STATISTICAL ANALYSIS OF THE TEST VARIANTS IN ADMISSION PROCESS

Jindřich Klůfa

Abstract

The aim of this paper is to study dependence of the results of entrance examinations in mathematics on test variants. The paper contains an analysis of the differences of number of points in the test in mathematics between test variants, which were used in the entrance examinations at the Faculty of Informatics and Statistics at University of Economics in Prague in 2015. The differences may arise due to the varying difficulty of variants, but also because of the different level of knowledge of students who write these variants. This problem we shall study in present paper. Results of this analysis will be used for improvement of the entrance examinations at University of Economics in coming years. To increase the homogeneity of the test variants, the database of examples of mathematics will be extended and divided into multiple groups. This problem will be solved in the framework of the project Internal development competition number IRS/MF/F4/3/2016.

Key words: Entrance examination in mathematics, statistical methods, differences between test variants

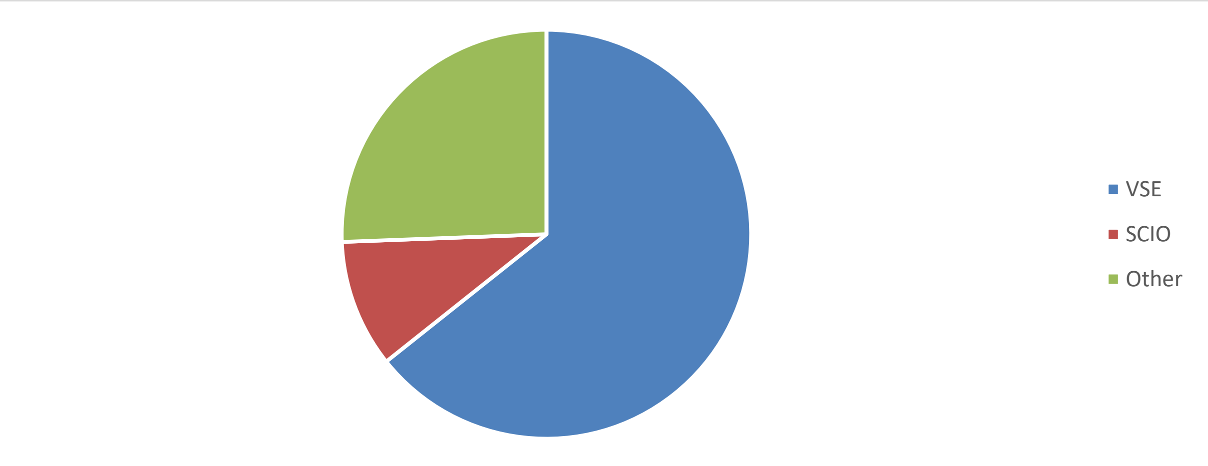
JEL Code: C12, I21

Introduction

Most students of the Prague University of Economics have been accepted to study on the basis of tests in mathematics and language tests. The math tests are prepared by the Department of Mathematics of the Faculty of Informatics and Statistics. These tests are the multiple choice question tests. Advantages and disadvantages of multiple choice question tests are described in (Klůfa, 2012), (Zhao, 2006), (Klůfa, 2013), (Premadasa, 1993). In (Klůfa, 2012) it was shown that risk of success of students with lower performance levels in multiple choice question tests in mathematics is negligible.

In this paper we shall analyse admission process at the Faculty of Informatics and Statistics in 2015. Students of the Faculty of Informatics and Statistics at University of Economics in Prague are accepted to study on the basis of tests in mathematics and English, which is used at University of Economics in Prague (VSE tests), or on the basis of the national comparative exams - the tests of general academic prerequisites (SCIO tests), or without entrance examinations (on the basis of results in mathematics and English at grammar school etc.). Most students of the Faculty of Informatics and Statistics have been accepted to study on the basis of tests in mathematics and English – see Fig. 1.

Fig. 1: Ways of acceptance students at the Faculty of Informatics and Statistics



Source: Own construction

For comparison of these ways of acceptance students see (Klůfa, 2015) - worst results from the point of view number of points in test in mathematics achieved by students which were accepted on the base of the national comparative exams (SCIO tests). Similar problem is solved in (Coufal and Tobíšek, 2015), (Klůfa, 2008), (Brožová and Rydval, 2013), (Kubanová and Linda, 2012) - students should be accepted on the basis of own admission.

The tests in mathematics at the Faculty of Informatics and Statistics have 10 questions for 5 points and 5 questions for 10 points. Therefore, the number of points in the test in mathematics can be: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, ... ,90, 95, 100. Test variants in mathematics are generated from a database created by the Department of Mathematics. The effort is to choose variants, which are equally difficult for students, but the difficulty of test variants for students is poorly measured.

1 Entrance exams at the Faculty of Informatics and Statistics

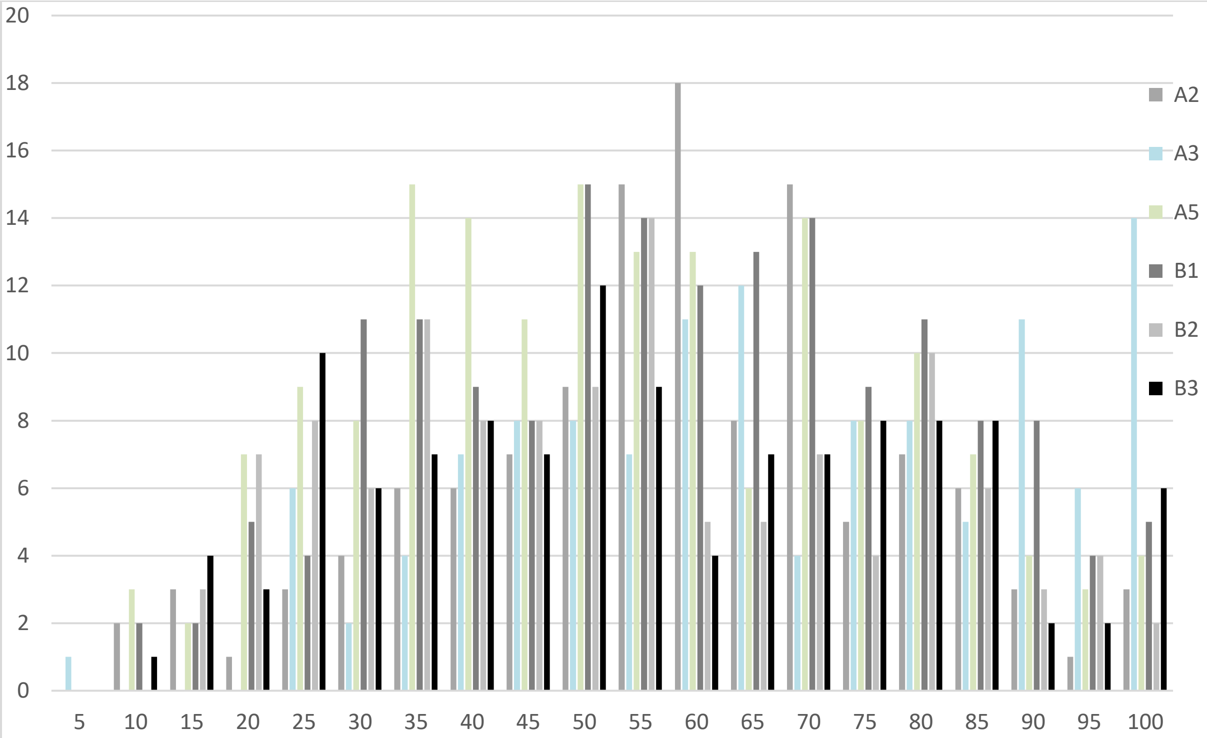
For entrance examinations in mathematics at the Faculty of Informatics and Statistics at University of Economics in Prague in 2015 it was used six test variants, denoted A2, A3, A5, B1, B2 and B3. The analysed data are the results in mathematics of the entrance examinations in 2015 of 814 students – see Tab. 1 and Fig. 2.

Tab. 1: Descriptive statistics for number of points in mathematics

Test variant	A2	A3	A5	B1	B2	B3
Frequency n_i	122	122	166	165	120	119
Average	58.0328	66.5984	53.6145	58.0909	53.5417	55.5462
Median	60	65	50	60	52.5	55
Modus	60	100	35	50	55	50
Variance	384.9411	524.0770	478.0686	479.4124	511.5109	559.2330
Std. Deviation	19.620	22.893	21.865	21.895	22.617	23.648
Kurtosis	-0.0585094	-0.8169274	-0.7486692	-0.7695973	-0.9520309	-0.9547594
Skewness	-0.2513542	-0.2083088	-0.1589383	-0.0531730	-0.2120466	-0.1026342

Source: Own construction

Fig. 2: Distribution of number of points in test in mathematics in 2015 – test variants A2, A3, A5, B1, B2, B3 (histogram)



Source: Own construction

Now we shall test null hypothesis "mean number of points in test variants A2, A3, A5, B1, B2, B3 is the same", i.e.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \tag{1}$$

where $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ is mean number of points in mathematics in test variants A2, A3, A5, B1, B2, B3. To verify the validity of the hypothesis we use ANOVA – see (Rao, 1973). In the first step we verify assumption (the same variance of number of points in test variants A2, A3, A5, B1, B2, B3) of this method by Bartlett's test. The hypothesis "variance of number of points in test variants A2, A3, A5, B1, B2, B3 is the same" is not rejected at 1% significance level, assumption of ANOVA can be considered to have been met. Results of ANOVA we got with MS Excel (e.g. (Marek, 2013)) – see Tab. 2. Since

$$F = 6.157 > 2.225,$$

null hypothesis (1) is rejected at 5% significance level. There are some differences between the test variants (the differences between average number of points in test variants A2, A3, A5, B1, B2, B3 (see Tab. 1) are statistically significant).

Tab. 2: Results of ANOVA

Source of variability	*Sum of Squares*	*Degrees of freedom*	*Fraction*	*F*	*P value*	*F crit*
Test variants	15024.19994	5	3004.83999	6.156655844	1.30656E-05	2.225185
Rezidual	394355.4377	808	488.06366			
Sum	409379.6376	813				

Source: Own construction

Finally we shall study which pairs of averages differ significantly. We use Scheffé's method – see e.g. (Anděl, 1978). Pairs of averages differ significantly if absolute value of difference in averages exceeds critical value (see Tab. 1 and Tab. 2)

$$\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\ x\ 5\ x\ 488.06366\quad x\ 2.225185} \tag{2}$$

Tab. 3: Absolute value of differences between average number of points in test variants A2, A3, A5, B1, B2, B3

Test variant	A2	A3	A5	B1	B2
A2					
A3	8.57				
A5	4.42	12.98*			
B1	0.06	8.51	4.48		
B2	4.49	13.06*	0.07	4.55	
B3	2.49	11.05*	1.93	2.55	2.00

Source: Own construction * Significant difference for α=0.05 (Scheffé's method)

From Tab. 3 it is seen that a significant difference is at 5% significant level between A3 and A5, A3 and B2, A3 and B3. All other pairs of averages are not significantly different. Greatest significant difference is between variant A3 and B2.

2 Difference between test variants A3 and B2

The differences between test variants may arise due to the varying difficulty of variants, but also because of the different level of knowledge of students who write these variants. This problem we shall study in following text.

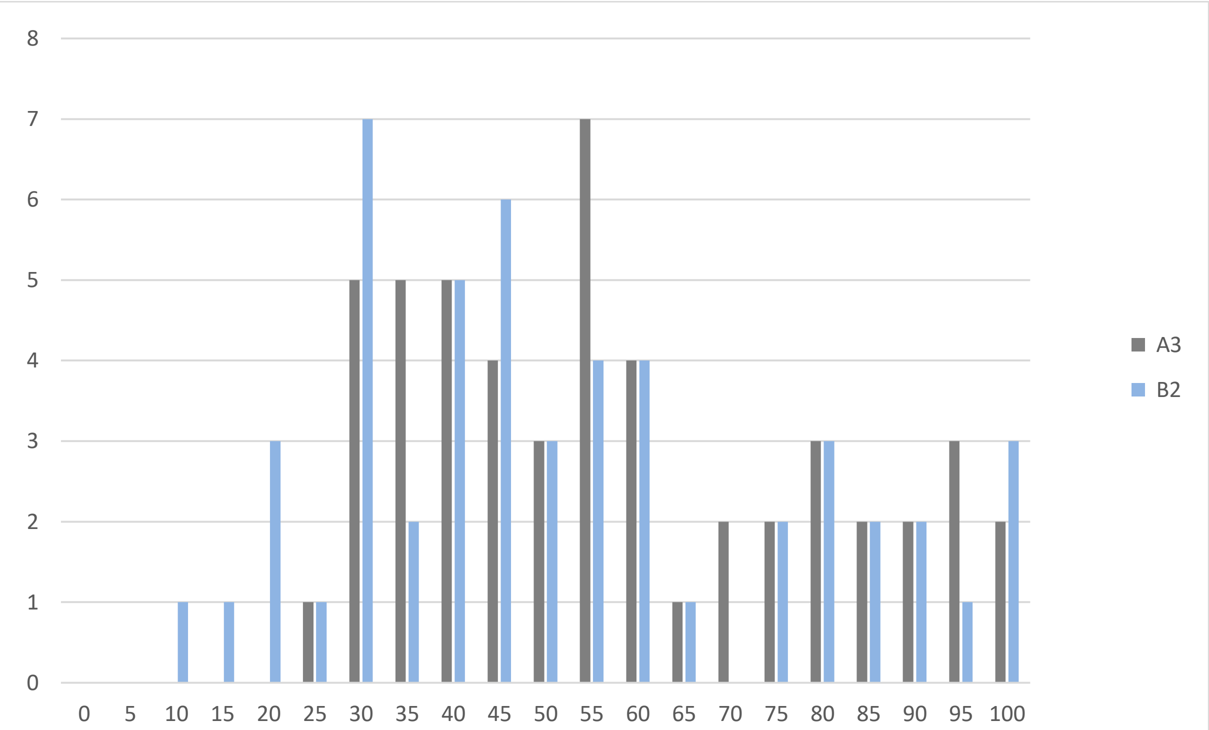
Therefore we shall now study results in mathematics of the other two groups of students, which we obtained in project "Entrance exams practice" in 2016. The analysed data are the results in mathematics of 102 students – see Tab. 4 and Fig. 3.

Tab. 4: Descriptive statistics for number of points in mathematics

Test variant	A3	B2
Frequency n_i	51	51
Average number of points	57.54901961	52.35294118
Median	55	45
Modus	55	30
Variance	474.372549	592.3529412
Std. Deviation	21.780	24.338
Kurtosis	-0.921162717	-0.71750832
Skewness	0.476547907	0.459545831

Source: Own construction

Fig. 3: Distribution of number of points in test in mathematics – test variants A3, B2 (histogram)



Source: Own construction

Average number of points in mathematics in test variant A3 is $\bar{x}_{A3} = 57.55,$ average number of points in mathematics in test variant B2 is $\bar{x}_{B2} = 52.35$ (Tab. 4). Now we shall test if the difference between average number of points in test variants A3 and B2 is statistically significant, i.e. null hypothesis

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

opposite $H_1 : \mu_1 \neq \mu_2$, where μ_1, μ_2 is mean number of points in mathematics in test variants A3 and B2. In the first step we verify the same variance of number of points in test variants A3, B2 by Fisher-Snedecor's test. We shall test $H_0 : \sigma^2_{A3} = \sigma^2_{B2}$ opposite $H_1 : \sigma^2_{B2} > \sigma^2_{A3}$. Since (see Tab. 4)

$$F = \frac{592{,}35}{474{,}37} = 1.25 < F_{0.05}(50{,}50),$$

where $F_{0.05}(50{,}50) = 1.60$ is critical value of Fischer-Snedecor distribution with 50 and 50 degrees of freedom, the hypothesis "variance of number of points in test variants A3 and B2 is the same" is not rejected at 5% significance level (p value is 0.218). To verify null hypothesis (3) we shall use Student's t test for the same variance. Results we got with MS Excel (e.g. (Marek, 2013)) – see Tab. 5. Since

$$|t| = 1.136 < t_{0.05}(100),$$

where $t_{0.05}(100) = 1.984$ is critical value of Student t distribution with 100 degrees of freedom ($P(|t| > t_\alpha(f)) = \alpha$) null hypothesis (3) is not rejected at 5% significance level (this hypothesis is not rejected also at 25% significance level, p value is 0.2586 – see Tab. 4). Average number of points $\bar{x}_{A3} = 57.55$ and average number of points $\bar{x}_{B2} = 52.35$ are not differ significantly.

For other two groups of students there are not significant differences between test variants A3 and B2 – see Student's t test. Therefore, significant differences between tests variants A3 and B2 may arise as a result of the different levels of knowledge of the students, who wrote these variants.

Tab. 5: Student‘s t–test for the same variance

Test variant	A3	B2
Average number of points	57,54901961	52,35294118
Variance	474,372549	592,3529412
Frequency	51	51
Common variance	533,3627451	
Degrees of freedom	100	
t Stat	1,136146799	
p value	0,258610608	
t crit	1,983971519	

Source: Own construction

Conclusion

The differences between average number of points in mathematics in 2015 in test variants A2, A3, A5, B1, B2, and B3 are statistically significant – see ANOVA. Greatest significant difference is between variant A3 and B2. But for other two groups of students there are not significant differences between test variants A3 and B2 – see Student’s t test. From results of this paper it follows that these significant differences between tests variants may arise due to different level of knowledge of students who wrote these variants.

Significant changes in test variants in mathematics in the coming years are not needed. But increase the homogeneity test variants would be very useful. Therefore the database created by the Department of Mathematics will be further modified (the database will be expanded and divided into more of the groups) in the framework of the project Internal development competition number IRS/MF/F4/3/2016.

Acknowledgment

The paper was processed with contribution of long term support of scientific work on Faculty of Informatics and Statistics, University of Economics, Prague (IP 400040).

References

Anděl, J. (1978) *Matematická statistika*. Praha: SNTL/ALFA.

Brožová, H., Rydval, J. (2013) *Analysis of the exam test quality, Efficiency and Responsibility in Education, Proceedings of the 10th International Conference*, Prague, pp. 47–54.

Coufal, J., Tobíšek, J. (2015) *Mathematics and e-learning, The 9th International Days of Statistics and Economics*, Prague, pp. 270-278.

Klůfa, J. (2008). Dodge-Roming AOQL plans for inspection by variables from numerical point of view. *Statistical Papers*, 49(1), 1-13.

Klůfa, J. (2012) *Tests from probability point of view, Efficiency and Responsibility in Education, Proceedings of the 9th International Conference*, Prague, pp. 229–233.

Klůfa, J. (2013) *Comparison of entrance examinations in mathematics, Efficiency and Responsibility in Education, Proceedings of the 10th International Conference*, Prague, pp. 270–275.

Klůfa, J. (2015) *Analysis of entrance examinations, Efficiency and Responsibility in Education, Proceedings of the 12th International Conference*, Prague, pp. 250–256.

Kubanová, J., Linda, B. (2012) Relation between results of the learning potential tests and study results, *Journal on Efficiency and Responsibility in Education and Science*, vol. 5, no 3, pp. 125-134.

Marek, L. (2013) *Statistika v příkladech,*, Prague: Professional Publishing

Premadasa, I. (1993) A reappraisal of the use of multiple-choice questions, *Medical Teacher,* vol. 15, no. 2-3, pp. 237-242.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, New York: John Wiley

Zhao, Y. (2006) How to design and interpret a multiple-choice-question test: A probabilistic approach, *International Journal of Engineering Education*, vo l. 22, no. 6, pp. 1281-1286.

Contact

Prof. RNDr. Jindřich Klůfa, CSc.
University of Economics in Prague,
Department of Mathematics,
W.Churchill Sq.4, 13067 Prague 3
klufa@vse.cz