = F_{0,025} = 0,59, F_{1-\frac{\alpha}{2}} = F_{0,975} = 1,87$$

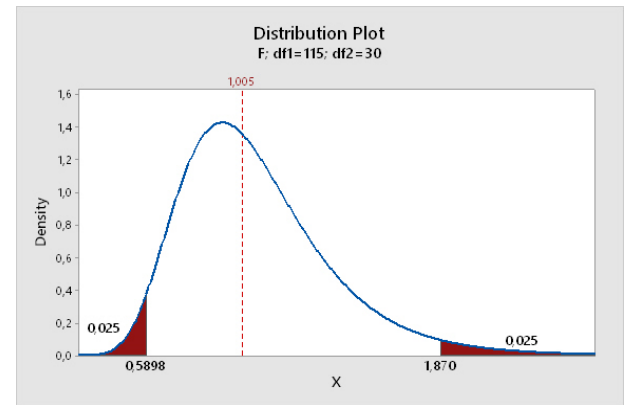


Figure 1. Distribution Plot – F-test

The value of the test criterion does not belong to the critical region (Fig. 1.)

$$v_{\alpha} = (0, F_{\frac{\alpha}{2}}) \cup (F_{1-\frac{\alpha}{2}}, +\infty) \quad (10)$$

therefore we accept the null hypothesis at the level of significance $\alpha = 0.05$.

A test criterion is appropriate for verifying the hypothesis $H_0: \mu_1 = \mu_2$. First, we calculate the combined standard deviation:

$$\begin{aligned} \tilde{s}_p &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)\tilde{s}_2^2}{(n_1 + n_2 - 2)}} \\ &= \sqrt{\frac{115 \times 211810.4563 + 30 \times 210709.1835}{(116 + 31 - 2)}} = 459.9811 \end{aligned}$$

Now we calculate our test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\tilde{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1258.5690 - 1149.8757}{459.9811 \sqrt{\frac{1}{116} + \frac{1}{31}}} = 1.1687$$

In Excel, using the TINV function, we find the 0.05 critical values for a T-distribution with 145 degrees of freedom to be -1.98 and 1.98 . So to accept the null hypothesis, we need out test statistic to satisfy $t < 1.98$ [12], [13].

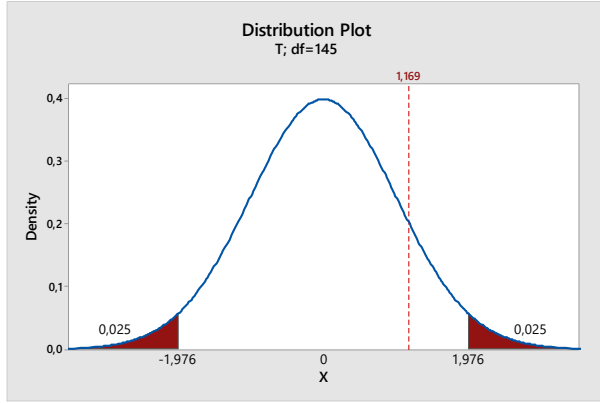


Figure 2. Distribution Plot – t-test

The value of the test criterion does not belong to the critical area (Figure 2.), therefore we assume the null hypothesis $H_0: \mu_1 = \mu_2$ for the significance level $\alpha = 0.05$. At the level of significance $\alpha = 0.05$ we assume that in the SR and the Czech Republic the average salary of drivers is not significantly different.

3.2. Example 2 – Salary Comparison of Slovak driver and drivers from the other EU countries

From research we know the average salaries of drivers in SR and the other EU countries. In the SR we determined the average salaries of drivers with random sample of size 116 (the same salaries as example 1 are considered).

In selected EU countries (Belgium, Denmark, Luxembourg, Netherlands, Norway, Poland, Austria, and Italy) we determined the following average salaries of drivers with random sample of size 60 (all entries are in euro):

500; 1750; 2450; 1500; 630; 2200; 2160; 1650; 1600; 900; 1500; 1750; 2500; 2500; 2100; 2180; 3690; 2500; 1850; 2300; 2100; 4301; 1400; 1600; 1600; 1700; 1750; 600; 2200; 1356; 573; 1600; 1146; 620; 1612; 1774; 2063; 1070; 1375; 554; 1135; 688; 1146; 1176; 722; 1031; 1713; 1135; 879; 1146; 825; 840; 722; 928; 1031; 877; 2475; 722; 933; 1719.

At the level of significance $\alpha = 0.05$, we assume that average Slovak salaries are the same as salaries in selected EU countries.

We want to verify the hypothesis $H_0: \mu_1 = \mu_2$ versus the alternative $H_1: \mu_1 \neq \mu_2$. The sample size is large with an un-known variances σ_1^2, σ_2^2 . Next, we calculate the mean and variance for each sample:

$$\bar{x}_1 = 1258.5690$$

$$\tilde{s}_1^2 = 211810.4563$$

$$\tilde{s}_1 = 460.2287$$

$$\bar{x}_2 = 1517.4500$$

$$\tilde{s}_2^2 = 240216.0465$$

$$\tilde{s}_2 = 490.1184$$

Now we can fit into the test criterion and get it:

$$F = \frac{\tilde{s}_1^2}{\tilde{s}_2^2} = \frac{211810.4563}{240216.0465} = 0.8817$$

Then we determine critical values of F-distribution with and degrees of freedom (115; 59) $\alpha = 0.05$:

$$F_{\frac{\alpha}{2}} = F_{0.025} = 0.65, F_{1-\frac{\alpha}{2}} = F_{0.975} = 1.59$$

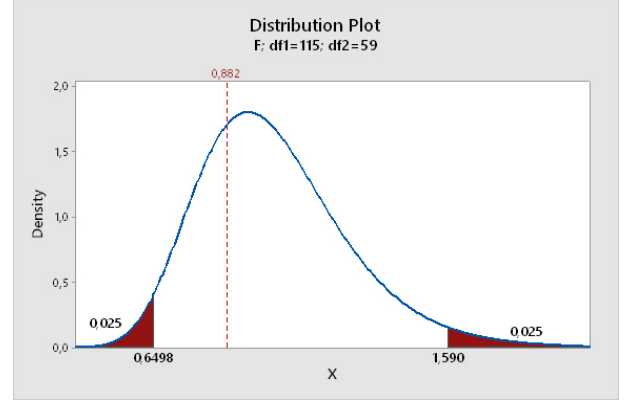


Figure 3. Distribution Plot – F-test

The value of the test criterion does not belong to the critical region (Fig. 3.)

$$v_\alpha = (0, F_{\frac{\alpha}{2}}) \cup (F_{1-\frac{\alpha}{2}}, +\infty) \quad (11)$$

therefore we accept the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ at the level of significance $\alpha = 0.05$.

A test criterion is appropriate for verifying the hypothesis $H_0: \mu_1 = \mu_2$. First, we calculate the combined standard deviation:

$$\begin{aligned} \tilde{s}_p &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)\tilde{s}_2^2}{(n_1 + n_2 - 2)}} \\ &= \sqrt{\frac{115 \times 211810.4563 + 59 \times 240216.0465}{(116 + 60 - 2)}} = 470.5765 \end{aligned}$$

Now we calculate our test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\tilde{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1258.5690 - 1517.4500}{470.5765 \sqrt{\frac{1}{116} + \frac{1}{60}}} = -3.4595$$

In Excel, using the TINV function, we find $\alpha 0.05$ critical values for a T-distribution with 174 degrees of freedom to be -1.97 and 1.97 [12], [13].

Since $t = -3.4595$ is inside of our critical region (Fig. 4), we accept the null hypothesis and we can conclude at $\alpha 0.05$ significance level that the average salary of the driver in the Slovak Republic and in the other analysed EU countries is significantly different.

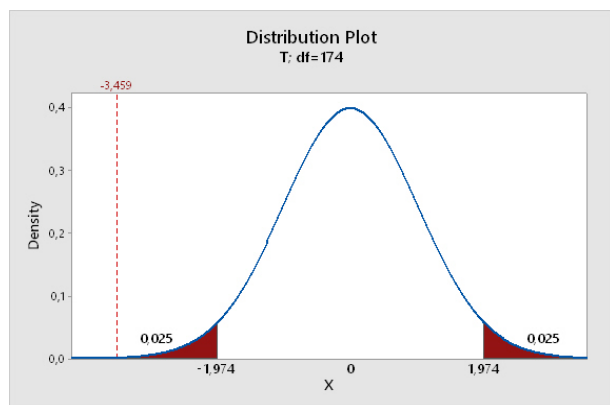


Figure 4. Distribution Plot – t-test

Since $t = -3.4595$ is inside of our critical region (Fig. 4.), we accept the null hypothesis and we can conclude at a 0.05 significance level that the average salary of the driver in the Slovak Republic and in the other analysed EU countries is significantly different.

5. Conclusions

On the basis of testing hypotheses, we can conclude that there are differences between driver salaries in the selected EU countries. Testing has proven that in the Slovak Republic and the Czech Republic the average salary of drivers is not significantly different at the level of significance $\alpha 0.05$ but on the other hand there are significant differences between the average salary of the driver in the Slovak Republic and in the other analysed countries, mostly Western European Union countries. For Western European Union countries, the level of salaries in road transport is higher in absolute terms compared to the salaries of drivers in the central and eastern parts of the European Union. Therefore, it can be expected that in the western part of the EU drivers are less interested in this profession, and companies are often looking for workers from the eastern part of the EU or from outside the EU. The harmonization of salaries in the field of road transport is specific because carriers offer transport throughout the whole European Union market. Every carrier established in the European Union and with a license issued by the Community has the same access to this market. Individual states put pressure on increase of salaries in international road transport with the use of national regulations in order to reduce the competitive pressure of lower prices of carriers located in the central and eastern parts of the European Union.

Acknowledgement

The contribution was elaborated with the support of the Ministry of Education of the Slovak Republic VEGA no. 1/0143/17 POLIAK, M.: Increasing the competitiveness of Slovak carriers providing road transport services in the common market of the European Union.

REFERENCES

- [1] BLENDON, R. et al (1997): Bridging the gap between the public' and economists' views of the economy. *Journal of Economic Perspectives* 11, p. 105–118.
- [2] BRENNAN, G. - BUCHANAN, J. (1980): *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge University Press, Cambridge U.K.
- [3] CAPLAN, B. (2002): Systematically biased beliefs about economics: robust evidence of judgemental anomalies from the Survey of Americans and economists on the economy. *The Economic Journal* 112, p. 433–458.
- [4] CYPRICH, O. - KONEČNÝ, V. – KILIANOVA, K. (2013): Short-term passenger demand forecasting using univariate time series theory. *Promet - Traffic&Transportation : scientific journal on traffic and transportation research*. Vol. 25, no. 6, p. 533-541.
- [5] PACÁKOVÁ, V. a kolektív: *Štatistika pre ekonómov*. ES VŠE, Bratislava 2003
- [6] PACÁKOVÁ, V. a kolektív: *Štatistika pre ekonómov, Zbierka príkladov*, ES VŠE, Bratislava 2005.
- [7] POLIAK, M. (2013): The relationship with reasonable profit and risk in public passenger transport in the Slovakia. *Ekonomický časopis/Journal of Economics*, 54, issue 7, p. 668 – 684.
- [8] POLIAK, M. – SEMANOVA, S. – POLIAKOVA, A. (2015): Risk allocation in transport public service contracts. *Ekonomski Pregled*. Vol. 66, iss. 4, p. 384-403.
- [9] DOROCIÁKOVÁ, B., POBOČÍKOVÁ, I.: *Zbierka úloh z pravdepodobnosti a matematickej štatistiky*, ŽU Žilina 2005.
- [10] LINCZÉNYI, A.: *Inžinierska štatistika*, Alfa, Bratislava 1973.
- [11] ZODROW, G. (2003) Tax competition and tax coordination in the European Union. *International Tax and Public Finance* 10, p. 651–671.
- [12] JURČEKOVÁ, M., MOLNÁROVÁ I.: *Štatistika s Excelom*. Dostupné na internete:<http://spu.fem.uniag.sk/cvicenia/kskov/palkovic/statistikaB/pr ednasky/statistika_s_excelom.pdf>
- [13] How to Create a Graphical Version in Program Minitab. Dostupné na internete:<<http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-create-a-graphical-version-of-the-1-sample-t-test-in-minitab>>