TECHNICS AND INFORMATICS IN EDUCATION
6th International Conference, Faculty of Technical Sciences, Čačak, Serbia, 28–29th May 2016
TEHNIKA I INFORMATIKA U OBRAZOVANJU
6. međunarodna konferencija, Fakultet tehničkih nauka, Čačak, Srbija, 28–29. maj 2016.

UDK: 378-057.875:519.2 Professional paper

Analysis of the results of the entrance exam and the first colloquium of Business informatics

Vladimir Kraguljac[1], Mladen Janjić[2] and Vera Lazarević[2]
[1] Faculty of Hotel Management and Tourism, Vrnjačka Banja, University of Kragujevac, Serbia
[2] Faculty of Technical Sciences, Čačak, University of Kragujevac, Serbia
e-mail vladimir.kraguljac@kg.ac.rs, mladen.janjic@ftn.kg.ac.rs, vera.lazarevic@ftn.kg.ac.rs

Abstract: *This paper presents the analysis of student performance on the entrance exam and the first colloquium on the subject Business Informatics teaching in the school year 2015/2016 at the Faculty of Hotel Management and Tourism Vrnjačka Banja. The analysis is conducted through several hypotheses. For the implementation of analysis, appropriate statistics and data were used and an adequate interpretation of the results is given.*

Keywords: *hypothesis testing; entrance exam; business informatics; mathematical statistics*

1. INTRODUCTION

This paper analyzes the success of the students in the entrance exam and the first colloquium on the subject of Business Informatics in the academic year 2015/2016 at the Faculty of Hotel Management and Tourism in Vrnjačka Banja. Today, in the digital era, information and communication technologies have a great impact on daily life. Considering that the results of this study can provide new perspectives on the links between secondary school, knowledge shown in the entrance exam and the knowledge shown in the subject which is directly related to information and communication literacy.

The data analyzed in the examples are taken from the official records of Student Services at the Faculty of Hotel Management and Tourism.

2. METHODOLOGY

Data for 107 students were obtained, which represent a statistically large sample due to their number greater than 30. For the analysis the program StatSoft Statistica was used.

2.1. Testing the mean value hypotheses

In a large sample, which is the case here, the arithmetic mean of the sample $\overline{X}$ has a normal distribution $\overline{X} \sim N(\mu_x, \sigma_x / \sqrt{n}) \approx N(\mu_x, s_x / \sqrt{n})$. This, of course, holds for the standardized random variable $\overline{T} = (\overline{X} - \mu) / (\sigma / \sqrt{n}) \sim N(0,1)$, too.
If the hypothesis $H_0(\mu = a)$ is tested, ie. the assertion that the mean value does not differ significantly from the value of a, we do not reject it, with the confidence level of 95%, if $(|\bar{x} - a|$

$\sqrt{n}) / \sigma \leq 1.96$. If this is not fulfilled, then we reject the null hypothesis with a significance level of 5%. To reduce this risk to 1% it is necessary that $(|\bar{x} - a| \sqrt{n}) / \sigma > 2.58$ is satisfied.

2.2. Equality testing of the two main data sets mean values

Let $\bar{X}_1$ and $\bar{X}_2$ be the mean of large number of samples with n_1 and n_2 elements which are subject to normal distributions $N(\mu_1,\sigma_1)$ and $N(\mu_2,\sigma_2)$. If we test the null hypothesis $H_0(\mu_1 = \mu_2)$, ie. that the mean values of the two observed sets are equal, then $\bar{X}_1 - \bar{X}_2$ has normal distribution with parameters $\mu_{\bar{X}1 - \bar{X}2} = 0$ and $\sigma_{\bar{X}1 - \bar{X}2} = \sqrt{(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)} \approx \sqrt{(s_1^2 / n_1 + s_2^2 / n_2)}$. For the standardized variable holds $t = (\bar{X}_1 - \bar{X}_2 - 0) / \sigma_{\bar{X}1 - \bar{X}2} \sim N(0,1)$. We verify $|t| < 1.96 = t_{0.05}$, and for the true inequality we do not reject the null hypothesis with the probability of 95%. Similarly, if we use $t_{0.01} = 2.58$ instead of $t_{0.05}$, the probability is 99%.

2.3. Testing hypotheses regarding the proportions of the main data set

We test the hypothesis that the probability of the observed characteristics of p (success) on the main set of elements is equal to a perceived value p_0, or $H_0(p = p_0)$. From the initial set we separate a sample of n elements, with condition that it is a large sample, i.e. $n > 30$. Let m elements have the required property. Now, the probability of their occurrence in the given sample is $\bar{p} = m / n$. Under these terms, the random variable $\bar{p}$ has a normal distribution, and holds that $\bar{p} \sim N(p,\sqrt{(pq / n)}) \approx N(\bar{p},\sqrt{(\bar{p}\bar{q} / n)})$.
We introduce standardized variable $t = (\bar{p} - p) / \sqrt{(\bar{p}\bar{q} / n)} = (\bar{p} - p_0) / \sqrt{(\bar{p}\bar{q} / n)} \sim N(0,1)$. The decision on acceptance or rejection of the null hypothesis is as follows:
If $|t| < 1.96$, hypothesis is not rejected. If $|t| > 2.58$, hypothesis is rejected as inaccurate because the difference between p and p_0 is highly significant.
In the case of $1.96 < |t| < 2.58$, difference between p and p_0 is substantial, and we can either reject the hypothesis or, better, carry out new testing on a larger sample.

2.4. Testing hypotheses about the equality of the dispersion of two samples of normal sets

When comparing two sets, comparison of their arithmetic means and dispersions is needed. Assuming that for the given sets are with normal distribution, let take two statistically large samples with n_1 and n_2 elements and dispersions of the samples s_1^2 and s_2^2. Let check hypothesis $H_0(\sigma_1^2 = \sigma_2^2)$. Random variable $F = (n_1 s_1^2 (n_2 - 1)) / (n_2 s_2^2 (n_1 - 1)) \approx s_1^2 / s_2^2$ has distribution with two degrees of freedom $k_1 = n_1 - 1$ and $k_2 = n_2 - 1$.
The decision on acceptance or rejection of the proposed hypotheses is made by comparison of the calculated value for F with the table value $(F^{(k1,k2)})_{0.05}$, , i.e. $(F^{(k1,k2)})_{0.01}$, so that if F is smaller than the table value, hypothesis is not rejected, and opposite if it is higher.

2.5. χ2 test for nonparametric hypothesis verification

Tests that examine the probability distribution of some characteristic in the general population are nonparametric hypotheses tests. The value of χ^2 is used as a measure of deviation between the empirical and theoretical frequencies and is calculated as follows:

$$\chi^2 = \sum((f_i - f_{ti})^2 / f_{ti}) = \sum(f_i^2 / f_{ti} - N)$$

where: fi – frequency of the i-th value of random variable in the given sample or the frequency of the i-th class; f_{ti} – corresponding theoretical frequency; $N = \sum f_i = \sum f_{ti}$ – the total number of elements; α – probability, critical coefficient of the concurrence of empirical and theoretical distribution, the risk of accepting the hypothesis (1 - α is the reliability of hypotheses); k – the number of degrees of freedom (for normal distribution k = number of classes – 3).

The decision on rejecting or accepting of the hypotheses is made by reading the value χ_α^2, according to the chosen value α and a corresponding number of degrees of freedom, from the table of values for which is $P(\chi^2 > \chi_\alpha^2) = \alpha$, and then, if the calculated value of χ^2 is greater than χ_α^2, reject the hypothesis. If the calculated value χ^2 is smaller than χ_α^2, hypothesis is not rejected, but for its acceptance testing on a few more samples is required.

2.6. The Kolmogorov test for verification nonparametric hypothesis

This is a test that requires less calculation compared to the χ^2 test, and often can be used instead. Kolmogorov introduced the size D_n as the maximum difference between the empirical distribution function and the assumed theoretical distribution, and found the probability distribution function of the random variable $D_n \sqrt{n}$

$$\lim_{n\to\infty} P(D_n \sqrt{n} < \lambda) = \lim_{n\to\infty} P(D_n < \lambda / \sqrt{n}) = Q(\lambda) = \sum_{k\in[-\infty,\infty]}((-1)^k e^{-2k^2 \lambda^2})$$

The process of deciding whether to accept or reject the hypothesis of presumed concurrence of the empirical and theoretical distributions goes as follows:

1. Assume that a characteristic X has a distribution function F(x) in general population;
2. Select a large enough sample and form a distribution function Fn(x) and find Dn;
3. Choose the required reliability 1 - α;
4. In the table of function Q(λ) values find the value of λα, for which Q(λα) = 1 - α;
5. Do not reject the hypothesis if Dn √n < λα, otherwise rejected it with the risk of α.

3. RESEARCH RESULTS

3.1. Testing hypothesis for the mean of the total score on the entrance exam

We test the hypothesis that the mean total score on the entrance exam is 86 - $H_0(\mu = 86)$. By using StatSoft Statistica we get the following values: n = 107; σ = 4.05; $\bar{x}$ = 85.36; a = 86; α = 5%; c = 85.36 ± 0.78 and p = 0.10.
Now we calculate $(|\bar{x} - a| \sqrt{n}) / \sigma = 1.64 \leq 1.96$. Since the calculated value of 1.64 is less than 1.96, the null hypothesis is not rejected with a significance level of α = 5%. The same conclusion is obtained by comparing assumed value with critical value c = 86.13. Similar conclusion follows from comparison of the P-value p = 0.10 with the value 0.05. Since p > 0.05, the null hypothesis is not rejected with a significance level of α = 5%.

	Descriptive Statistics (podaci in kolokvijum 4.stw)						
Variable	Valid N	Mean	Confidence -95,000%	Confidence 95,000	Minimum	Maximum	Std.Dev.
ukupno prijemni	107	85,35738	84,58095	86,13381	74,40000	95,16000	4,050979

Figure 1. *Descriptive statistics – the complete sample*

	Test of means against reference constant (value) (podaci in kolokvijum 4.stw						
Variable	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
ukupno prijemni	85,35738	4,050979	107	0,391623	86,00000	-1,64091	106	0,103781

Figure 2. *Testing hypotheses about the mean value*

3.2. Testing hypothesis of equality of mean values

We test the equality of mean values of the total number of points at the colloquium for students on budget and self-financing. By using StatSoft Statistica we get the following values (index 1 – students on budget, index 2 – self-financing):

$$n_1 = 57; n_2 = 50; \sigma_1 = 3.39; \sigma_2 = 3.61; \bar{x}_1 = 14.11 \text{ and } \bar{x}_2 = 12.19.$$

We calculate $t = (\bar{X}_1 - \bar{X}_2 - 0) / \sqrt{(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)} = 2.84 > 1.96$. Since the calculated value of 2.84 is greater than 1.96, the hypothesis is rejected with the probability of error 5%. This means that there is a statistically significant difference in the results achieved by students enrolled in the budget and self-financing students.

Variable	T-tests; Grouping: status (podaci in kolokvijum 4.stw) Group 1: Budžet Group 2: Samofinansiranje **Mean Budžet**	Mean Samofinansiranje	t-value	df	p	Valid N Budžet	Valid N Samofinansiranje	Std.Dev. Budžet	Std.Dev. Samofinansiranje	F-ratio Variances	p Variances
ukupno kolokvijum	14,11404	12,19000	2,839641	105	0,005424	57	50	3,391188	3,613876	1,135646	0,642451

Figure 3. *Testing the equality of mean values*

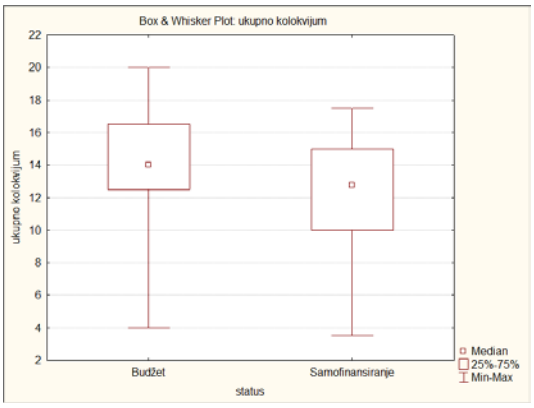

Figure 4. *Box & Whisker Plot*

3.3. Testing hypothesis of the probability of passing the colloquium

We test the hypothesis that the passing level at the colloquium, ie. the number of students with more than 10 points, is 75% - $H_0(p = 0.75)$. By using StatSoft Statistica we get the following values: $n = 107$; $m = 86$; $\bar{p} = m / n = 86 / 107 = 0.80$; $\bar{q} = 1 - \bar{p} = 0.20$ and $p_0 = 0.75$. We calculate $t = (\bar{p} - p_0) / \sqrt{(\bar{p}\bar{q} / n)} = 1.40 < 1.96$. Since the calculated value of 1.40 is less than 1.96 we have no reason to reject the hypothesis.

3.4. Testing hypotheses of the equality of two samples dispersion

We compare number of points at the colloquium for the students on budget and for self-financing students.

Variable	status=Budžet Descriptive Statistics (Spreadsheet in kolokvijum 4.stw) Valid N	Mean	Minimum	Maximum	**Std.Dev.**
ukupno kolokvijum	57	14,11404	4,000000	20,00000	3,391188

Figure 5. *Descriptive statistics - budget*

Variable	status=Samofinansiranje Descriptive Statistics (Spreadsheet in kolokvijum 4.stw) Valid N	Mean	Minimum	Maximum	**Std.Dev.**
ukupno kolokvijum	50	12,19000	3,500000	17,50000	3,613876

Figure 6. *Descriptive statistics - self-financing*

With $n_1 = 57$; $n_2 = 50$; $\sigma_1 = 3.40 = s_1$ and $\sigma_2 = 3.61 = s_2$ we have that

$$F = (n_1\, s_1^2\, (n_2 - 1)) / (n_2\, s_2^2\, (n_1 - 1)) = 0.88.$$

The variable F has a distribution with two degrees of freedom $k_1 = n_1 - 1 = 56$ and $k_2 = n_2 - 1 = 49$. In the table we have no value $(F^{(56,49)})_{0.05}$, but we see that it is in the range from 1.44 to 1.76. We test the hypothesis $H_0(\sigma_1^2 = \sigma_2^2)$, and since calculated F is less than $(F^{(56,49)})_{0.05}$ hypothesis is not rejected with significance level of $\alpha = 5\%$. This means that the dispersion around the mean value is equal for both the budget students and for the self- financing students. In other words, the results of this research are equally reliable for both of those samples and in the future we cannot expect significant changes in the ratio of their achievements, ie. budget students will always achieve better results.

3.5. Testing the hypothesis of independance

Here, we use χ^2 test and make hypothesis that the type of school and its status are independent of each other. By using StatSoft Statistica we get frequencies and values (Figures 7 and 8).

Summary Frequency Table (podaci in kolokvijum 4.stw)
Marked cells have counts > 10

tip škole	status Budžet	status Samofinansiranje	Row Totals
tehnička	13	7	20
ostale	14	9	23
ekonomska	14	15	29
gimnazija	8	8	16
ugostiteljsko-turistička	6	9	15
medicinska	2	0	2
trgovačka	0	2	2
All Grps	57	50	107

Figure 7. *Frequency of data for different types of schools*

	Statistics: tip škole(7) x status(2)		
Statistic	Chi-square	df	p
Pearson Chi-square	7,093856	df=6	p=,31226
M-L Chi-square	8,649167	df=6	p=,19430

Figure 8. χ^2 *test – type of the school and status*

From these Figures 7 and 8 we get the values $\chi^2 = 7.09$ and $p = 0.31$. As for the P-value holds $p > 0.05$, the hypothesis is not rejected with a significance level of $\alpha = 5\%$. This means that there is no significant correlation between types of the school which students attended and enrollment status.

3.6. Testing the hypothesis that the achieved results are in accordance with the normal distribution

The Kolmogorov test is similar to χ^2 test and the hypotheses are similar to the former – if the high school points, points obtained at the entrance exam, total points at the enrollment, the theoretical part and the practical part of the colloquium follow the normal distribution.

Variable	Tests of Normality (poda		
	N	max D	K-S p
bodovi srednja	107	0,087150	p > .20
bodovi prijemni	107	0,073172	p > .20
ukupno prijemni	107	0,092580	p > .20
bodovi teorija	107	0,160116	p < ,01
bodovi praktično	107	0,119716	p < ,10
ukupno kolokvijum	107	0,101548	p > .20

Figure 9. *Test whether the data follow a normal distribution*

By looking at the table of values for the function Q(λ) we find that the limit is $\lambda_\alpha \approx 1.36$. After checking whether the requirement $D_n \sqrt{n} < \lambda_\alpha$ is fulfilled, we conclude that only the points at the theoretical part of the colloquium do not have a normal distribution. The same conclusion would come by observing P-values, because only for the points on the theoretical part of the colloquium we have $p < 0.05$. Further, as an example, the detailed data table is given, which is obtained from testing the same hypothesis for a single value - the total number of points at the colloquium observed by classes of data.

Upper Boundary	Variable: ukupno kolokvijum, Distribution: Normal (podaci in kolokvijum 4.stw) Kolmogorov-Smirnov d = 0,06224, Chi-Square = 18,71057, df = 12, p = 0,09576								
	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 2,40000	0	0	0,00000	0,0000	0,14703	0,1470	0,13741	0,1374	-0,14703
3,80000	1	1	0,93458	0,9346	0,34173	0,4888	0,31938	0,4568	0,65827
5,20000	2	3	1,86916	2,8037	0,92725	1,4160	0,86659	1,3234	1,07275
6,60000	5	8	4,67290	7,4766	2,16882	3,5848	2,02693	3,3503	2,83118
8,00000	3	11	2,80374	10,2804	4,37296	7,9578	4,08688	7,4372	-1,37296
9,40000	5	16	4,67290	14,9533	7,60087	15,5587	7,10362	14,5408	-2,60087
10,80000	9	25	8,41121	23,3645	11,38916	26,9478	10,64408	25,1849	-2,38916
12,20000	10	35	9,34579	32,7103	14,71182	41,6596	13,74936	38,9342	-4,71182
13,60000	17	52	15,88785	48,5981	16,38287	58,0425	15,31110	54,2453	0,61713
15,00000	23	75	21,49533	70,0935	15,72767	73,7702	14,69876	68,9441	7,27233
16,40000	10	85	9,34579	79,4393	13,01633	86,7865	12,16480	81,1089	-3,01633
17,80000	14	99	13,08411	92,5234	9,28668	96,0732	8,67914	89,7880	4,71332
19,20000	6	105	5,60748	98,1308	5,71184	101,7850	5,33817	95,1262	0,28816
20,60000	2	107	1,86916	100,0000	3,02851	104,8136	2,83039	97,9566	-1,02851
< Infinity	0	107	0,00000	100,0000	2,18644	107,0000	2,04340	100,0000	-2,18644

Figure 10. *The Kolmogorov test and χ^2 test*

Here it is $D_n \sqrt{n} = 0.64 < 1.36 = \lambda_\alpha$, so the hypothesis is not rejected with a significance level of $\alpha = 5\%$.

4. CONCLUSION

Results of this analysis lead to some new conclusions. It is evident that the entrance exam is well devised, because it does not favor any type of school from which the candidates come. Then, it is encouraging that the average number of points on the entrance exam, 86 of 100, show that the colleges located outside the large university centers can attract good quality students. This further provides opportunity for better quality work with them, and then both budget and self-financing students achieve good consistent results at the colloquium with high success rate. The success of the budget students is slightly better, as it would be expected.
As for the colloquium on the subject of Business Informatics, it can be noted that there are some deviations from the expected achievement in the theoretical part. It is positive that the achievement at practical work, which is particularly important for further study and subsequent integration in the business environment, meets the high expectations.

REFERENCES

[1] Janjić, M., Lazarević, V. (2014). Kovarijansa i korelacija srednjoškolskog uspeha i pokazanih rezultata na prijemnom ispitu. Technics and Informatics in Education, 5th International Conference, FTS Čačak, 30-31st May 2014.
[2] Lazarević, V., Đukić, M. (2010). Inženjerska Matematika. Čačak: Tehnički Fakultet.
[3] Merkle, M. (2010) Verovatnoća i statistika za inženjere i studente tehnike. Beograd: Akademska misao.
[4] Vukadinović, S. (1988) Elementi teorije verovatnoće i matematičke statistike. Beograd: Privredni pregled.