

NON-PARAMETRIC ANOVA METHODS APPLIED ON STUDENTS' PERFORMANCE DEVELOPMENT IN COURSE OF STATISTICS

Martina Zámková¹, Martin Prokop¹, Radek Stolín¹

¹ Department of Mathematics, College of Polytechnics Jihlava, Tolstého 16, 586 01 Jihlava, Czech Republic

Link to this article: <https://doi.org/10.11118/actaun202068010281>

Received: 29. 6. 2019, Accepted: 22. 11. 2019

To cite this article: ZÁMKOVÁ MARTINA, PROKOP MARTIN, STOLÍN RADEK. 2020. Non-Parametric Anova Methods Applied on Students' Performance Development in Course of Statistics. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 68(1): 281–289.

Abstract

This study intends to use non-parametric statistical methods for comparison of mean values in assessing the changes of overall grades in basic statistics at the College of Polytechnics in Jihlava. The data covering students' grades were transferred to the ordinal scale and evaluated with the use of the corresponding non-parametric statistical methods, due to failing the data normality test. In order to identify statistically significant grade differences over the years 2006–2018, we have used the test of equality of means as well as the relevant analyses of variance, i.e. the Mann-Whitney U test to compare two samples, the Kruskal-Wallis test to compare multiple samples, and the post hoc analysis. All tests that were carried out confirm that on average, women and Finance and Management students tend to achieve better grades in statistics, as opposed to men, and Travel and Tourism students. Mean grades have rather fluctuated over the years. In the beginning (2006–2009) and in the end (2014–2017) of the observed period the grades have been significantly superior to the grades in the in-between stage (2010–2013). The identified fluctuation in grades shall lead to implementation of measures aimed at achieving an overall increase in statistics' education quality and students' grades.

Keywords: College of Polytechnics Jihlava, statistics, success rate, Mann-Whitney U test, Kruskal-Wallis test, non-parametric tests, post hoc analysis

INTRODUCTION

The aim of this article is to explore the possibilities of non-parametric tests for the comparison of mean values to assess the changes in the students' performance in the introductory course to statistics at the College in Jihlava in the past twelve academic years. As a reaction to the results, corresponding measures (summer and winter schools, new interactive e-learning module, new learning material, a database of solved and unsolved exercises, auto-correct tests etc.) will be adopted.

The authors of this study have already performed a similar analysis at the College in Jihlava in the past, only with regard to Mathematics I. Their findings on

this subject have been published in separate papers by Zámková *et al.* (2016, 2017). The first paper by Zámková *et al.* (2016) focused on the student success rate timeline in Mathematics I, and the next year Zámková *et al.* (2017) proved that the implementation of follow-up measures helped to improve the success rate in the two subsequent years.

Others, such as Chechi and Bhalla (2017), discussed a similar topic – factors influencing the students' university performance in mathematics. Furthermore, Muollo *et al.* (2018) described the implementation of modern methods in education and conducted a students' survey to assess their impact on students' satisfaction. In his research, Yue

(2011) analysed the potential to improve students' math skills in technology and engineering programs at the Newark University. Similarly Čipková *et al.* (2019) described scientific literacy levels achieved by Slovak secondary school graduates.

Reyneke *et al.* (2018) explored the effects of educational innovations, such as on-line homework system in combination with a “flipped” classroom, on the performance of university statistics students. Innovative approaches in college statistics and their effects were examined also by Groeneveld (2014). Different team of researchers, Hildreth *et al.* (2018), compared the success rates of university students attending courses with different curricula. Another comparison was suggested by Chow (2015), who looked into the success rates of students doing a computer-based homework and students working without the computer-based application.

MATERIALS AND METHODS

Our data was retrieved from the College of Polytechnics information system. The research focused on the students' success rate development in Statistics over the past 12 academic years. For each student, we took into consideration the grades and final grade, the study programme, mode of attendance and the date of their completion of the course.

The Mann-Whitney *U* test is a non-parametric test used to compare two independent samples. This test can be used to determine whether two independent samples were selected from populations having the same distribution. Null hypothesis assumes that the mean ranks of the groups are the same. It can represent an alternative to the *t*-test for independent samples when the population cannot be assumed to be normally distributed. The Mann-Whitney *U* test may be used for ranked or ordinal data.

Test statistic with asymptotically normal distribution (large samples) is

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum_{i=n_1+1}^{n_1+n_2} R_i, \quad (1)$$

R_i is the rank of the sample size, n_1 , n_2 is the sample size. SPSS software was used for the data processing.

The Kruskal-Wallis test by ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of

equal or different sample sizes. It extends the Mann-Whitney *U* test when there are only two groups. And like before, null hypothesis assumes that the mean ranks of the groups are the same. It can represent an alternative to the parametric one-way analysis of variance (ANOVA) in case the population cannot be assumed to be normally distributed. For more details see Wonnacott (1990). The Kruskal-Wallis test may be used for ranked or ordinal data. A significant Kruskal-Wallis test indicates that at least one sample stochastically dominates one other sample. The test does not identify where this stochastic dominance occurs or for how many pairs of groups stochastic dominance obtains. For analysing the specific sample pairs for stochastic dominance post hoc tests are applicable.

Test statistic with asymptotically chi-square distribution with $c-1$ degrees of freedom is

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^c \frac{T_j^2}{n_j} \right] - 3(n+1), \quad (2)$$

n is the sum of sample sizes for all groups, c is the count of groups, T_j is the sum of ranks in j -th group and n_j is the size of the j -th group. SPSS software was used for the data processing. For more details see Hendl (2006) or Anděl (2005).

RESULTS

Students' grades were transferred to the ordinal scale of 1(A) to 6(F). Therefore – the lower the number, the better the grade. In order to analyse the mean grades, the students were divided into groups according to their gender, programme, mode of attendance, and academic year. Finally, the mean grade changes in individual years were analysed in more detail in relation to the programme of study.

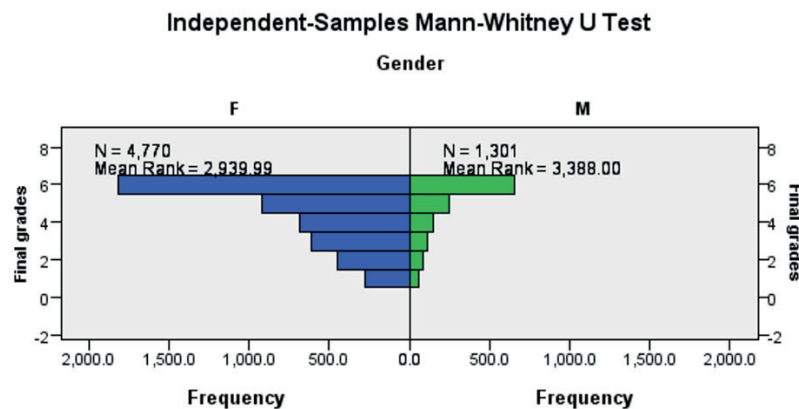
The descriptive statistics table (see Tab. I) illustrate that on average, women tend to achieve better results.

The frequency histogram's shape is (see Fig. 1) similar in both cases, however, the Mann-Whitney *U* Test *p*-value is less than 0.001 (see Tab. II), proves statistically significant difference between the mean results of men and women. Therefore, not only the descriptive statistics table and the figure of mean values indicate women's better performance, but the test too suggests that the mean women's grade is statistically significantly better than men's

I: Descriptive statistics table – grades by gender

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Female	4 770	4.46	1.591	0.023	4.42	4.51	1	6
Male	1 301	4.85	1.489	0.041	4.77	4.93	1	6
Total	6 071	4.54	1.578	0.020	4.50	4.58	1	6

Source: own calculation



1: Frequency histogram for the corresponding groups – grades by gender (F – Female, M – Male)

Source: own calculation

II: Test statistics table – Independent-Samples Mann-Whitney U Test – grades by gender

Total N	6 071.000
Mann-Whitney U	3 560 841.000
Wilcoxon W	4 407 792.000
Test Statistic	3 560 841.000
Standard Error	53 760.986
Standardized Test Statistic	8.518
Asymptotic Sig. (2-sided test)	< 0.001

Source: own calculation

mean grade. Plus, the shaping of histograms shows significantly asymmetrical distribution, different from normal distribution – a result that was repeated in other analyses – hence the decision to go with non-parametric tests. The second reason behind that decision is the discrete ordinal scale.

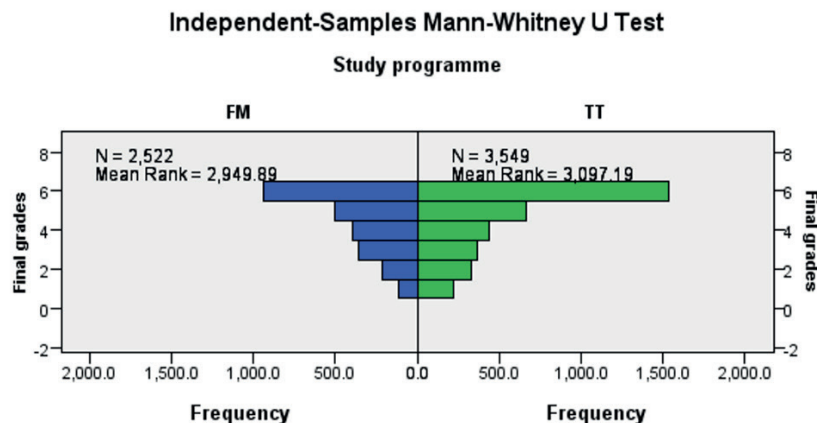
The descriptive statistics table (see Tab. III) illustrate that on average, Finance and Management students tend to achieve better results.

The frequency histogram (see Fig. 2) clearly shows higher frequency of F grades in the Travel and Tourism programme. The Mann-Whitney U Test p-value is less than 0.001 (see Tab. IV) proves

III: Descriptive statistics table – by programme (FM – Finance and Management, TT – Travel and Tourism)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
FM	2 522	4.49	1.528	0.030	4.44	4.55	1	6
TT	3 549	4.58	1.612	0.027	4.53	4.63	1	6
Total	6 071	4.54	1.578	0.020	4.50	4.58	1	6

Source: own calculation



2: Frequency histogram for the corresponding groups – by programme (FM – Finance and Management, TT – Travel and Tourism)

Source: own calculation

IV: Test statistics table – Independent-Samples Mann-Whitney U Test – by programme

Total N	6 071.000
Mann-Whitney U	4 692 450.000
Wilcoxon W	10 991 925.000
Test Statistic	4 692 450.000
Standard Error	64 564.669
Standardized Test Statistic	3.363
Asymptotic Sig. (2-sided test)	< 0.001

Source: own calculation

statistically significant difference between mean results of both groups. Therefore, not only the descriptive statistics table and the figure of mean values indicate better performance of Finance and Management students, but the test too suggests that the mean FM students' grade is statistically significantly better than TT students' mean grade.

The descriptive statistics table (see Tab. V) illustrate that on average, full-time students tend to achieve considerably better results.

The frequency histogram (see Fig. 3) clearly shows higher frequency of F grades in the part-time study mode. The Mann-Whitney U Test p -value is less than 0.001 (see Tab. VI) proves statistically significant difference between mean results of both groups. In other words, not only the descriptive statistics table and the figure of mean values

indicate full-time students' better performance, but the test too suggests that the mean FT students' grade is statistically significantly better than PT students' mean grade.

Looking at the descriptive statistics table (see Tab. VII) and the figure of mean values (see Fig. 4), we may say that the mean grades clearly fluctuated a lot in individual years. Let us say that all in all, there was a gradual significant deterioration between the years 2006 and 2010, then this rather unfavourable situation lasted until 2013, and since then we have experienced certain improvement.

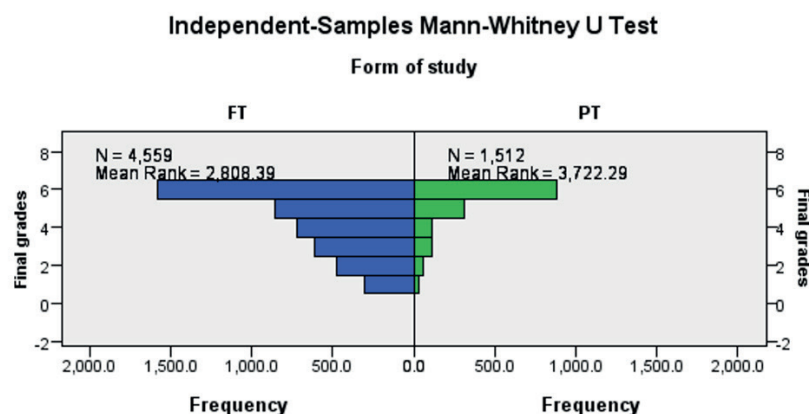
The boxplot (see Fig. 5) as well as the figure of mean values show that there has been a major fluctuation in the mean grade value over the years. The Kruskal-Wallis Test p -value is less than 0.001 (see Tab. VIII) is an indication that in some years the mean grade is statistically significantly different. Considering the test significance, we have decided to perform the post hoc analysis, in order to identify pairs of years with statistically different grades (see Fig. 6). The figure shows that in the first years (2006–2009) and in the last years (2014–2017) the grades have been significantly superior to the grades in the in-between stage (2010–2013).

Afterwards, the grade changes in individual years were further analysed with regard to the programme of study. Between the two programmes, the differences are rather minor (see the figure of mean values Fig. 7 and the boxplot Fig. 8); one may only say, that as for the FM programme, the

V: Descriptive statistics table – by mode (FT – full-time, PT – part-time)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
FT	4 559	4.34	1.618	0.024	4.29	4.39	1	6
PT	1 512	5.16	1.264	0.033	5.10	5.23	1	6
Total	6 071	4.54	1.578	0.020	4.50	4.58	1	6

Source: own calculation



Source: own calculation

VI: Test statistics table – Independent-Samples Mann-Whitney U Test – by mode

Total N	6 071.000
Mann-Whitney U	4 484 277.000
Wilcoxon W	5 628 105.000
Test Statistic	4 484 277.000
Standard Error	56 660.458
Standardized Test Statistic	18.314
Asymptotic Sig. (2-sided test)	< 0.001

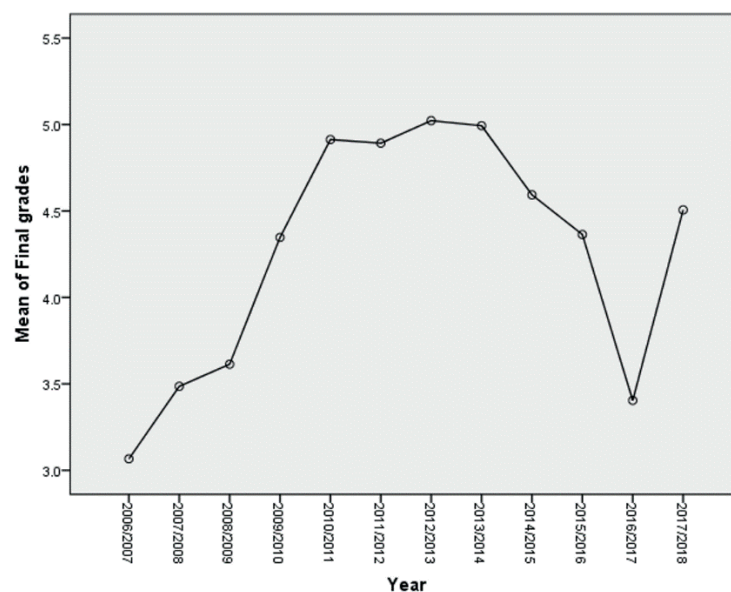
Source: own calculation

deterioration in grades came more gradually in the first years (2006 to 2010), while in the TT programme, the grades went downhill rather quickly between 2008 and 2009. In 2016, we have identified a considerable (although nonrecurring and only temporary) TT improvement. The differences between some of the years are yet again statistically significant, see the test statistics tables for the Kruskal-Wallis test, see Tab. IX.

VII: Descriptive statistics table – by years

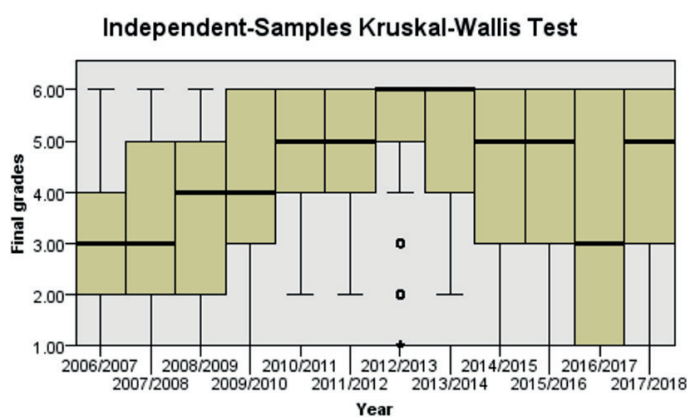
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
2006/2007	15	3.07	1.534	0.396	2.22	3.92	1	6
2007/2008	323	3.49	1.547	0.086	3.32	3.66	1	6
2008/2009	391	3.61	1.574	0.080	3.46	3.77	1	6
2009/2010	498	4.35	1.334	0.060	4.23	4.46	1	6
2010/2011	644	4.91	1.155	0.045	4.82	5.00	1	6
2011/2012	862	4.89	1.142	0.039	4.82	4.97	1	6
2012/2013	797	5.02	1.524	0.054	4.92	5.13	1	6
2013/2014	743	4.99	1.511	0.055	4.88	5.10	1	6
2014/2015	767	4.59	1.655	0.060	4.48	4.71	1	6
2015/2016	470	4.36	1.664	0.077	4.21	4.51	1	6
2016/2017	324	3.40	2.020	0.112	3.18	3.63	1	6
2017/2018	237	4.51	1.469	0.095	4.32	4.69	1	6
Total	6071	4.54	1.578	0.020	4.50	4.58	1	6

Source: own calculation



4: Mean values – by years

Source: own calculation



5: Boxplot: grades over the years

Source: own calculation

VIII: Test statistics table – Independent-Samples Kruskal-Wallis Test – grades over the years

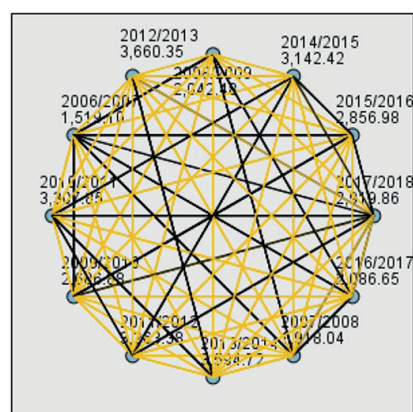
Total N	6 071.000
Test Statistic	650.259
Degrees of Freedom	11.000
Asymptotic Sig. (2-sided test)	< 0.001

Source: own calculation

DISCUSSION

This is not the first analysis of its kind. In the past, the collective of authors from the College of Polytechnics in Jihlava have focused on the changing grades in the introductory course to mathematics. This research by Zámková *et al.* (2016) confirmed that the success rate in 2006–2015 in Mathematics I at the College of Polytechnics in Jihlava had decreased in all of the study programmes in question. Technical programmes (Applied Computer Science and Computer Systems) have been proved to have lower success rate in mathematics than economic programmes (Finance and Management

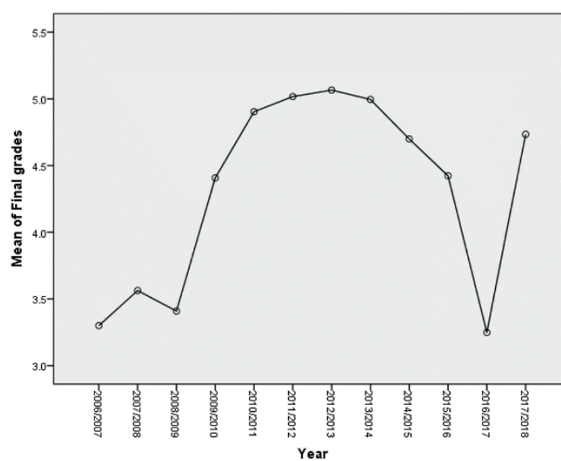
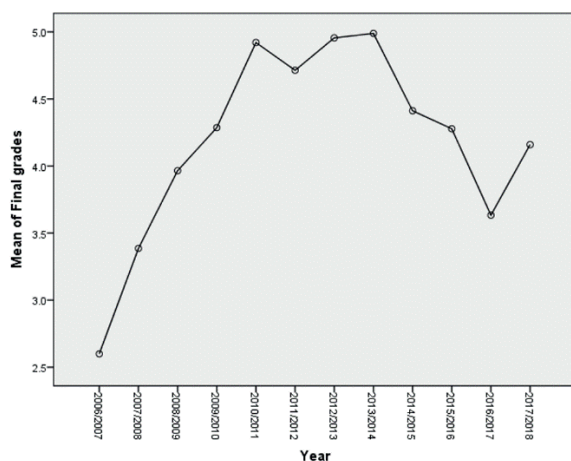
Pairwise Comparisons of Year



6: Pairwise Comparisons of Year

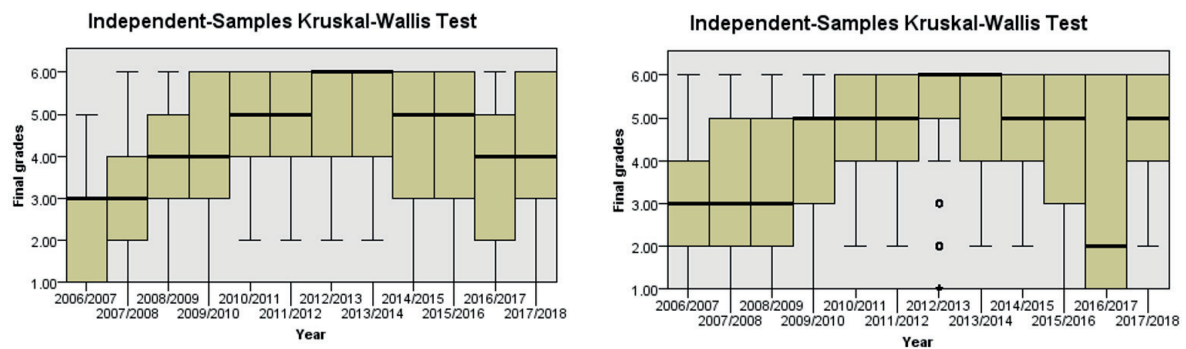
Source: own calculation

and Travel and Tourism). In an attempt to reverse this trend, the College has adopted many support measures. The following year, Zámková *et al.* (2017) published a paper mapping the improvement of students' success rate in mathematics at the College



7: Mean values – grades over the years – Finance and Management (left); Travel and Tourism (right)

Source: own calculation



8: Boxplot: grades over the years – Finance and Management (left); Travel and Tourism (right)
Source: own calculation

IX: Test statistics table – Independent-Samples Kruskal-Wallis Test – grades over the years – Finance and Management (left); Travel and Tourism (right)

Finance and Management		Travel and Tourism	
Total N	2,522	Total N	3,549
Test Statistic	264.877	Test Statistic	400.078
Degrees of Freedom	11.000	Degrees of Freedom	11.000
Asymptotic Sig. (2-sided test)	< 0.001	Asymptotic Sig. (2-sided test)	< 0.001

Source: own calculation

of Polytechnics in Jihlava achieved in the last two years thanks to the aforementioned measures. Some of these measures, *inter alia*, are – an additional seminar in mathematics, new textbook, a database of solved and unsolved exercises, new e-learning module, and the winter and summer schools of mathematics. Based on the non-parametric tests comparing the mean success rates, the authors have confirmed the change for the better.

In the same year, Chechi and Bhalla (2017) revealed that the success rate in mathematics was constantly poor, and that the majority of students thought of mathematics as boring and dull. Čipková *et al.* (2019) used tests to discover that the grammar school students' understanding of the basics of statistics was poor, thus indicating an insufficient level of mathematical literacy in the field. Overall, the students graduating from natural sciences had better scientific literacy, in comparison with the liberal arts students. The research by Chow (2015) looked into the matter from another perspective and proved that students who used a computer app to complete mathematical and statistical homework were just as successful as the students who did the homework traditionally. The major advantages of the computer-based application included immediate feedback and quick evaluation. Reyneke *et al.* (2018) approached this issue from a similar point of view: they compared statistics students' success rate in traditional education, and in education using an on-line homework system. They identified statistically significant success rate improvement in the latter group. Using mixed logistic regression Hildreth *et al.* (2018) calculated that the statistics students of

a simulation-based curriculum achieved a better understanding of the issue than the students of a traditional curriculum. Additionally, Groeneveld (2014) identified statistics students' success rate improvement after introducing a special tool from a gaming environment. Yue (2011) managed to improve students' mathematical skills through intensive mentoring and tutoring during a summer program.

This analysis of students' grades in statistics shows that the students of the Finance and Management programme have statistically significant better mean grade than the Travel and Tourism students. Since the students of both programmes have access to similar learning materials and the curricula are basically the same, it is recommended to focus more closely on the TT programme and to make every effort to provide students with the best education possible. In order to successfully master statistics, one needs certain mathematical knowledge and skills; it is therefore possible that the reason behind worse grades may be slightly less complex mathematical curriculum of this programme, or overall poorer mathematical skills of the TT students.

Furthermore, the authors came to the conclusion that full-time students had statistically significant better mean grades than part-time students. Poor performance of part-time students may be caused by the insufficient time allocation, as more time may be needed to master the subject. There is apparently a need to introduce new clear and simple learning materials containing large amounts of examples with solutions that could offset the insufficient time

and space allocation when it comes to exercises in the part-time study mode.

Considering the high fluctuation of grades over the years, it would be only fitting to implement support measures in order to stabilize the students'

performance in statistics in the upcoming years. Planned measures include winter and summer schools of statistics, new interactive e-learning module, more study materials, a database of solved and unsolved exercises, and auto-correct tests.

CONCLUSION

This research attempted to contemplate the uses of statistical tests in comparing the mean values of several samples and apply them on an example in order to determine whether and how the students' grades in the basic course of statistics at the College of Polytechnics in Jihlava changed in the last twelve years. Students' grades A-F were transferred to the ordinal scale of 1–6. In order to analyse the mean grades, the students were divided into groups according to their gender, programme, mode of attendance, and academic year. The mean grade changes in individual years were analysed in more detail in relation to the programme of study. The shaping of the students' grades frequency histogram showed that the frequencies did not indicate normal distribution. Hence the decision to employ non-parametric statistical methods for the assessment of mean grade changes over the years. The second reason behind this decision is the fact, that only ordinal data were available.

Data analysis confirmed that women tend to achieve better mean grades in statistics than men. The tests also confirmed that Finance and Management students tend to achieve better mean grades in statistics, as opposed to the Travel and Tourism students. Furthermore, the authors came to the conclusion that full-time students had statistically significant better mean grades than part-time students.

Detailed analysis suggested that the mean grades had rather fluctuated over the years. There was a gradual, yet significant deterioration between the years 2006 and 2010, which lasted until 2013, and since then we have experienced certain improvement. The post hoc analysis showed that in the beginning (2006–2009) and in the end (2014–2017) of the observed period the grades had been significantly superior to the grades in the in-between stage (2010–2013). Further analyses of the grade changes in individual years in relation to the programme of study have not established any significant difference between the two.

REFERENCES

- ANDĚL, J. 2005. *Mathematical statistics* [in Czech: *Základy matematické statistiky*]. Praha: Matfyzpress.
- ČIPKOVÁ, E., KAROLČÍK, Š. and SCHOLZOVÁ, L. 2019. Are secondary school graduates prepared for the studies of natural sciences?—evaluation and analysis of the result of scientific literacy levels achieved by secondary school graduates. *Research in Science and Technological Education*, In Press.
- GROENEVELD, C. M., 2014. Implementation of an Adaptive Training and Tracking Game in Statistics Teaching. In: KALZ, M. and RAS, E. (Eds.). *Computer Assisted Assessment. Research into E-Assessment*. CAA 2014. Communications in Computer and Information Science, vol 439. Cham: Springer, pp. 53–58.
- HENDL, J. 2006. *Summary of statistical methods: analysis and meta-data* [in Czech: *Přehled statistických metod: analýza a metaanalýza dat*]. Praha: Portal.
- HILDRETH, L. A., ROBISON-COX, J. and SCHMIDT, J. 2018. Comparing student success and understanding in introductory statistics under consensus and simulation-based. *Statistics Education Research Journal*, 17(1): 103–120.
- CHECHI, V. K. and BHALLA, J. 2017. Invigorating self-regulated learning strategies of mathematics among higher education students. In: *Proceedings of the International Conference on Recent Advances in Fundamental and Applied Sciences – RAFAS 2016*. 25–26 November, Punjab, India. AIP Conference Proceedings volume 1860. American Institute of Physics Inc., Article number 020036.
- CHOW, A. F. 2015. Online homework impact in undergraduate mathematics and business statistics courses. *Educational Studies*, 41(3): 244–248.
- MUOLLO, K., BASAVARAJ, P. and GARIBAY, I. 2018. Understanding students' online reviews to improve college experience and graduation rates of stem programs at the largest post-secondary institution: A learner-centered study. In: *Proceedings of the International Frontiers in Education Conference, FIE 2018*. 3–6 October, San Jose, United States. Institute of Electrical and Electronics Engineers Inc., pp. 1–7.

- REYNEKE, F. E., FLETCHER, L. and HARDING, A. 2018. The Effect of Technology-based Interventions on the Performance of First Year University Statistics Students. *African Journal of Research in Mathematics, Science and Technology Education*, 22(2): 231–242.
- YUE, J. 2011. Improving math skills through intensive mentoring and tutoring. In *Proceedings of the ASEE Annual Conference and Exposition*. 26–29 June, Vancouver, British Columbia, Canada. American Society for Engineering Education, pp. 1–12.
- WONNACOTT, T. H. and WONNACOTT, R. J. 1991. *Introductory statistics for business and economics*. USA: John Wiley & Sons.
- ZÁMKOVÁ, M., PROKOP, M. and STOLÍN, R. 2016. Development of the success rate in mathematics 1 at the college of polytechnics Jihlava (Czech Republic) in 2006–2015. *Turkish Online Journal of Educational Technology*, (Special issue 2016): 700–708.
- ZÁMKOVÁ, M., PROKOP, M. and STOLÍN, R. 2017. Increasing the success rate in mathematics at the college of polytechnics Jihlava (Czech Republic) as a result of the implementation of support measures. *Turkish Online Journal of Educational Technology*. (Special issue 2017): 142–150.

Contact information

Martina Zámková: martina.zamkova@vspj.cz
Martin Prokop: martin.prokop@vspj.cz
Radek Stolín: radek.stolin@vspj.cz