

# COMPARISON OF THE TEST VARIANTS IN ENTRANCE EXAMINATIONS

Jindřich Klůfa✉

✉Department of Mathematics, Faculty of Informatics and Statistics, University of Economics, W. Churchill Sq.4, Prague, 13067, Czech Republic, +420 224 094 244, klufa@vse.cz

## Highlight

- *Dependence of the test results on the test variants*

## Abstract

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 Business Administration at University of Economics in Prague in 2015. The differences may arise due to the varying difficulty of variants for students, but also because of the different level of knowledge of students who write these variants. This problem we shall study in present paper. The aim of this paper is to study dependence of the results of entrance examinations in mathematics on test variants. The results obtained will be used for further improvement of the admission process at University of Economics.

## Keywords

Entrance examinations, test variants, mathematics, statistical methods

## Article type

*Full research paper*

## Article history

*Received: September 19, 2016*

*Received in revised form: November 21, 2016*

*Accepted: November 28, 2016*

*Available on-line: December 15, 2016*

*Klůfa J. (2016) "Comparison of the Test Variants in Entrance Examinations", Journal on Efficiency and Responsibility in Education and Science, Vol. 9, No. 4, pp. 111-116, online ISSN 1803-1617, printed ISSN 2336-2375, doi: 10.7160/eriesj.2016.090404.*

## Introduction

Students of the Faculty of Business Administration are 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 (Klůfa, 2012), (Zhao, 2006), (Klůfa, 2013), (Premadasa, 1993), (Klůfa, 2015b). Multiple choice question tests are suitable for entrance examinations at university. These tests are objective, results can be evaluated easily for large number of students. On the other hand, a student can obtain certain number of points in the test purely by guessing the right answers. This problem is addressed in education research Premadasa (1993), Zhao (2005, 2006) - the probabilistic analysis shows that the optimum number of choices of answers for the multiple choice question tests is four, and for a four-choice question test, increasing from 8 questions to 18 and 48 questions reduces the probability of obtaining a good result by pure guesswork from about 5% to below 1% and 0.01%, respectively. In Klůfa (2012) it was shown that risk of success of students with lower performance levels in entrance exams at

University of Economics in Prague is negligible (approximately one student in million successfully makes the entrance exams by pure guessing the answers), i.e. the multiple choice question tests are optimal for admission process. The multiple choice question tests from probability point of view with similar results are also in Klůfa (2013).

The tests in mathematics at the Faculty of Business Administration at University of Economics in Prague have 10 questions for 5 points and 5 questions for 10 points, i.e. 100 points total. Questions are independent. Each question has 5 answers, one answer is correct, wrong answer is not penalized. 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. Test variants, which were

used in the entrance examinations at the Faculty of Business Administration at University of Economics in Prague in 2015 we can find in Klůfa and Langhamrová (2015), part of one of these variants is in Figure 2 in Appendix. The database of the Department of Mathematics is divided into more of the groups, e.g. goniometric equations, sequences etc. From the selected groups is generated a question. Finally, the generated variants are chosen which are used for entrance examinations. The effort is to choose variants, which are equally difficult for students.

The aim of this paper is to analyse 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 Business Administration in 2015. Similar problems are solved in Brožová and Rydval (2013), Hrubý (2013), Kaspříková (2012), Mošna (2013), Klůfa (2015c), Kubanová and Linda (2012), Coufal and Tobíšek (2015), Otavová and Sýkorová (2014). The dependence of study results and results of the entrance exams in mathematics is solved in Kubanová and Linda (2012). Analogous problem (the dependence of study results in mathematics on ways of acceptance students at university) is analysed in Klůfa (2015c). From results of these papers follows that students should be accepted to study on the basis of own admission process. University study results as related to the admission exam results we can find also in Kučera, Svatošová and Pelikán (2015). Analysis of the study results in basic courses in mathematics at University of Economics is in Kaspříková (2012) and Otavová and Sýkorová (2014). There is studied whether the score from final test depends on the score from mid-term test. Obtained results show that dependence between the score from final test and the score from mid-term test exists. The exam results in mathematics at Czech University of Life Sciences in Prague from the last 13 years have been analysed in Brožová and Rydval (2013). The reasons of low grades of students are discussed in this paper. Mathematics is generally said to be one of the unpopular school subjects. Popularization

of mathematics (e-learning) is described in Coufal and Tobíšek (2015). E-learning and teaching of mathematics is also in Mošna (2013).

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 present paper. The results obtained will be used to further improve of the preparation of test variants in coming years.

This paper is an extended version of the paper Klůfa (2016) – results of other group of students, obtained in project “Entrance exams practice” in 2016, was analyzed.

**Material and Methods**

The analysed data are the results of the entrance examinations of 1514 students in mathematics at the Faculty of Business Administration in 2015. Six test variants, denoted A0, A8, A9, B0, B4, B6, were used for the entrance examinations in mathematics at the Faculty of Business Administration in 2015, other test variants were not used at this faculty. Differences between genders are not analysed in present paper.

On the other hand, the Department of Mathematics organizes preparatory courses for entrance examinations in mathematics. The results of one randomly selected parallel class (17 students) of these courses in 2016 will be analysed in this paper as well.

Furthermore, other results of 58 students, which were obtained in project “Entrance exams practice” in 2016, will be analysed in present paper.

For study the differences of number of points in the test in mathematics between 2 test variants we shall use paired t-test and t-test for independent samples. Statistic *t* for paired test is

$$t = \frac{\bar{d}}{s_d} \sqrt{n}, \tag{1}$$

where  $d_i = x_i - y_i$ , and  $x_i, y_i$  is number of points in the test in mathematics of a student *i* in 1st and 2nd test variant,  $\bar{d}$  is average of values  $d_i$ ,  $s_d$  is standard deviation,  $n$  is sample size (17). When

$$|t| > t_\alpha(n-1), \tag{2}$$

where  $t_\alpha(n-1)$  is critical value of student t distribution with  $(n-1)$  degrees of freedom, the hypothesis “mean number of points in 2 test variants is the same” is rejected at significance level  $\alpha$ .

Statistic *t* for t-test for independent samples (under the same variance of samples) is

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3}$$

where  $\bar{x}, \bar{y}$  is average number of points in the test in mathematics in 1st and 2nd sample,  $n_1, n_2$  is sample size in 1st and 2nd sample (in our case is  $n_1 = n_2 = 29$ ) and  $s$  is standard deviation ( $s_x, s_y$  is standard deviations in 1st and 2nd sample) given by relation

$$s = \sqrt{\frac{1}{n_1+n_2-2} [(n_1-1)s_x^2 + (n_2-1)s_y^2]}. \tag{4}$$

When

$$|t| > t_\alpha(n_1 + n_2 - 2), \tag{5}$$

where  $t_\alpha(n_1 + n_2 - 2)$  is critical value of student t distribution with  $(n_1 + n_2 - 2)$  degrees of freedom, the hypothesis “mean number of points in 2 test variants is the same” is rejected at significance level  $\alpha$ .

For comparison of 6 test variants at the Faculty of Business Administration in 2015 we shall use ANOVA and Scheffé’s method. We shall verify the validity of the null hypothesis: mean number of points in test variants A0, A8, A9, B0, B4, B6 is the same. When the test statistic (Rao, 1973)

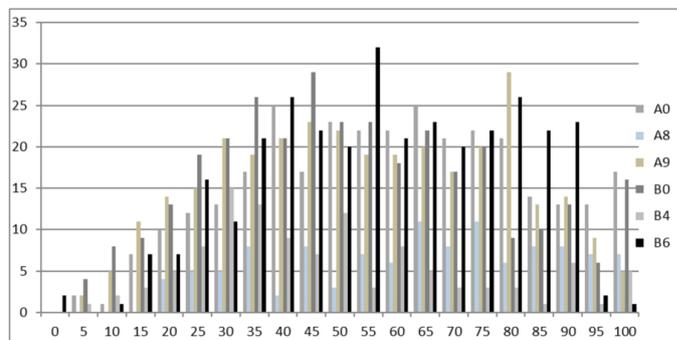
$$F > F_\alpha(k-1, n-k), \tag{6}$$

where  $F_\alpha(k-1, n-k)$  is critical value of Fischer-Snedecor distribution with  $(k-1)$  and  $(n-k)$  degrees of freedom, the hypothesis is rejected at significance level  $\alpha$ . In our case is  $k=6$  (number of variants) and  $n=1514$  (sample size for ANOVA).

**Results**

**Differences between the test variants**

The results of the entrance examinations of 1514 students in mathematics at the Faculty of Business Administration in 2015 are in Table 8, Table 9, Table 10, Table 11, Table 12, Table 13 in Appendix. Now we shall compare distributions of number of points in the test in mathematics in test variants A0, A8, A9, B0, B4, B6 - see Figure 1 and Table 1.



**Figure 1: Distribution of number of points in test in mathematics in 2015 – test variants A0, A8, A9, B0, B4, B6 (histogram) (source: own calculation)**

Test variant	Frequency $n_i$	Average number of points	Variance
A0	317	59.23	543.94
A8	114	64.17	540.23
A9	318	54.61	559.70
B0	327	52.54	584.03
B4	113	47.92	544.97
B6	325	57.31	462.71

**Table 1: Distribution of number of points in test – test variants A0, A8, A9, B0, B4, B6 (source: own calculation)**

We shall test null hypothesis “the differences between average number of points in test variants A0, A8, A9, B0, B4, B6 in Table 1 are not statistically significant”.

To verify the validity of the hypothesis we use ANOVA. In the first step we verify assumption of this method by Bartlett’s test, i.e. we verify the hypothesis “variance of number of points in test variants A0, A8, A9, B0, B4, B6 is the same”. Test statistic  $B$  (see e.g. Anděl (1978)) is  $B = 4.9$ . Critical value of  $\chi^2$  distribution for 5 degrees of freedom and significance level  $\alpha = 0.05$  is  $\chi^2_{0.05}(5) = 11.1$ . Since  $B < 11.1$ , the hypothesis “variance of number of points in test variants A0, A8, A9, B0, B4, B6 is the same” is not rejected at 5% significance level, assumption of ANOVA can be considered to have been met.

Source of variability	Sum of Squares	Degrees of freedom	Fraction	F	p value	F crit
Test variants	23365.02	5	4673.00	8.68	3.99E-08	2.22
Residual	811706.13	1508	538.27			
Sum	835071.15	1513				

**Table 2: Results of ANOVA (source: own calculation)**

Results of ANOVA we got with MS Excel (Marek, 2013) – see Table 2. Since

$$F = 8.68 > 2.22,$$

the null hypothesis is rejected at 5% significance level. There are some differences between the test variants, the differences between average number of points in test variants A0, A8, A9, B0, B4, B6 in Table 1 are statistically significant.

Finally we shall study which pairs of averages differ significantly. We use Scheffé’s method (Anděl, 1978). Pairs of averages differ significantly if absolute value of difference in averages exceeds critical value

$$\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) \times 5 \times 538.27 \times 2.22} \quad (7)$$

where 538.27 is the residual variance and 2.22 is the critical value from Table 2.

Test variant	A0	A8	A9	B0	B4	B6
A0		4.94	4.62	6.69	11.31*	1.92
A8			9.56*	11.63*	16.25*	6.86
A9				2.07	6.69	2.70
B0					4.62	4.77
B4						9.39*
B6						

\*Significant difference for  $\alpha=0.05$  (Scheffé’s method)

**Table 3: Absolute value of differences between average number of points in test variants A0, A8, A9, B0, B4, B6 (source: own calculation)**

From Table 3 it is seen that a significant difference is at 5% significant level between A0 and B4, A8 and A9, A8 and B0, A8 and B4, B4 and B6. All other pairs of averages are not significantly different. Greatest significant difference is between the test variants A8 and B4.

### Difference between A8 and B4 – paired t test

Significant differences between test variants may arise due to the varying difficulty of variants for students, but also because of the different level of knowledge of students who write these variants. Therefore we shall now study results of the same group of students – see results of 17 students in preparatory course for entrance examinations in 2016 in Table 4.

Student	A8	B4	d
1	100	95	5
2	60	45	15
3	70	80	-10
4	35	20	15
5	40	35	5
6	25	20	5
7	40	45	-5
8	60	50	10
9	55	50	5
10	45	50	-5
11	60	55	5
12	55	60	-5
13	45	45	0
14	70	80	-10
15	45	40	5
16	55	40	15
17	80	85	-5

**Table 4: Number of points in mathematics in test variants A8 and B4 (source: own calculation)**

From Table 4 we have average number of points in mathematics in test variants A8  $\bar{x}_{A8} = 55.29$  and average number of points in mathematics in test variants B4  $\bar{x}_{B4} = 52.65$ .

Now we shall test null hypothesis “the difference between these average number of points in test variants A8, B4 is not statistically significant”.

We have two results for the same student. It means that the samples in Table 4 are not independent. Therefore, to verify the validity of the hypothesis we use paired t test. According to (1) we have

$$t = 1.31$$

Critical value of  $t$  distribution for 16 degrees of freedom and significance level  $\alpha = 0.05$  is  $t_{0.05}(16) = 2.12$ . Since

$$|t| < 2.12,$$

the null hypothesis is not rejected at 5% significance level. Because  $t_{0.20}(16) = 1.34$ , this hypothesis is not rejected also at 20% significance level. The difference between average number of points in test variants A8 and B4 in preparatory course for entrance examinations in 2016 is not statistically significant.

### Difference between A8 and B4 – t test for independent samples

Now we shall compare other results of 58 students, which were obtained in project “Entrance exams practice” in 2016

(two different groups of students, each group has 29 students, i.e.  $n_1 = n_2 = 29$ ). These students wrote test variants A8 and B4 once more, results are in Table 14 in Appendix, descriptive statistics for distributions of number of points in the test in mathematics in test variants A8 and B4 are in Table 5.

Test variant	A8	B4
Average number of points	47.931	40.517
Median	45	30
Modus	15	25
Variance	588.42	572.04
Kurtosis	-0.854	-0.616
Skewness	0.181	0.714

**Table 5: Descriptive statistics for number of points in mathematics in test variants A8 and B4 (source: own calculation)**

From Table 5 we have average number of points in mathematics in test variants A8  $\bar{x}_{A8} = 47.93$  and average number of points in mathematics in test variants B4  $\bar{x}_{B4} = 40.52$ .

Now we shall test null hypothesis “the difference between these average number of points in test variants A8, B4 is not statistically significant”.

We have results of two different groups of students, i.e. the results are independent. Therefore, to verify the validity of the hypothesis we use t-test for independent samples. In the first step we verify assumption of the same variance of samples by Fisher-Snedecor F-test. The hypothesis “variance of number of points in test variants A8 and B4 is the same” is not rejected at 5% significance level (p-value is 0.47), assumption of the t-test for independent samples can be considered to have been met.

Results of the t-test for independent samples we got with MS Excel (Marek, 2013) – see Table 6. According to (3) we have

$$t = 1.172$$

Critical value of  $t$  distribution for 56 degrees of freedom and significance level  $\alpha = 0.05$  is  $t_{0.05}(56) = 2.003$ . Since

$$|t| < 2.003,$$

the null hypothesis is not rejected at 5% significance level. Because p-value is 0.246 (see Table 6), this hypothesis is not rejected also at 24% significance level. The difference between average number of points in test variants A8 and B4 in project “Entrance exams practice” in 2016 is not statistically significant.

Alfa=0.05	A8	B4
Average	47.931	40.517
Variance	588.42	572.04
Sample size	29	29
Standard deviation (see (4))	24.088	
Degrees of freedom	56	
t Stat	1.172	
p- value	0.246	
Critical value	2.003	

**Table 6: Results of the t-test for independent samples (source: own calculation)**

## Discussion

From results of this paper it follows that the difference between average number of points in mathematics in test variants A8 and B4 in entrance exams in 2015 is statistical significant – see also second row of Table 7. Therefore, we ask whether these test variants are equally difficult for students.

Test variant	A8	B4
2 different groups of students in entrance exams in 2015	$\bar{x}_{A8} = 64.17$	$\bar{x}_{B4} = 47.92$
1 group of students in preparatory course in 2016	$\bar{x}_{A8} = 55.29$	$\bar{x}_{B4} = 52.65$
2 different groups of students in project “Entrance exams practice” in 2016	$\bar{x}_{A8} = 47.93$	$\bar{x}_{B4} = 40.52$

**Table 7: Average number of points in mathematics (source: own calculation)**

For the same group of students in preparatory course in 2016 the difference between average number of points in mathematics in test variants A8 and B4 is not statistical significant – see also third row of Table 7. For two different groups of students in project “Entrance exams practice” in 2016 the difference between average number of points in mathematics in test variants A8 and B4 is not statistical significant, either. It means that the difference between test variants A8 and B4 in entrance exams in 2015 could be caused by other factors, e. g. by the different level of knowledge of students who wrote these variants in entrance exams in 2015.

Entrance exams in mathematics at the University of Defence in Brno with similar problems are analysed in Hořková-Majerová and Račková (2010) - examples in mathematics with the same level of difficulty. Analysis of the entrance examination in mathematics at University of Pardubice we can find in Linda and Kubanová (2013) – correlation between results of the entrance examination test in mathematics and examination in mathematics at the university. The aim of these papers was a little different. Analysis of the entrance tests in mathematics at Faculty of mathematics, physics and informatics at Comenius University in Bratislava we can find in Kohanová (2012). The focus of the paper is to find what types of tasks should be included in the entrance test if we want to select students who have best predispositions for study. Similar statistical methods here were used as in present paper.

The problem of the same difficulty of tests variants in entrance examination, which is mentioned in this paper, occurs in scientific papers only rarely. One of them is paper written by Klůfa (2015a). There is on the basis of test of independence in contingency table shown that results of entrance examinations at the Faculty of Informatics and Statistics at University of Economics in Prague do not depend on the test variants, i.e. the analogous result as in present paper.

## Conclusion

The differences between average number of points in mathematics in test variants A0, A8, A9, B0, B4, B6, which were used for the entrance examinations in mathematics at the Faculty of Business Administration in 2015, are statistically significant. 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. From results of this paper it follows that these significant differences between tests variants may arise due to different level of knowledge of

the students who wrote these variants. On the other hand, the difficulty of test variants for students is poorly measured. This problem will be solved in the following paper.

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.

## Acknowledgements

This paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

## 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.

Hošková-Majerová, Š, Račková, P. (2010) 'Problematic examples at the entrance exams of mathematics at the University of Defence', *Procedia - Social and Behavioral Sciences*, vol. 9, pp. 348–352. <http://dx.doi.org/10.1016/j.sbspro.2010.12.162>

Hrubý, M. (2013) 'Extended elements of the question objects', *Efficiency and Responsibility in Education, Proceedings of the 10th International Conference*, Prague, pp. 204–209.

Kaspříková, N. (2012) 'Data analysis of students' performance', *Efficiency and Responsibility in Education, Proceedings of the 9th International Conference*, Prague, pp. 213–218.

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. (2015a) 'Dependence of the Results of Entrance Examinations on Test Variants', *Procedia - Social and Behavioral Sciences*, vol. 174, pp. 3565–3571. <http://dx.doi.org/10.1016/j.sbspro.2015.01.1073>

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

Klůfa, J. (2015c) 'Comparison of the Ways of Acceptance Students at University', *Journal on Efficiency and Responsibility in Education and Science*, vol. 8, no. 3, pp. 72–76. <http://dx.doi.org/10.7160/eriesj.2015.080204>

Klůfa, J., Langhamrová, J. (2015) *Matematika - příprava k přijímacím zkouškám na VŠE*, Prague: Ekopress.

Klůfa, J. (2016) 'Analysis of the differences between results of test variants', *Efficiency and Responsibility in Education, Proceedings of the 13th International Conference*, Prague, pp. 279–285.

Kohanová, I. (2012) 'Analysis of university entrance test from mathematics', *Acta Didactica Universitatis Comenianae Mathematics*, vol. 12, pp.31-46.

Kučera, P., Svatošová, L., Pelikán, M. (2015) 'University study results as related to the admission exam results' *Efficiency and Responsibility in Education, Proceedings of the 12th International Conference*, Prague, pp. 318–324.

Kubánová, 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. <http://dx.doi.org/10.7160/eriesj.2012.050302>

Linda, B., Kubánová, J. (2013) 'Relation between results of the entrance examination test in mathematics and examination in mathematics at the university', *Efficiency and Responsibility in Education, Proceedings of the 10th International Conference*, Prague, pp. 370–373.

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

Mošna, F. (2013) 'E-learning and interactive dynamical figures for teaching of mathematics', *Efficiency and Responsibility in Education, Proceedings of the 10th International Conference*, Prague, pp. 449–454.

Otavová, M., Sýkorová, I. (2014) 'Analysis of Scores from Mid-Term and Final Test by a Contingency Table', *Efficiency and Responsibility in Education, Proceedings of the 11th International Conference*, Prague, pp. 527–533.

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. (2005) 'Algorithms for converting raw scores of multiple-choice question tests to conventional percentage marks', *International Journal of Engineering Education*, vol. 21,, no. 6, pp. 1189-1194.

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

