Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

PREFACE

Peter Kovacs, IPC chair

Citizens need sophisticated ways of thinking in order to understand complex real social and economic phenomena and to interpret relationships among social and business data correctly. Huge amounts of data, data sources, and visualization tools provide an opportunity to illustrate complex relations with real data. At the same time, the misuse of these tools can lead to misinterpretations, and perhaps to poor decision making.

The conference entitled Challenges and Innovations in Statistics Education, which was a multiplier event of ProCivicStat project, was organized on 7–9 September, 2018. ProCivicStat (http://www.procivicstat.org), a strategic partnership of six universities funded by the Erasmus+ program of the European Union, is developing new methods in statistics instruction for high-schools and universities, this way contributing to young people's ability to understand quantitative evidence about the key social phenomena that permeate civic life. These materials use authentic large scale data in topics such as migration, quality of life, sustainable development goals, and social inequality, often presented in innovative ways.

The conference was sponsored by the International Association for Statistical Education (IASE) and the Hungarian Statistical Association (MST).

For the participants, the goal of the conference was to learn about ProCivicStat materials and other relevant resources, to share their ideas and current practices, and to work on challenges and innovations in statistics education. Use of real data, databases, IT solutions, visualizations, innovative teaching, learning and examining methods, solutions were the main topics of the conference in English and Hungarian languages.

The plenary talk was given by Jim Ridgway on *Statistics for Empowerment: opportunities and challenges.* In course of the event, one special session on *Resources to Support Innovative Teaching: conceptual maps, data sources, and visualization tools,* 8 workshops on *Codap, R, digital tools, Jupyterhub, Answerminer, Understanding statistics about society and gamification*, and 7 presentation sessions *on ProCivicStat approach, issues and innovative teaching methods, misuse and IT support of Statistics and Developing (Official) Statistical Literacy* with 22 talks were organized, 72 colleagues from 13 countries (43 Hungarians and 29 from other countries, mainly from Eastern Europe) participated. The materials of the sessions are available at http://www.eco.u-szeged.hu/procivicstat.

This proceeding contains 13 papers in the above mentioned topics.

- Jim Ridgway, Durham University, UK
- Pedro Campos, Statistics Portugal & University of Porto, Portugal
- Rolf Biehler, University of Paderborn, Germany

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

DEVELOPING OFFICIAL STATISTICS LITERACY: A PROPOSED MODEL AND IMPLICATIONS

Iddo Gal[1] and Irena Ograjenšek[2]

[1] University of Haifa, Department of Human Services, Israel
[2] University of Ljubljana, Faculty of Economics, Slovenia
iddo@research.haifa.ac.il
irena.ograjensek@ef.uni-lj.si

This paper points to a gap in the literature regarding the lack of a widely accepted conceptual framework about the knowledge bases that adults at large and non-specialists need in order to be critically aware of social and economic phenomena by understanding key aspects of official statistics. This negatively affects work by official statistics providers and statistics educators. We thus propose a model encompassing six key knowledge elements needed for official statistics literacy and highlight its implications. We emphasize the need to develop a digital textbook about official statistics literacy as well as a modular online course, and point to other directions official statistics providers and statistics educators could take.

Keywords: statistical literacy; competencies; official statistics literacy; dissemination; adult education.

INTRODUCTION

In recent years, both national and international statistical offices as well as other producers of official statistics have been paying increasing attention to the formal training of professional statisticians who work in national and international statistical systems, and sometimes to the training of other user groups (e.g., MacCuirc 2015). However, the provision of training or resources related to official statistics for wider, non-professional audiences and adults at large has been largely left aside.

Further, there is a surprising lack of solid educational materials in official statistics designed for professionals, i.e., statistics or economics majors (Pfeffermann, 2015). A literature search we conducted did not identify a single current textbook that describes key knowledge bases which have to be emphasized in detail when educating statistics majors about official statistics aside from Citro and Straf (2013). This US-based text focuses both on key aspirations or expectations from an official statistics provider (for example, relevance to policy issues, credibility among data users, trust among data providers, independence from political and external influences), and on numerous important administrative and organizational practices and roles (such as mission clarity, confidentiality, continuous development of useful data, openness about sources, data limitations transparency, and more). These are core issues for all official statistics providers around the world, yet they are not related to the comprehension of the actual products from the content point of view.

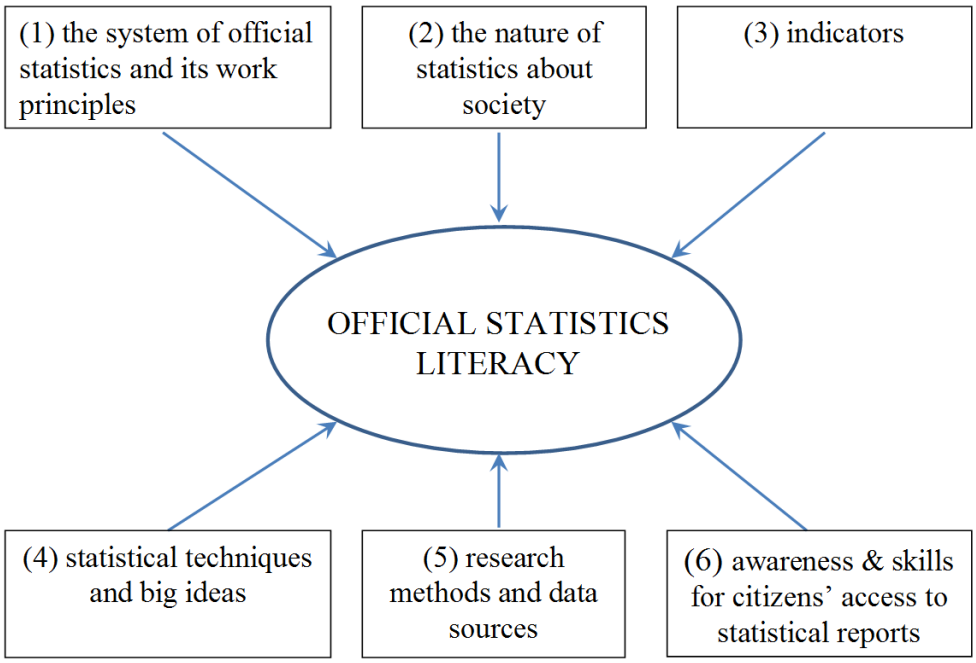
Hence, we focus our contribution on specific issues that official statistics providers may face if they want to help non-specialist users or citizens develop the aspects of statistical literacy (Gal 2002) that pertain to knowledge of, and engagement with, official statistics. For brevity, we refer to this desired knowledge base as *official statistics literacy* or *OSL*. We outline a model comprised of six possible building blocks of the desired knowledge base that is specific to official statistics literacy. Due to space limits we only elaborate on two of its most critical elements in this paper. A detailed presentation of the model and its elements is available in Gal & Ograjenšek (2017). We conclude the paper by discussing selected implications of our proposed model for official statistics providers and statistics educators.

PROPOSED MODEL

In Gal & Ograjenšek (2017), we argue that the six elements in the model outlined in Figure 1 are unique or specific knowledge bases and skills that citizens at large and non-specialists, as well as specialists in statistics, need in order to make sense of official statistics, in addition to having the knowledge bases and skills subsumed under the more generalized constructs reviewed in the previous

section. A specific point of comparison pertains to the knowledge expected of students who have taken an introductory statistics course at the undergraduate level, which may be the last, and for some students the only, structured exposure to statistics (Moore 1998; Meng, 2009).

Fig. 1. *Proposed model of six building blocks (areas) of official statistics literacy*



Given space limitations we elaborate here briefly only on two elements in the model that we deem key for adults at large and non-specialists, as well as for specialists, when they engage in their everyday or work lives with statistical information from official sources. Such information may be encountered, e.g., in the digital media, when listening to news on TV, reading a magazine article, or browsing the Internet and seeking information on topics of interest.

Regarding *the nature of statistics about society* (2nd element in the model), we point to recent work by the ProCivicStat project, a collaboration by six universities in five countries (Germany, Hungary, Israel, Portugal, and the United Kingdom) funded by the European Commission's ERASMUS+ program (see http://community.dur.ac.uk/procivic.stat), which aims to promote civic engagement and understanding among young adults regarding 'civic statistics' about key societal phenomena. As part of the ProCivicStat work, Engel, Gal, and Ridgway (2016) claim that to be fully engaged, citizens need to understand 'civic statistics' with regard to past trends, present situations, and possible future changes in diverse areas of importance to society such as demographics, employment, wages, migration, health, poverty, access to services, education, human rights, and other domains**.** The ProCivicStat analysis points to five general characteristics of civic statistics: They relate to *multivariate phenomena* and often to *aggregated data*. They also involve *dynamic data* that change or is updated over time. Furthermore, since data and findings about social phenomena are multivariate, aggregated at multiple levels, and dynamic, their description across time or comparison units requires the use of diverse types of representations and may often be delivered through *rich texts* and rich *visualizations* that are broader and at times more sophisticated compared with the limited range of ideas and representations included in introductory statistics classes.

Regarding *indicators* (3rd element in the model), the literature (e.g., Haack, 1979) suggests that official statistics providers create key messages to decision makers and to the general public regarding levels or changes in dozens of indicators, such as unemployment level, child mortality, gross domestic product, or income inequality (e.g., Gini coefficient). These and many other indicators in use by official statistics providers are often not raw variables, such as those encountered in introductory

statistics, but rather combinations of data elements that may be expressed as percentages, ratios, or numbers on arbitrary scales. They may be computed as simple rates, or be derived as complex aggregates of weighted elements. They may be based either on objective (e.g., consumer spending) or subjective data (e.g., consumer confidence), and their definitions may develop and change over time to reflect society's needs for information.

Whatever their definition, indicators are widely used by official statistics providers to report on a wide range of issues, hence their understanding is critical for all citizens. Although they are seemingly included in the broad description of the prior element "the nature of statistics about society", we highlight indicators as a separate aspect of official statistics, because while they are prevalent in public and political discourse, indicators are hardly ever described or analyzed in textbooks and statistics curricula for non-specialists, or in resources related to teaching research methods (Gal, 2007). This lack of attention to indicators may in part be a result of the fact that some indicators are comprised of *qualitative* variables or derived in part via qualitative methods, or otherwise their meaning and interpretation may require qualitative thinking. Possibly, such qualitative issues are not receiving the same attention from statistics educators, compared to quantitative variables and quantitative issues (see Ograjenšek and Gal, 2016).

IMPLICATIONS AND ACTION PLAN

To date, discussions of the connections between official statistics providers and statistics educators have focused in large part on how official statistics providers can facilitate improvement of generic statistics education at the school or university level. Within this framework, official statistics providers have been contributing to teachers' professional development by offering datasets, lesson plans, ideas for projects and poster competitions, and other resources that can inform class activities or highlight the importance of official statistics. Additional directions are noted by Gal (2002) and Sanchez (2008). Further, de Smedt (2016) describes directions related to enhancing support and explanations for users regarding interpretation of displays and published statistics, and other options.

Going beyond the directions for action noted above, in Gal & Ograjensek (2017) we discuss various implications of the proposed model with its six elements. We conclude from our analysis of the literature that supports the proposed model that unique efforts are needed to promote official statistics literacy. This is because regular statistics education, normally does not highlight the unique aspects of data about society, the complex nature of indicators, and other elements in our model.

Hence, it is important to continue existing collaborations between official statistics providers and school-level educators as noted by sources discussing the development of statistical literacy at school level as illustrated above. However, we believe the vision of systematically promoting official statistics literacy within the general adult population (including actions in countries with characteristics that differ from the few that have spearheaded educational services and activities in statistics education) requires an examination of additional directions – from a long-range future collaborative perspective.

With the above in mind, we outline two possible initiatives and some additional ideas that could be implemented in the international collaborative setting.

Firstly, we propose the development of a textbook on official statistics geared towards statistics majors as well as non-majors who may study selected topics in statistics. We note that there are many more non-majors than majors who take only introductory statistics, and the provision of an accessible textbook may be the first step to helping educational institutions develop new modules or whole courses related to official statistics that are currently lacking.

Secondly, we propose the development of an MOOC or a collection of digital (video and audio) teaching modules for entry-level majors, non-majors, and other groups of interest among the general public. It is hard to expect a single official statistics provider to shoulder responsibility and allocate resources related to both initiatives outlined above, although it would be technically possible. Both initiatives thus call for an international collaborative effort of official statistics providers, statistics educators, specialists in applied fields that rely on official statistics when discussing major concepts inherent to their disciplines, and other stakeholders. Such an effort can, of course, benefit from existing materials and frameworks developed in the context of existing diploma and degree programs listed in the previous sections of this article. Textbook developers participating in this collaborative

effort could build on experiences gained within the framework of the already mentioned Phare project, which resulted in the modular online Course on the European Economic Statistics (Bregar et al. 2000).

In addition, large professional associations with an international outreach and long-standing interest and activities in statistics education can also facilitate collaborations and the long-term development of a textbook and a MOOC. Key actors may be the International Statistical Institute (ISI) and its relevant divisions (the International Association for Statistics Education - IASE and the International Association for Official Statistics - IAOS) as well as the Royal Statistical Society (RSS), the American Statistical Association (ASA), and others.

Finally, it is essential to accompany the introduction of such new educational tools with appropriate research that will examine users' perceptions, value judgments, and reactions, as well as the many ways in which users operate within new digital learning spaces. As Gal and Ograjenšek (2010) argue, research that aims to support learners and users of statistics has to combine both quantitative *and* qualitative research methods, in order to provide effective information to the agencies or actors that develop statistical information products.

Acknowledgment: The first author was partially supported by a grant from the ERASMUS+ program of the European Commission for the *ProCivicStat* project. However, the opinions expressed in this article reflect the author's own views and not necessarily those of the sponsoring agency.

REFERENCES

Citro, C. F & Straf., M. E. (2005). *Principles and Practices for a Federal Statistical Agency. National Research Council. 3rd ed*. Washington: National Academy Press.

de Smedt, M. (2016). European Statistics and Eurostat's Contribution to Improving Statistical Literacy. In Engel, J. (Ed.), Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE. http://iase-web.org/Conference_Proceedings.php.

Engel, J., I. Gal, & J. Ridgway. (2016). *Mathematical Literacy and Citizen Engagement: The Role of Civic Statistics.* Paper presented at the 13th International Congress on Mathematics Education (ICME13).

Gal, I. & Ograjenšek, I. (2017). Official statistics and statistics education: Bridging the gap. *Journal of Official Statistics*, *33*(1), 79–100.

Gal, I. (2002). Adult Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review, 70*(1) 1-25.

Gal, I. (2007). Research Methods: Reflections on Teaching Frameworks and Research. In Murtonen, M., Rautopuro, J. & Väisänen, P (Eds), *Learning & teaching of Research methods at University*. Turku: Finnish Educational Research Association.

Gal, I. and I. Ograjenšek. (2010). Qualitative Research in the Service of Understanding Learners and Users of Statistics. *International Statistical Review, 78*(2) 287-298.

Haack, D. G. (1979). *Statistical Literacy: A Guide to Interpretation.* Wadsworth: Duxbury Press.

MacCuirc, E. (2015). You Don't Teach, Students Learn: Lessons Learned in Statistical Literacy and Statistical Education in Ireland. *Austrian Journal of Statistics, 44*(2) 73-83.

Meng, X. (2009). Desired and Feared—What Do We Do Now and Over the Next 50 Years? *The American Statistician, 63* (3) 202–210.

Moore, D. S. (1998). Statistics Among the Liberal Arts. *Journal of the American Statistical Association, 93*(444) 1253–1259.

Ograjenšek, I. & I. Gal (2016). Enhancing Statistics Education by Including Qualitative Research. *International Statistical Review, 84*(2) 165–178.

Pfeffermann, D. (2015). Methodological Issues and Challenges in the Production of Official Statistics. (24th Annual Morris Hansen Lecture). *Journal of Survey Statistics and Methodology, 3*(4) 425-483.

Sanchez, J. (2008). *Government Statistical Offices and Statistical Literacy*. International Statistical Literacy Project and ISI. http://iase-web.org/islp/Publications.php?p=Books.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

THE IMPORTANCE OF STATISTICAL LITERACY FOR DEMOCRACY – CIVIC–EDUCATION THROUGH STATISTICS

Achim Schiller and Joachim Engel
Ludwigsburg, University of Education
Schiller01@ph-ludwigsburg.de Engel@ph-ludwigsburg.de

Civic Education has had the same objective ("Mündigkeit") for the last 50 years, but the conditions to achieve it have changed. Mündigkeit is a prerequisite for citizen's participation, to strengthen and stabilize democratic structures. In the information age, Mündigkeit regarding statistics means having an orientation in the confusion of the modern information jungle and the deluge of quantitative information and statistics. The requirements for understanding and evaluating information about societal developments have changed: Statistical skills are becoming increasingly important for an evidence-based judgment in today's society. They entail understanding data-related arguments and representations, questioning possible conclusions as well as uncovering opinions and already made decisions.

NEEDS OF DEMOCRACY

Democracy is not only a form of government but also a political system. It has an impact on the society, the culture, and everyday life. Himmelmann (2001), a German political scientist, distinguishes between the individual level, the societal level and the institutional level. Democracy impacts all three of these levels and affects them sustainably. Negt (2011) argues that a consistent democratization of all the three levels is the only way to keep the democratic system as a whole alive. To be able to function, democracy needs *Mündigkeit* and the people in the respective levels need to be *mündig*. If a person is, for example, a patient (individual level), *Mündigkeit* could mean he should not make his way of life dependent on the doctor (Böhme, 2010). *Mündigkeit* in the case of a consumer (societal level) could mean he should know what goods and services he needs for his life (Buchholz, 2010). *Mündigkeit* for a citizen (institutional level) could mean to know how to participate within the system (see Negt, 2011).

The German word *Mündigkeit* has various meanings. It is attributed to Immanuel Kant (see Böhme, 2010), "to use one's own mind without the guidance of another person." It refers to a cluster of ideas around the goals of the enlightenment movement: empowerment to represent one's own interests; emancipation in the sense of detachment from the authorities; self-determination to decide freely about one's own life; taking responsibility about one's own life; maturity of being able to reflect the effect of one's own acting; autonomy in the sense of recognition of opinion-making and already made decisions as well as making independent informed decisions. All these concepts are possible characteristics of *Mündigkeit*.

Mündigkeit is a prerequisite for participation in the democratic system (Dammer & Wortmann, 2014). The people should know about what they want to achieve and how to realize it. For further development of the democratic system, it is necessary to reflect what could be done better and where change is needed. Participation in democratic processes contributes to the stabilization of the democratic system (Dammer & Wortmann, 2014). *Mündigkeit* is also important for the development and maintenance of the democratic culture (Frech & Richter, 2013). In relation to societal life, a reflection of the personal needs and interests is important for shaping society and everyday life. Dammer & Wortmann (2014) emphasize that democracy can also function or exist without *Mündigkeit*, but only if citizens understand themselves as an audience and are satisfied with the production of its appearance.

REQUIREMENTS FROM THE PEOPLE

Demanding the concept of *Mündigkeit* as a characteristic among citizens is not new. Immanuel Kant already called *Mündigkeit* an important characteristic for people in 1794 (see Böhme, 2010; Villhauer, 2010). Theodor Adorno formulated the evolvement towards *Mündigkeit* as a basis of democracy at the end of the 1950s (see Adorno, 2015). Also in modern times political scientists in Germany have described *Mündigkeit* as a bundle of qualities to shape their live (see Dammer & Wortmann, 2014), as a prerequisite for participatory acting (see Breit & Massing,

2013) or the ability to distinguish (see Bünger, 2013). In recent years, the critical handling of data has become an important part of *Mündigkeit*. Gramm (2010), for example, specifies *Mündigkeit* as "having an orientation in the confusion of the modern information jungle". According to Fischer (2012), *Mündigkeit* "enables the people to obtain the necessary information from the media and make sense of data." (p.11).

Requirements on the different levels of democracy

The change described above is also reflected in the requirements on the respective levels of democracy. In each of the levels, the people have to deal with the data deluge, with the associated difficulties and possibilities. People in their private and public life are constantly challenged to make decisions that go beyond their knowledge and competences (see Lengnink, Meyerhöfer & Vohns, 2013). For this reason "it is important to learn to get the necessary information, to ask questions, to understand" (Lengnink, Meyerhöfer & Vohns, 2013, p.4) to be able to make decisions based on *Mündigkeit*.

As the three levels have some features in common, it is also important to understand, verify, interpret, and critically evaluate data-related arguments and various representations (see Engel, 2014; Ridgway, 2015) at each of the three levels.

Requirements at the individual level are, for example, as a patient (see Böhme, 2010):

- Handle risk as an essential ingredient of life.
- Understand and evaluate medical test-results.
- Make decisions in the therapeutic process.

Requirements at the societal level are, for example, as a consumer (see Buchholz, 2010):

- Understand and evaluate advertising based on various study results.
- Evaluate recommendations from different institutions.
- Make adequate evaluation of one's own needs

Requirements at the institutional level are for a citizen:

- Understand concepts that describe social developments (see Engel, 2014).
- Take a critical attitude towards data-based /statistical statements (see Schiller, 2016).
- Evaluate and classify opinion surveys, political forecasts and status reports

Needed knowledge about information

Another important issue is to differentiate between information and opinion (see Schiller & Engel, 2016), not only since 'post-factual' has been the word of the year 2016. Data-based statements are increasingly used. Figure 1 shows the stages how data-based information is produced,

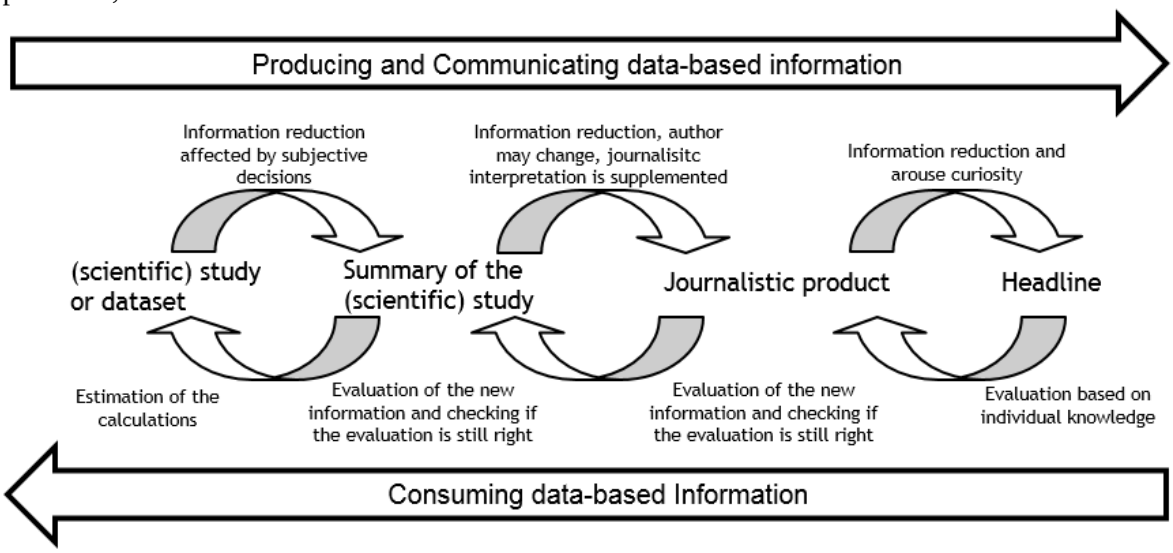


Figure 1

communicated and consumed. The terms refer to Gal (2002) and Rumsey (2002), which distinguish between data-producers, data-communicators and data-consumers.

The upper part shows the producing and communicating of data based information, from the (scientific) study over a summary of the (scientific) study to a journalistic product with the headline as an eye-catcher. There can be more steps if the first journalistic product is the basis for further ones. In this context, people should be aware of the fact that at every step of produced and communicated information there is an information reduction that is influenced by subjective decisions. A change of the author strengthens the subjective character of the transmitted information since further decisions were made. This makes it difficult to keep opinions and information apart at first sight which should, however, be considered while consuming data based information. The interpretation and evaluation of the headline is the first step (see Figure 1 lower part from right to the left). The headline is judged based on the own knowledge. After the decision that it is worth searching for further information, the additional information of the journalistic product is evaluated and judged again by means of one's own knowledge. In times of data-driven journalism, links to further information and references are found in more and more journalistic products. Sometimes the reader can even access the original datasets and check calculations or perform own analyses. In order to be able to carry out each of these individual steps, a wider knowledge of statistics is required (more than is normally taught at schools), in addition to critical evaluation and reflection as well as positive dispositions (see Ridgway, Nicholson & Gal, 2017).

Needed knowledge of statistics

To understand data-based arguments or to uncover opinions and already made decisions with regard to societal phenomena in the media, citizens who are *mündig* need knowledge of statistics that goes beyond traditional statistics courses.

- Statistics and risk
  In addition to the classical themes of statistics and risk like variability, describing and comparing distributions, correlation or regression and the concepts of probability and conditional probability (including Bayes theorem), the citizen also needs to know about statistical topics such as samples, representativeness, signal and noise, Bayesian inference or effect size. Understanding ideas around Big Data such as variety of data sources or techniques of analysis (see Ridgway, Nicholson & Gal, 2017) is also necessary.
- Modelling and representation
  Complex societal phenomena are not easy to measure. In addition, there is more than just one way to model a particular phenomenon: The decision for the model mostly depends on the background of the researchers. Representations can help in understanding social phenomena. *Mündigkeit* requires a familiarity with static and dynamic visualizations (see Ridgway, Nicholson & Gal, 2017). The way a latent variable is defined influences the measurement and, consequently, the result. The "scientific" definition is not always consistent with the everyday definition, which can lead to misunderstandings.
- Methodology of data collection
  Different issues require different approaches. A critical analysis of the strengths and weaknesses of a research method should be conducted (see Ridgway, Nicholson & Gal, 2017). In some cases, however, ethical reasons also determine whether an experiment, an observation study or a survey is appropriate as a research method.
- Contextual knowledge
  Contextual knowledge is of essential importance for the interpretation of social phenomena. The more background knowledge is available in the corresponding reference framework, the better can the information about the phenomena be evaluated regarding its plausibility. Likewise, prior knowledge helps to continue the examination with more detailed information.

Under *Civic Statistics*, various facets are subsumed and divided into three groups (engagement and action, knowledge and enabling processes), which are of importance in the daily handling with statistics within the three democratic levels (see Engel, Gal & Ridgway, 2016;

Ridgway, Nicholson & Gal, 2017). Ridgway, Nicholson & Gal (2017) described all of the above-mentioned elements as part of the group *knowledge* of Civic Statistics.

CONCLUSION

Democracy is inconceivable without *Mündigkeit*. Consequently, *Mündigkeit* has always been an important educational goal within democracy. In recent years, however, due to the change into an information society, the requirements for *Mündigkeit* have changed. Knowledge and skills from the field of Civic Statistics are becoming increasingly important for *Mündigkeit*. Above all, as far as the information procurement within the media is concerned, a large number of the arguments in the social field are data-based. That makes *Mündigkeit* inconceivable without Civic Statistics. If we follow the analogy, democracy is inconceivable without Civic Statistics.

There is obviously a need for training Civic Statistics as a component of civic education. However, Civic Statistics are hardly provided, neither in the school curriculum nor in teacher training. Due to its importance to democracy, we claim that Civic Statistics should have its own place in the school curriculum and not just be handled as part of teaching mathematics, social sciences, politics or geography.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Böhme, G. (edit.). (2010). *Der mündige Mensch. Denkmodelle der Philosophie, Geschichte, Medizin und Rechtswissenschaft*. Darmstadt: WBG

Breit, G. und Massing, P. (edit.) (2002). *Bürgergesellschaft – Zivilgesellschaft – Dritter Sektor.* Schwalbach: Wochenschauverlag

Buchholz, K. (2010). Der mündige Konsument. In G. Böhme (Hrsg.), *Der mündige Mensch. Denkmodelle der Philosophie, Geschichte, Medizin und Rechtswissenschaft*. 131 - 141. Darmstadt: WBG

Dammer, K.-H. & Wortmann, E. (2014). *Mündigkeit. Didaktische, bildungstheoretische und politische Überlegungen zu einem schwierigen Begriff* (Didactica nova, Bd. 23). Baltmannsweiler: Schneider-Verlag Hohengehren.

Engel, J. (2014): Open data, civil society and monitoring progress: challenges for statistics education. *In Makar, K., de Sousa, B. & Gould, R. (Eds.). Sustainability in statistics education: Proceedings of ICOTS9.* Voorburg, The Netherlands: Intern. Statistical Institute. http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_4F4_ENGEL.pdf

Engel, J.; Gal, Iddo; Ridgway, J. (2016), Mathematical Literacy and Citizen Engagement: the role of Civic statistics. Proceedings 13th International Congress on Mathematical Education, 13th International Congress on Mathematical Education, Hamburg.

Fischer, R. (2012). Fächerorientierte Allgemeinbildung: Entscheidungskompetenz und Kommunikationsfähigkeit mit ExpertInnen. In Fischer, R., Greiner, U. & Bastel, H. (Hrsg.), *Domänen fächerorientierter Allgemeinbildung*. 9-17. Linz: Trauner.

Frech, S. & Richter, D. (edit.). (2013). *Politische Kompetenzen fördern* (Didaktische Reihe). Schwalbach/Ts.: Wochenschauverlag

Himmelmann, G. (2001). Demokratie Lernen. Schwalbach: Wochenschauverlag

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70, 1-25, Netherlands.

Lengnink, K., Meyerhöfer, W. und Vohns, A., (2013): Mathematische Bildung als staatsbürgerliche Erziehung?, in Der Mathematikunterricht 59, Heft 4, Friedrichverlag

Negt, O. (2011). *Der politische Mensch. Demokratie als Lebensform* (2. Aufl.). Hannover: Steidl

Ridgway, J. (2015): Implications of the Data Revolution for Statistics Education. *International Statistical Review, 84(3), 528-549.* *http://onlinelibrary.wiley.com/doi/10.1111/insr.12110/abstract*

Ridgway, J., Nicholson, J. & Gal, I. (2017): Conceptual framework of civic statistics. *http://community.dur.ac.uk/procivic.stat/wp-content/uploads/2017/08/Conceptual_Framework_-_Aug_17.pdf*

Rumsey, D.J. (2002). Discussion: Statistical Literacy: Implications for Teaching, Research, and Practice. *International Statistical Review*, 70, 32-36.

Schiller, A. (2016). Entwicklung von Modulen zur Förderung von statistical literacy an der Hochschule. *Gesellschaft für Didaktik der Mathematik (Hrsg.) – Beiträge zum Mathematikunterricht 2016,* 851-854.

Schiller, A. & Engel, J. (2016). Civic statistics and the preparation of future secondary school mathematics teachers. IASE 2016 Roundtable Paper. *http://iase-web.org/documents/papers/rt2016/Schiller.pdf (28.3.2017)*

STATISTICAL EDUCATION NEEDS AND EXPERIENCES IN THE HUNGARIAN CENTRAL STATISTICAL OFFICE

Gábor Lovics, Katalin Szép and Zoltán Vereczkei
Hungarian Central Statistical Office
Keleti Károly utca 5-7. Budapest 1024, Hungary
gabor.lovics@ksh.hu
katalin.szep@ksh.hu
zoltan.vereczkei@ksh.hu

The statistical knowledge of official statisticians is very heterogeneous, there is no solid standard basic knowledge in statistics. This paper summarises the needs from statistical education, from the perspective of official statistics, based on experience gathered at the Hungarian Central Statistical Office. There are key areas where official statistics misses general knowledge and cooperation of official statistics and the teachers of statistics is absolutely necessary.

Official statisticians need sophisticated thinking in understanding and translating real social and economic phenomena into statistical concepts, a general overview of the statistical business process model, and special knowledge in certain tasks (like sampling, seasonal adjustment) and/or in some subject matter areas. A system of courses is offered, some examples (e.g. EMOS, ESTP) are mentioned in the paper.

OFFICIAL STATISTICS AND ITS CHALLENGES

Official statistics "*provide the European Union, the world and the public with independent high-quality information on the economy and society on European, national and regional levels and make the information available to everyone for decision-making purposes, research and debate"* (ESS, 2011, p. 6). With such role defined for the whole European Statistical System (ESS), the members of the ESS have safeguards in place to ensure the independence and high quality of the information provided by official statistics. The Hungarian Central Statistical Official (HCSO) is the leader institution of the Hungarian Official Statistical System, a member of the ESS. As an ESS member, the most important safeguards for the quality and independence of official statistics are the following:

- Legal safeguards: the main principles on the mandate, role of official statistics and the quality criteria of official statistics are ensured by legal acts both on European and national level.
- Statistical standards: the ESS, the UNECE, the OECD and the UNSD is adopting, using and promoting methods, classifications, handbooks, tools, structures and other instruments based on consensus and professional debate to increase comparability, transparency and credibility of official statistics.
- Demonstrated high quality of official statistics: quality of the statistical information provided by official statistics is available in the forms of quality reports and additional descriptive metainformation, accessible to the users and the producers of official statistics.
- Information provided by official statistics is public, therefore accessible to everyone. The main users of official statistics are the general public and decision-makers.

There are some constant and somewhat new challenges to official statistics (ESS, 2014; Vukovich, 2015). Some major challenges:

- More and more alternative statistical data: it is a common saying that everyone can produce their own statistics nowadays. For example, electronic forms and storage of information provide possibilities to build databases and produce statistics (for service providers like commerce, banks, mobile, cameras, etc.); the internet is a gold mine of information, access to databases and tools that helps us producing statistics is easier than ever before. Publishing these datasets produced by persons, market researchers and other companies can also be done

without real efforts. This is clearly a challenge for official statistics where the aforementioned safeguards give the edge for official statistics, even if it is constantly challenged by these alternative producers and solutions.

- Growing information need: the society always needs fast, more detailed, more frequent and generally, more kinds of data. This phenomenon is very significant now but it has been always a challenge for official statistics.
- Reduction of reporting burden: institutions developing, producing, disseminating official statistics generally gather information using statistical surveys, they are expected to always look for opportunities to lower the burden official statistics imposes on its respondents.
- The "timeliness versus accuracy" issue: users demand "immediate" results. Statistics produced faster usually means less accurate information than possible; increasing timeliness is therefore against the quality criterion of producing more accurate information.

These challenges are also contradictory in nature; it is not possible to meet all these challenges to full extent at the same time. Still, official statistics must provide solutions for these issues. One possible answer is differentiation between different users and different statistical outputs. In official statistics, we traditionally speak of three main types of users:

- Basic users: they typically understand the type of study or sources of official statistical information, the basic statistical measures and graphical representations. We also expect them to be able to find and understand underlying definitions behind the numbers.
- Intermediate users: they can understand but also handle the limitations of the methods used for the production of official statistics, are very well aware of the most commonly used statistical concepts and understand phenomena such as variability, uncertainty and probability.
- Advanced users: they understand the sophisticated terminology, are aware of sampling and non-sampling errors and can critically evaluate the statistical information and know what constitutes a valid statistical study.

EXPECTATION IN KNOWLEDGE AND SKILLS

In the era of Big Data, open data and open-source freely accessible IT tools it is also common to call the "ideal" official statisticians data scientists. In this regard, based on our experience and international references (ProCivicStat, GAISE), we identified five pools of knowledge and skills:

The first pool is the *sound statistical background* providing the foundation on which the official statisticians can perform the statistical tasks.

- Knowledge of different data sources and the methods to combine them to produce inputs, throughputs and outputs for official statistics;
- Good understanding of methods of cleaning, editing data and adjustment in large datasets;
- Methods to detect and manage atypical data in datasets (identify and handle outliers);
- Imputation techniques in order to choose from imputations methods and apply them on datasets;
- Creation of new statistical units, variables from collected, cleaned, processed data;
- Aggregation techniques to compile new aggregates;
- Modelling techniques.

Another pool is the *knowledge and skill basis needed both for the development and the production of official statistics*.

- Understand questions, raised in reality and translate them into statistical concepts (define statistically suitable concepts, its relationships with other similar concepts, identify and justification of divergences from international or national concepts, etc.);

- Make use of and define metadata (apart from concepts, be able to define and describe codelists, nomenclatures, classifications, correspondence tables, data sources, etc. in a way that users can interpret them when disseminated and also in order to serve as descriptive, structural, reference information in the statistical business process);
- Understand methods and make and evaluate important assumptions in estimations, data integration, modelling and data analysis;
- Select adequate methods and manage their implementation;
- Manage large (uncleaned/unstructured) datasets;
- Select relevant information from huge volumes of data;
- Draw conclusions based on quantitative analysis;
- Interpret statistical parameters, outputs.

A third pool is what we describe as *"affinity for IT"* that basically contains all the knowledge elements but more prominently skills to apply IT solutions in official statistics.

- Be able to use statistical software, not only Excel;
- Be able to understand outputs produced by statistical software;
- Using, building, maintaining databases;
- Basic SQL knowledge;
- Affinity for new technologies (open source tools, Hadoop);
- Deal with computational time issues (feasible algorithms, choice of software);
- Use of tools used for visualisation.

Another important pool is the collection of elements needed to *visualise and disseminate official statistics*.

- Sound knowledge on how to use charts, tables, maps;
- Defining concepts, methods in a clear and understandable way in the form of metadata;
- Innovation and contextual awareness;
- Story-telling skills;
- Ability to convert information from one form to another to fulfil different user needs (basic tables, graphs, microdata sets, infographics);
- Ability to analyse the statistical outputs in a complex way.

There is a fifth pool of knowledge and skills for official statisticians that is usually *specific to official statistics*. This knowledge is usually acquired in-house of the statistical institutions.

- System and fundamental principles of official statistics;
- Legal background, most important legal instruments;
- European Statistical System (ESS) and cooperation within the ESS and beyond (UNECE, OECD, etc.);
- ESS, international and national standards of official statistics; sources of information and value added;
- Quality criteria and quality management (ability to measure and assess product and process quality);
- Statistical confidentiality, data protection, IT security.

The "ideal" official statistician has a mixture of this knowledge and skillset. It is also important to mention that it is not one person that has all these skills but usually a small group of people, working together. Nevertheless, these knowledge and skill elements determine what is expected from official statisticians.

GAPS IN KNOWLEDGE AND SKILLS

In our experiences, *the knowledge base of the majority of the new graduate students are very far from the "ideal" official statistician in Hungary*. Based on our experiences, this problem can be a result of different kinds of gaps.

First of all, we would like to underline the fact that *there are some specific knowledge elements for official statistics, as listed above, which naturally cannot be a part of a usual university education* (maybe with the exception of courses being specific to official statistics). This knowledge is typically to be learned within the National Statistical Systems and not in the universities.

It is very hard to discuss in general about the statistical teaching in the Hungarian higher education system since every university is different. Nevertheless, *the second type of gap is between the ideal curricula of universities and the current ones*. Using big (cleaned & uncleaned / structured & unstructured) datasets instead of small well-prepared data; problem-solving instead of mechanic use of methodology; understanding the purpose, methods of estimations and analysis instead of memorising and using some steps of the process; ability to learn new methodology and to choose methodology are fields which should be more prominent parts of the curricula.

The biggest gap we can identify is *between the curriculum of Hungarian universities and the real knowledge of some students who have just finished their studies*. In economic and sociological education (where most of the staff of the HCSO conducted their studies) everybody learns statistics, nevertheless a lot of them does not know more than at the end of their high school education (example: they understand mean but even the concepts of median and modus might cause problems in practice). The main problem within this situation is that *if someone has a degree it will not give us any information about his/her real knowledge*. This person can either be someone well aware of the knowledge acquired in the education, or someone who can barely remember a thing from his/her studies.

Although we understand that the knowledge of the graduated students cannot be the same, but some minimal standards about the graduated should exist. In our point of view *the minimal set of knowledge and skills should contain the following*:

- Ability to understand questions raised in reality and translate them into statistical concepts, forms;
- Basic data manipulation techniques;
- Data collection, missing data, outliers;
- Using, building, maintaining databases;
- Descriptive statistics knowledge;
- Basics of hypothesis testing;
- Basics of regressions;
- Be able to use at least one statistical software;
- Be able to understand outputs produced by statistical software.

MANAGING GAPS

From a long-term perspective, the best way is to be proactive to prevent future problems and *have a collaboration in place between official statistics and teachers of statistics*. More specifically:

- Identify needs in knowledge and skills; based on these, make recommendations to universities to change, add/remove material and incorporate the needs of official statistics into the curricula. Official statistics should provide input, reference material and even datasets to help to incorporate these needs into statistical education;
- Collaborate with universities in a close way to share knowledge, human resources, look for applications of emerging methods and tools.

There are several reasons, why gaps – to some extent – are inevitable to always be present. *Therefore the collaboration should focus on the minimalisation of gaps*. Some inevitable sources of gaps are the following:

- Adequate knowledge becomes obsolete fast, due to quick changes in available methods and tools;
- For special jobs/work areas especially deep knowledge is needed in specific areas;
- There are some specific internal tools and standards in institutions that can be known and taught to employed people.

In official statistics, there are several initiatives and courses to overcome the shortage of knowledge and skills. Here are some examples.

HCSO School

As in many statistical institutions, in the HCSO a yearly updated system of internal training courses is present; it is intended to cover gaps mentioned before. The courses of the so-called "HCSO School" are provided by experts of the HCSO for the employees of the HCSO. The short training courses (mainly 1-3 days long) cover general statistical topics – to upgrade basic knowledge and deepen analytical skills – and general professional knowledge (language, IT, personal behavior, etc.), institution-specific IT systems and tools, statistical standards and methods.

As the international standards in official statistics are more and more important and there is a shortage of experts in new emerging methods – especially in small countries –, the international cooperation in training is a must, presented in the next two examples.

European Statistical Training Programme[1]

This system is managed and outsourced by Eurostat, courses are held by experts from different Member States, the participation is free for employees from official statistical authorities. The 2-3 days trainings in a number of topics (43 in 2017) aims to meet the challenges of relevant and comparable statistics at European and international level (ESTP, 2017). This is a flexible tool, easy to address fast changes of needs in the supply of courses. Some topics/examples:

- Standards and subject-matter domain-specific knowledge: European standards like national accounts, environmental accounts, Labour Force Survey, classifications, quality reporting, metadata solutions;
- Statistical methods and tools: seasonal adjustment, JDemetra+, output checking, data analysis;
- Information standards: SIMS, SDMX, R, EDAMIS;
- Emerging topics for official statistics: use of Big Data, how to become a data scientist, data integration (including geospatial information) methods and tools.

All of these can be considered as complementary short courses to the more compehensive and systematic training program for official statisticians – in the next example.

European Master in Official Statistics[2]

The main objective of the European Master in Official Statistics (EMOS) is to develop a network of EMOS master programmes providing post-graduate education in official statistics at European level. EMOS is a joint project of universities and data producers in Europe.

As the key starting issue, the learning outcomes (EMOS, 2015) have been identified in five groups: the system of official statistics; data production models and methods; specific topics; statistical methods; dissemination. This range of knowledge and skills represents the ideal foundation for the development of professionals able to interpret the fast-changing official data

[1] For more information on the ESTP programme, please visit http://ec.europa.eu/eurostat/web/european-statistical-system/training-programme-estp.

[2] Detailed information on EMOS is available at https://ec.europa.eu/eurostat/cros/content/emos_en.

production system of the 21st century. The Board of EMOS manage the use of EMOS label. EMOS continues to provide services in developing learning materials, providing literature and knowledge base, producing webinars, and organizing conferences, traineeships.

CONCLUSIONS

Based on the challenges of, expectations from official statistics and the aforementioned reflections of expected knowledge and skills, we identify three main conclusions:

- *Development of an MA curricula is necessary*. It is very important to build more bridges between reality and the education of statistics. Based on our experiences it is important to cover the whole statistical business process of official statistics, from raising a question, through developing and conducting data collection, processing, analysis to finding answers for the initial questions. In this regard awareness of international and national standards (concepts, classifications, of international and national bodies) is a must. Students should be able to manage large datasets (structured, unstructured; cleaned, uncleaned), select and apply adequate methods and tools on them. Graduated students also need to be open and adaptive in a fast changing environment.
- *Development of after-graduation courses is necessary*. It is important that graduates make themselves familiar with new data sources, new tools and methods. International cooperation is important for such after-graduation courses to cover niches, new areas where there are only a limited number of experienced professionals.
- *Development of internal training courses in institutions is necessary*. Institutions need to upgrade and level up the general statistical skills and knowledge of their employees. There are always special, institution-specific knowledge and skills for which training courses need to be provided in-house. These trainings should naturally offer special knowledge for the given jobs (in some cases even international cooperation is preferable).

As most of the issues expressed also in this paper is mostly common for teachers of statistics and official statisticians, *it is important to find the opportunities to discuss these common issues, arrive at common statements and start cooperation on more specific issues*. The Hungarian Statistical Association[3] and the Hungarian Academy of Sciences – Scientific Committee of Statistics and Futures Studies[4] could provide these opportunities for people interested in the teaching of statistics.

REFERENCES

EMOS (2015). Learning outcomes of the EMOS Programmes. December 2015. *https://ec.europa.eu/eurostat/cros/system/files/emos_learning_outcomes.pdf*

ESTP (2017). European Statistics Training Programme. October 2017. *http://ec.europa.eu/eurostat/web/european-statistical-system/training-programme-estp*

European Statistical System (ESS) (2011). *European Statistics Code of Practice* – revised edition. September 2011. *http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15*

European Statistical System (ESS) (2014). ESS Vision 2020. *http://ec.europa.eu/eurostat/documents/7330775/7339647/ESS+vision+2020+brochure/4baffcaa-9469-4372-b1ea-40784ca1db62*

Gabriella Vukovich (2015). Big Data and Official Statistics. *Hungarian Statistical Review. Volume 93* (number 8-9), 745-758. (available in Hungarian only)

Guidelines for Assessment and Instruction in Statistics Education (GAISE, USA). August 2005. *http://www.amstat.org/education/gaise/*

[3] Please visit http://www.mstnet.hu/

[4] Please visit http://mta.hu/english/scientific-sections-105963

ProCivicStat material (2017). October 2017. *http://community.dur.ac.uk/procivic.stat/index.php/publications/*
UNECE (2014). Fundamental Principles of Official Statistics. July 2013. *https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx*

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

LIFELONG LEARNING DEVELOPMENT LEVEL IN SELECTED EUROPEAN COUNTRIES: A PERSPECTIVE FOR IMPROVING STATISTICAL LITERACY

Ksenija Dumičić and Berislav Žmuk
Faculty of Economics and Business, University of Zagreb, Croatia
Trg J. F. Kennedyja 6, HR-10000 Zagreb, Croatia
kdumicic@efzg.hr
bzmuk@efzg.hr

The higher the lifelong learning development level, or the higher the participation rate of adults in education and training in a country, the higher is the chance that the employment rate will be higher, since the more persons are competitive on the labour market and the greater the chance for a national economy to be more competitive. The aim of the paper is to inspect whether European countries with higher average lifelong learning development level have also higher average economic development level. The main variable under study is the Participation rate in education and training of people aged from 25 to 64 years for 2014, based on Eurostat data. The descriptive statistics analysis had shown significant differences in the participation rates in education and training between the observed European countries in 2014. An increasing lifelong learning of adults' development level performs a great chance for improving statistical knowledge in all the European countries studied.

Keywords: *cluster analysis, economic development level, descriptive statistics, European countries, adult participation in lifelong learning*

JEL classification: *C38, D83*

INTRODUCTION

Involvement in education and training is a measure of lifelong learning, which includes all learning activities started throughout life with the goal of improving knowledge, competences and skills, regarding personal, civic, social or employment-associated views. The goal to learn is the basic point that differentiates these activities from non-learning activities.

The research hypothesis of the paper is that the European countries with higher average lifelong learning development level have also higher average economic development level. Also, the higher level the lifelong learning of adults, the higher opportunity for statistical education to be included into it and statistical literacy and knowledge of adults to be improved.

Recently, issues as statistical literacy, knowledge and thinking are discussed widely. Statistical literacy is a key ability expected of citizens in informed societies, and is often taken as an expected outcome of educating and as a necessary component of adults' literacy. In Gal (2002) statistical literacy is portrayed as the ability to interpret, critically evaluate, and communicate about statistical information and messages. It is argued that statistically literate performance is predicated on the joint activation of five areas: literacy, statistical, mathematical, context, and critical thinking, together with a collection of supporting natures and enabling beliefs. Educational and research implications are discussed, and responsibilities facing educators, statisticians, and other stakeholders are outlined. Gould (2017) promotes statistical literacy as necessary because the role and nature of data have changed, so the definition includes prerequisite for "algorithmic culture", which stands in contrast to the traditional inference culture. Schield (2017) explains the 2016 revision of Guidelines for Assessment and Instruction in Statistics Education (GAISE), which eliminated pure statistical literacy as a stated goal by including multivariate thinking and promoting statistical thinking being part of statistical literacy.

According to Eurostat, adult participation in lifelong learning is defined as the share of population aged 25-64 who take part in education and lifelong learning. Statistics about adult participation in lifelong learning can be drawn from four main datasets. Detailed comparison among the Labour Force Survey (LFS), Adult Education Survey (AES), Continuing Vocational Training Survey (CVTS) and OECD Programme for the International Assessment of Adult Competencies (PIAAC) is given in the report of the European Commission (2014). This technical

briefing deals with adult participation in lifelong learning. In particular, it focuses on the implications associated to the use of different statistical sources (LFS, AES, CVTS and PIAAC), characterized by different reference periods and different definitions of lifelong learning.

According to Dumičić (2017), where statistical literacy is treated as a unique language for a better world, meaning that the enhancing of the statistical literacy may improve knowledge needed for better citizenship, and according to the analysis shown in this paper, lifelong learning may be considered as an opportunity for improving statistical literacy and knowledge, and all this is related to the development level in considered countries.

LIFELONG LEARNING DATA, METHODS AND ANALYSIS RESULTS

In the paper the lifelong learning as participation rate in education and training of people aged from 25 to 64 years is observed. In order to better understand the participation rate in education and trainings, it is inspected by observing it from different angles: by sex (male, female), by employment status (employed, unemployed), by educational attainment level (primary, secondary, tertiary), and by degree of urbanisation (city, rural areas). The full list of 10 lifelong learning variables is listed in Table 1.

Table 1 *List of selected lifelong learning variables, for population 25 to 64 years (last 4 weeks)*

No.	Code	Variable brief description	Source
1.	LLTotal	Participation rate in education and training, percentage of total population	Eurostat (2015a)
2.	LLMale	Participation rate in education and training, percentage of males	
3.	LLFemale	Participation rate in education and training, percentage of females	
4.	LLEmp	Participation rate in education and training, percentage of employed persons	Eurostat (2015d)
5.	LLUnemp	Participation rate in education and training, percentage of unemployed persons	
6.	LLEduc1	Participation rate in education and training, percentage of persons with less than from 25 to 64 years.	Eurostat (2015c)
7.	LLEduc2	Participation rate in education and training, percentage of persons with upper secondary and post-secondary non-tertiary education (levels 3 and 4)	
8.	LLEduc3	Participation rate in education and training, percentage of persons with tertiary education (levels 5-8)	
9.	LLCity	Participation rate in education and training, percentage of persons living in cities	Eurostat (2015b)
10.	LLRural	Participation rate in education and training, percentage of persons living in rural areas	

In the analysis data for overall 33 European countries are used. So, the lifelong learning data are included for all 28 European Union member states, plus the Former Yugoslav Republic of Macedonia (FYROM), Iceland, Norway, Switzerland, and Turkey. Furthermore, the lifelong learning variables are observed in the period from 2005 to 2014. Still, in the descriptive statistics and outlier analysis more emphasis is given to the most recent data (the year 2014). Unfortunately, data for all countries for the whole observed period are not available.

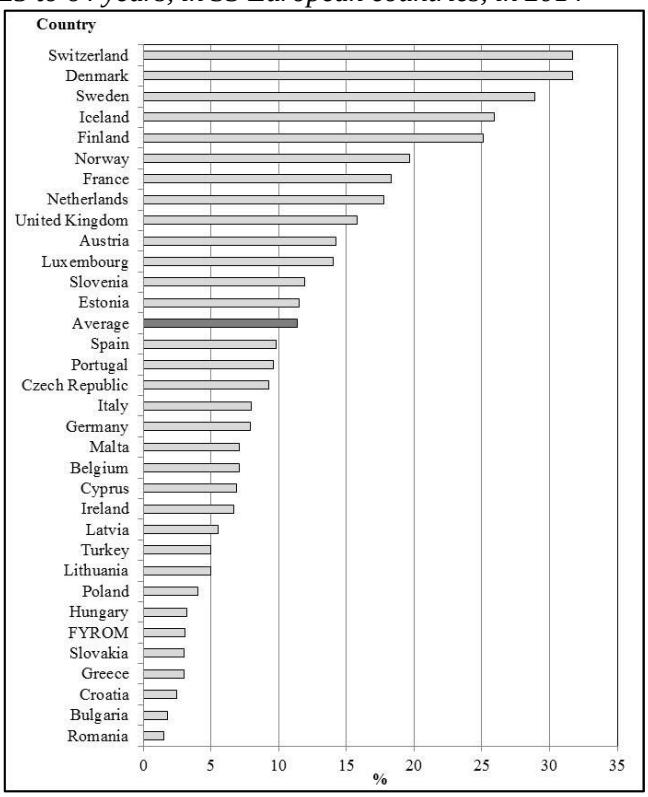
After conducted descriptive statistics and outlier analysis, using the selected lifelong learning variables, a cluster analysis is performed. The cluster analysis will enable recognizing of European countries groups which have similar lifelong learning development level. In the cluster analysis the non-hierarchical approach is used.

In order to observe the strength of impacts of lifelong learning change on the economic development, the correlation and regression modelling is conducted. In the analysis as a measure of the economic development level variable Gross domestic product (GDP) per capita in purchasing power parity (PPP) (World Bank, 2015) was used. Lifelong learning development level or participation rate in education and training of adults should be as high as possible.

The higher the participation rate is, the more persons have got new, relevant and specific knowledge. Consequently, these persons became more competitive on the labour market and they are more valued than before. Furthermore, the competitiveness level of the national economy also becomes higher. All these lead to higher development level in a country and to higher well-being level of its citizens. In Figure 1 the lifelong learning (variable LLTotal) in the observed 33 European countries for the most recent year for which data are available is shown.

Figure 1 revealed great disproportion in the participation rates in education and training between the observed European countries. In the 33 European countries together in average 11.4% of total population aged from 25 to 64 years participated in education and training in 2014.

Figure 1 *Lifelong learning as participation rate in education and training, percentage of total population aged from 25 to 64 years, in 33 European countries, in 2014*



Source: Authors' creation, Eurostat.

Table 2 *Basic descriptive statistics results of Participation rate in education and training, percentage of total population aged from 25 to 64 years, in 33 European countries, in period from 2005 to 2014*

Statistics	Year									
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
N	31	33	33	33	33	33	33	33	33	33
Mean	10.48	9.95	9.92	10.22	10.20	10.55	10.87	11.08	11.45	11.41
Std. dev.	8.23	8.27	8.07	8.21	8.05	8.75	8.57	8.68	8.74	8.93
Coef. var.	78.55	83.06	81.38	80.34	78.85	82.92	78.81	78.35	76.28	78.27
Skewness	1.03	1.12	1.11	1.06	1.01	1.11	1.06	1.06	0.98	1.05
Kurtosis	-0.16	0.09	0.13	0.00	0.03	0.28	0.28	0.15	-0.06	0.10
Minimum	1.3	1.3	1.3	1.4	1.4	1.2	1.3	1.3	1.7	1.5
1st quartile	5.1	4.2	4.4	4.7	4.4	3.9	4.4	4.5	4.3	5.0
Median	7.4	6.8	7.0	6.9	6.8	7.2	7.5	7.4	7.8	8.0
3rd quartile	15.6	15.0	14.8	13.9	14.6	16.2	15.7	14.1	16.1	15.8
Maximum	27.6	29.2	29.0	29.9	31.2	32.5	32.3	31.6	31.4	31.7

Source: Authors' calculation, Eurostat.

From Table 2 it could be concluded that the average of percentage of total population aged from 25 to 64 years which participated in education and training steadily increase from 9.95% in 2006 and 9.92% in 2007 to 11.45% in 2013 and 11.41% in 2014. The average participation rate increased from about 10%, at the beginning of the observed period, to the about 11.5% in the recent years. However, the coefficients of variation are higher than 75% in all observed periods. According to median, in 50% of the observed countries 8.0% or less of total population aged from 25 to 64 years participated in education and training in 2014 whereas in 50% of the countries more

than 8.0% of total population aged from 25 to 64 years participated in education and training in 2014. The lowest values of variable LLTotal convincingly had Bulgaria and Romania in the observed period. In all 10 observed years the value of variable LLTotal was not higher than 1.8%. Denmark and Switzerland have not the value of variable LLTotal lower than 22.5% in all periods.

Table 3 *Basic descriptive statistics results of selected lifelong learning variables, percentages, in 33 European countries, in 2014*

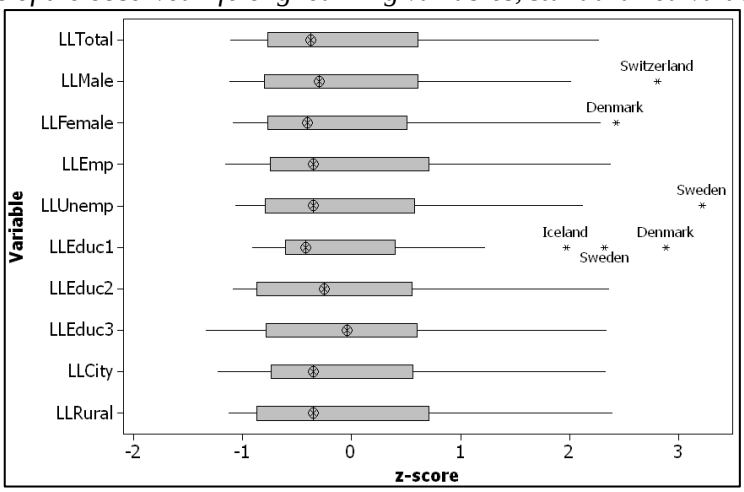
Statistics	Variable								
	LLMale	LLFemale	LLEmp	LLUnemp	LLEduc1	LLEduc2	LLEduc3	LLCity	LLRural
N	33	33	33	33	30	33	33	31	31
Mean	10.30	12.53	11.99	11.29	5.72	10.01	18.00	14.09	9.94
Std. dev.	7.79	10.28	9.28	9.82	5.97	7.74	11.23	9.74	8.27
Coef. var.	75.62	82.05	77.39	86.99	104.32	77.37	62.40	69.12	83.28
Skewness	1.10	1.14	0.96	1.40	1.56	0.98	0.72	0.99	0.94
Kurtosis	0.62	0.33	-0.02	2.11	1.86	0.02	-0.24	0.01	-0.06
Minimum	1.6	1.3	1.2	0.8	0.3	1.6	3.0	2.1	0.6
1st quartile	4.5	4.9	5.5	3.7	2.1	3.3	9.4	7.1	2.9
Median	8.0	8.3	8.7	7.9	3.2	8.1	17.5	10.7	7.0
3rd quartile	14.2	17.4	17.4	16.2	7.8	13.1	24.6	19.5	15.3
Maximum	32.2	37.5	34.1	42.9	23.0	28.3	44.3	36.8	29.7

Source: Authors' calculation, Eurostat.

Table 3 shows basic descriptive statistics results of the other 9 selected lifelong learning variables when all selected European countries are observed together in 2014. Unfortunately, there are some missing data. So, there are no available data for the variables LLCity and LLRural for the FYROM and Turkey. Considering variable LLEduc1 there are no available data for Bulgaria, Lithuania and Slovakia whereas for Croatia it was used the most recent available data from 2009. Similar, at the variable LLUnemp for Lithuania data from 2013 was used as an estimate for 2014.

According to the coefficients of variation values, which are presented in Table 3 and which are considerably higher than 30%, it can be concluded that data variation level at each variable is high. Consequently, the median should be consulted instead of the average.

Figure 2 *Box plots of the observed lifelong learning variables, standardized values, in 2014*

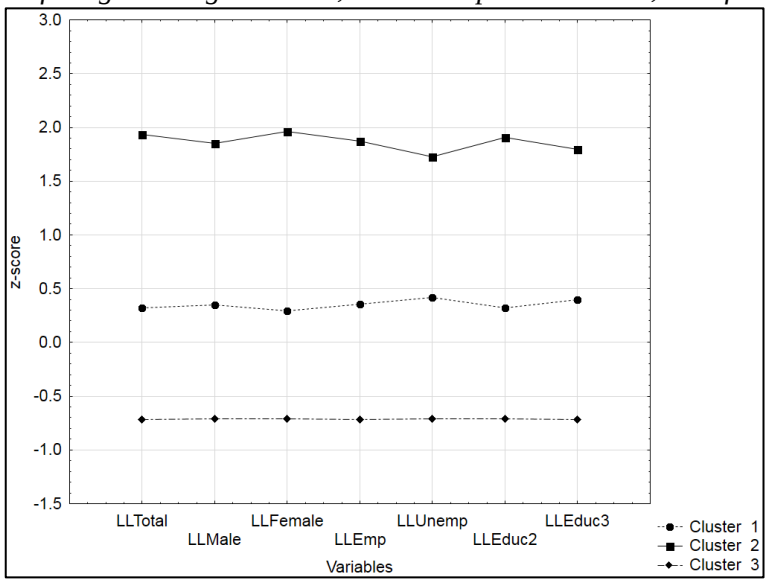


Source: Authors' calculation, Eurostat.

The box plots of the observed lifelong learning variables, which are shown in Figure 2, suggest that there are six outliers. However, only one of emphasized data can be defined as serious outlier. Namely, the LLUnemp variable value for Sweden deviates more than three standard deviations from the variable average (z=3.22). In Sweden 42.9% of unemployed aged from 25 to 64 years participated in education and training in 2014. It has been estimated that this data should not have significant impact on the analyses which, so, Sweden was not omitted from the further analysis. Whereas the outliers are not a problem here, the missing data are. Namely, the question

here is to omit from the further analysis countries for which data is missing (five countries) or to omit variables with missing data (three variables). It has been decided to conduct non-hierarchical cluster analysis by using both approaches and then to compare results to see if some significant differences exist.

Figure 3 *Plot of means for each cluster, non-hierarchical cluster analysis, K-means clustering method, 7 selected lifelong learning variables, in 33 European countries, data from 2014*



Source: Authors' creation, Eurostat.

Table 4 *Classification of countries in the clusters, non-hierarchical cluster analysis, K-means clustering method, 7 selected lifelong learning variables, in 33 European countries, data from 2014*

Cluster 1 10 countries	Cluster 2 5 countries	Cluster 3 18 countries
Austria, Estonia, France, Luxembourg, Netherlands, Norway, Portugal, Slovenia, Spain, United Kingdom	Denmark, Finland, Iceland, Sweden, Switzerland	Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, FYROM, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Turkey

Source: Authors' creation, Eurostat.

Table 5 *Classification of countries in the clusters, non-hierarchical cluster analysis, K-means clustering method, 10 selected lifelong learning variables, in 28 European countries, data from 2014*

Cluster 1 5 countries	Cluster 2 7 countries	Cluster 3 16 countries
Denmark, Finland, Iceland, Sweden, Switzerland	Austria, France, Luxembourg, Netherlands, Norway, Slovenia, United Kingdom	Belgium, Croatia, Cyprus, Czech Republic, Estonia, Germany, Greece, Hungary, Ireland, Italy, Latvia, Malta, Poland, Portugal, Romania, Spain

Source: Authors'.

According to Table 4 and Table 5 there are slight movements of European countries between clusters regarding their economic development.

CONCLUDING REMARKS

According to the analysis shown in this paper, increasing lifelong learning development level, which is positively related to and all this is related to the development level in considered countries, may be considered as an opportunity for improving statistical literacy and knowledge. The conducted analysis has shown that there are great differences in lifelong learning achieved

level over the observed European countries. The increasing trend in lifelong learning level in the observed European countries overall is noticed. The lifelong learning development level in the 33 observed European countries increased in last 10 years from 10% to 11.5%. The two conducted cluster analyses have shown that three groups of countries according to achieved lifelong learning development level can be recognized. Unfortunately, the same or similar distinction between countries could not be made when economic development level, measured by Gross domestic product (GDP) per capita in purchasing power parity, was taken into account. Consequently, the research hypothesis of the paper could be only partially accepted because strong relationship between achieved lifelong learning development level and achieved economic development level was not found.

The higher the participation rate of adults in education and training in a country, the higher is the chance that the employment rate will be higher, since the more persons are competitive on the labour market and the greater the chance for a national economy to be more competitive. All these lead to higher well-being level of citizens. The statistical education for adults, employed or unemployed, should be improved in the European countries. Since, statistics is considered as a new key competence for lifelong learning enhancing employability and the ability to remain employable throughout life, the role of statistics educators is important, related to government support, ministry of education, etc., and the educators become very responsible, especially for those that design and organize educational programs in statistics for all society segments. In future research lifelong learning perspective for improving statistical knowledge should be investigated and implemented more.

REFERENCES

Dumičić, K. (2017). *Enhancing Statistical Literacy as a Unique Language for a Better World*. Plenary Speech. In: Kremić, E., Kozarić, K. (Eds.), Book of Abstracts ICOS2017- The International Conference on Official Statistics, Sarajevo, B&H, 30.-31.03.2017, School of Economics and Business in Sarajevo, University of Sarajevo, B&H, pp. 19-21.

European Commission (2014). *Adult Participation in Lifelong Learning - The impact of using a 12-months or 4-weeks reference period*. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92330/lbna26918enn.pdf (June 1, 2017)

Eurostat (2015a). *Participation rate in education and training (last 4 weeks) by sex and age*. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=trng_lfse_01&lang=en (November 4, 2015).

Eurostat (2015b). *Participation rate in education and training (last 4 weeks) by sex, age and degree of urbanisation*. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=trng_lfs_14&lang=en (November 4, 2015).

Eurostat (2015c). *Participation rate in education and training (last 4 weeks) by sex and educational attainment level*. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=trng_lfse_03&lang=en (November 4, 2015).

Eurostat (2015d). *Participation rate in education and training (last 4 weeks) by sex and labour status*. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=trng_lfse_02&lang=en (November 4, 2015).

Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: SAGE Publications.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review, 70*(1), 1-5.

Gould, R. (2017). Data Literacy is Statistical Literacy. *Statistics Education Research Journal, 16*(1), 22-25.

Schield, M. (2017). GAISE 2016 Promotes Statistical Literacy. *Statistics Education Research Journal, 6*(1), 50-54.

World Bank (2015). *GDP per capita, PPP (current international $)*. http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (November 5, 2015).

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

ACTIVITIES FOR PROMOTING CIVIC STATISTICAL KNOWLEDGE OF PRESERVICE TEACHERS

Daniel Frischemeier, Susanne Podworny and Rolf Biehler
Paderborn University
dafr@math.upb.de

Critical statistical thinking and a sustainable knowledge in civic statistics is inevitably to become a concerned citizen. This kind of thinking and knowledge can already be enhanced in secondary school. For the implementation of civic statistics in mathematics classrooms in secondary school, teachers themselves have to be well educated in the field of civic statistics. For this purpose, we have designed and realized a university course about civic statistics in the winter term 2016/2017 at the University of Paderborn, where preservice teachers worked on projects and activities in regard to civic statistical contexts. For instance, they have analyzed official open data of the German Statistical Office on the German gender pay gap with digital tools, explored the distribution of net assets in German households and investigated the unemployment situation in different countries in the European Union via Google Public data files. In this paper, we will present some activities and derive first implications for re-designing these activities.

INTRODUCTION

Since decisions in society, politics and economy are often based on civic statistics, responsible citizens in the sense of "Mündigkeit" (responsibility, emancipation) (see Schiller 2017) need statistical and also, more specifically, civic statistical skills. When we talk about civic statistics we mean statistics about key phenomena in society such as employment, health, education, social welfare or inequality (Ridgway 2016; Engel 2017). The process of critical thinking in regard to civic statistics contexts is preferably supposed to start already at school level. In Germany mathematics teachers are not necessarily educated in this special area of statistics education. To teach civic statistics, teachers need not only statistical content knowledge but also pedagogical content knowledge, technological knowledge and a positive stance towards civic statistics. The project ProCivicStat, funded by the ERASMUS+ program of the European Commission, aims at supporting teachers with specific courses, materials, tools, and datasets for teaching civic statistics (www.procivicstat.org).

At the University of Paderborn, we have the following situation. There is a compulsory course on elementary statistics and probability and a compulsory course on didactics of statistics in students' bachelor studies. In their master studies, preservice teachers can choose a seminar which is supposed to deepen and expand their knowledge they have gained in the compulsory courses.

GENERAL INFORMATION

We have designed the seminar "Statistical literacy in mathematics classroom" in the frame of a Design-Based-Research setting (Cobb, Confrey, diSessa, Lehrer, & Schauble 2003) and have realized the first cycle in winter term 2016/2017, the next -second- cycle will take place in winter term 2017/2018. In this paper, we refer to the cycle of winter term 2016/2017 only. Our seminar had 21 participants in the winter term 2016/2017. All participants have been preservice teachers for mathematics in lower secondary school and have been at the end of their studies, having successfully attended the course on "Elementary statistics" and "Didactics of statistics". The seminar consisted of 15 sessions, each session lasted 90 minutes. The main idea of the seminar was to build on the statistical content knowledge our participants have gained in the previous courses so that the participants on the one hand can apply their statistical content knowledge in civic statistics contexts and on the other hand develop a specific pedagogical content knowledge, so that the preservice teachers get prepared to implement civic statistics ideas into mathematics classroom at secondary school. For distinguishing the several knowledge domains, see for example (Wassong and Biehler (2010)).

So, there are learning goals on two dimensions learning goals in regard to statistical content knowledge, and learning goals in regard to pedagogical content knowledge.

Learning goals of the course

Our learning goals with regard to statistical content knowledge are

- to deepen students´ knowledge about reading and interpreting summary statistics and graphical displays (also in the sense of reading beyond data of Friel, Curcio, & Bright 2001).
- to introduce students into statistical concepts and constructs (like correlation and causality or Simpson´s paradox) relevant in civic statistics.
- to introduce into the definition and operationalization of concepts such as unemployment.
- to explore multivariate datasets on the base of given and self-generated statistical questions.

Our learning goals with regard to the pedagogical content knowledge are

- to consider contents in civic statistics across subjects.
- to get to know relevant material (articles, links, tools, datasets, etc.).
- to learn to "simplify" complex situations in civic statistics for classroom use.
- to develop ideas for implementing civic statistics activities in classrooms.

The general goal is to evoke a critical thinking towards statistics and analyses given for example in media reports.

General design principles

The underlying general design principles of our course are similar to the principles of the Statistical Reasoning Learning Environment from Garfield and Ben-Zvi (2008, p.48). For instance, we focus on the development of central statistical ideas, we use real and motivating data sets, we use classroom activities to support the development of our students´ reasoning, we integrate the use of technological tools, we promote classroom discourse and we also promote assessment to monitor the cognitive development of our participants.

REALIZATION OF THE COURSE

In this paragraph, we will describe the realization of some of the sessions in our course.

Sessions 1-3: Introduction

To confront our participants with civic statistics issues immediately we started the course with the task ("Interpret the statistical display in the context of inequality in German net assets") to interpret a complex statistical display showing the distribution of net assets in Germany in the years 2003 and 2012 (see Figure 1).

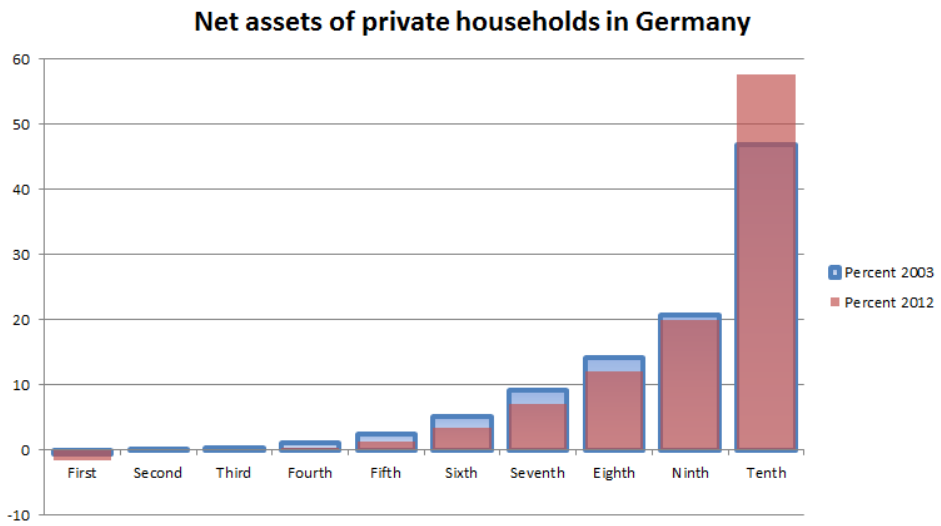
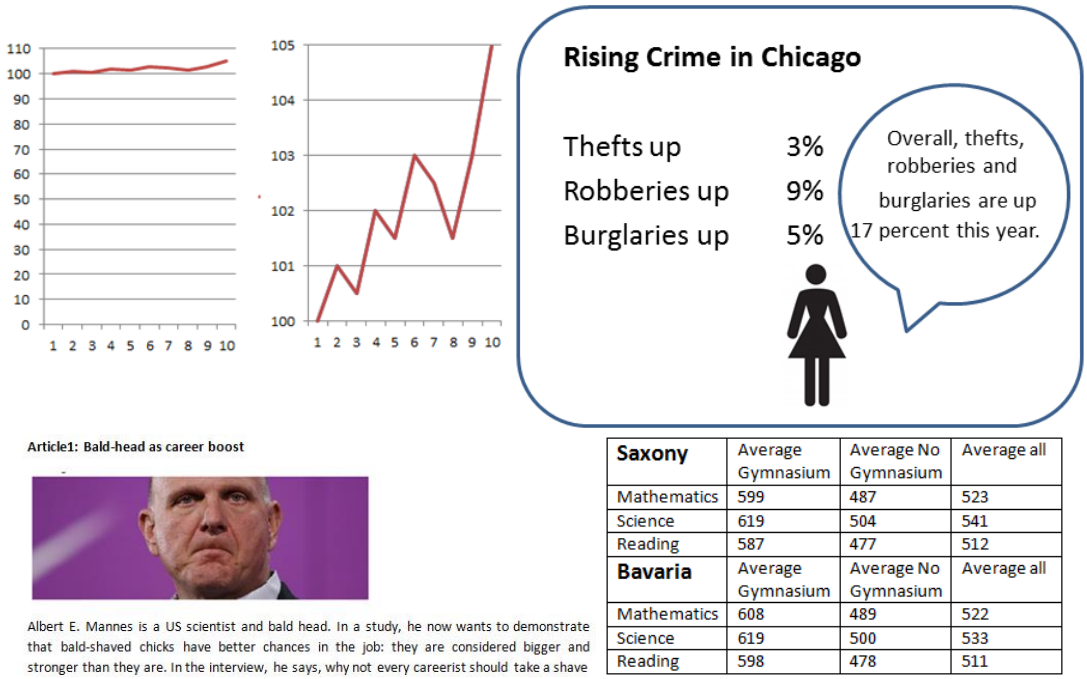


Figure 1: Diagram of the task on German net assets, similar to the diagram in https://crp-infotec.de/deutschland-vermoegensverteilung/

The display in Figure 1 shows the distribution of net assets in households in Germany in 2003 (see blue bars) and 2012 (see red bars). One first competence learners need is to recognize that the bars displaying the situation in year 2003 and 2012 are overlapping. The distribution is divided in ten deciles which are ordered ascending from left to right. First interpretations might be that in most deciles the blue bars are higher than the red bars – but having a look at the tenth decile it is the other way around: the rich people have become even richer in the nine years from 2003 to 2012. Another interpretation might be that there is a big difference in net assets between 90 percent of the German population and the richest 10 percent. Our participants worked on this task in session 1 and 2 and when observing the working processes of our students, it was obvious that this task was very challenging for the students. There is a need to understand the definition of "net assets", of "household", and of "deciles". A difficulty was to compare two overlapping bar graphs showing the growing inequality between 2003 and 2012. In the next session 3, we wanted to refresh the technological Fathom knowledge. The students were familiar with using the German version of the software in their previous courses (Biehler, Hofmann, Maxara, & Prömmel 2011). Our idea was that our participants use Fathom for their explorations and that they refresh their technological Fathom knowledge when exploring a real dataset on leisure time activities of German 11th grade students (Biehler, Kombrink, & Schweynoch 2003). So, for instance in session 3, our participants had to work in pairs to investigate the question "in which way do boys and girls differ in interest with regard to different leisure time activities (e.g., in playing games on the computer)".

Sessions 4-9: Students sessions

In sessions 4-9, students as session leaders were responsible to design and moderate the sessions. We have had sessions on representation of data (session 4), percentages (session 5), percentages II (session 6), correlation & causation (session 7), Simpson's paradox (session 8) and on the concept of unemployment (session 9). In Figure 2, we get an impression on the different tasks and topics in these sessions.



Saxony	Average Gymnasium	Average No Gymnasium	Average all
Mathematics	599	487	523
Science	619	504	541
Reading	587	477	512
Bavaria	Average Gymnasium	Average No Gymnasium	Average all
Mathematics	608	489	522
Science	619	500	533
Reading	598	478	511

Figure 2. Examples of four different activities in sessions 4-9

In the upper-left corner we see an example diagram of session 4 on manipulating statistical displays similar to the diagram in Krämer (2007, p.38-39) – both graphs show the same data with different axes. In the upper-right corner we see a diagram implemented in session 5 showing the wrong use of percentages (see Bauer, Gigerenzer, and Krämer (2014, p.19)). In the lower-left

corner we see an excerpt of a German online magazine (*Spiegel online*) article (see http://www.spiegel.de/karriere/interview-mit-albert-e-mannes-glatze-hilft-der-karriere-a-876879.html - in Figure 2 we present our English translation of it) assuming a relationship between bald-heads and career boost in the session on correlation and causation. Finally in the lower-right corner we see PISA data in the subjects mathematics, science and reading from two German federal states, which turns out to be an instructive example of Simpson's paradox. The common structure of all sessions was that the student session leaders began with an introduction (~ 5 minutes), followed by a presentation that was supposed to refresh relevant statistical knowledge (~ 10 minutes) and providing examples for the theme (~15 minutes). After these inputs there was a working phase where all participants worked on activities in small groups (~ 30 minutes) followed by a plenary discussion of the results (~20 minutes). Each session 4-9 concluded with a reflection on the session (~10 minutes).

Sessions 10-13: Gender Pay Gap project

After attending to the sessions 1-9 and after gaining statistical knowledge in regard to representation of data, percentages, correlation & causation, Simpson's paradox and the concept of unemployment, our aim was to provide our participants with a more complex task. We have chosen the gender pay gap situation in Germany as we identified this as a meaningful topic for young adults.

Specifically, we wanted our students to explore the causes of the unequal pay situation between male and female employees in Germany and we wanted them to become familiar with the concept of the gender pay gap so that our participants are able to distinguish between the adjusted and unadjusted gender pay gap. The unadjusted pay gap means that it is about 23% that men earn more than women in Germany. Furthermore, we wanted our participants to explore the German income structure data set from the German statistical office, to learn to reflect reports in the media critically and to relate them to their own data explorations. In total, we had four sessions dedicated to the gender pay gap project. In session 10 our participants informed themselves by reading media and internet articles about the definition and explanations of the gender pay gap in Germany. For session 11 and 12 we provided our participants with a random stratified sample of all German employees downloaded from the German statistical office and containing about 60,000 cases with variables like gender, wage per month, region of Germany, kind of employment, age, etc. In addition, we provided five topics (profession, function, age, economy and region) according to the variables in the dataset and asked our participants to choose one of these topics – for example the topic "age" (see the precise task in Figure 3).

Project on the Gender Pay Gap - Aspect: Age

Work in teams of two!

Now, you are to carry out a project work on the gender pay gap with your knowledge gained in the seminar. In doing so, you should independently explore the data set for the 2006 Income Structure Survey and get insights into possible explanations for the gender pay gap on the basis of the available data.
You have learned that the differences in income between male and female workers, which are published in the media, have to be interpreted with caution because of the different factors that determine the difference.

Your TASK

In this article (see link below), the focus is on the factor "age", which has an influence on the differences in income. Under this perspective, examine the present data set and work out the extent to which income differences are caused by the aspect mentioned above. In addition, try to explore other aspects that affect differences in income.

Source/Link: http://www.bild.de/ratgeber/job-karriere/gehalt/wie-alter-und-geschlecht-ihr-gehalt-bestimmen-44537794.bild.html
Write a short article and create a PowerPoint presentation that you will present to your fellow students.

Figure 3. Task G*ender Pay Gap Project* (aspect: age)

In this task our participants were asked to explore the gender pay gap data in peers with Fathom and to work out in which way e.g. the aspect age might have an influence on the gender pay gap in Germany. To document their results and to be able to present them to their classmates we asked our participants to prepare a PowerPoint-presentation.

In session 13 the participants have presented their findings to their classmates via their PowerPoint presentations.

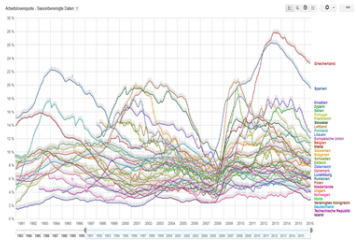
Sessions 14-15: Mini projects

Our course concluded with the sessions 14 and 15 where our participants worked in small groups on mini projects. These mini projects have the intention that the participants can apply all their competencies they have gained during the course in small projects using interactive graphs and tools that can be found on the Internet for free. One example of a mini project can be seen in Figure 4. It uses an interactive visualization from Google Public Data.

Unemployment in the EU

Please go to:

https://www.google.com/publicdata/explore?ds=z8o7pt6rd5uqa6_&ctype=l&strail=false&bcs=d&nselm=h&met_y=unemployment_rate&fdim_y=seasonality:sa&scale_y=lin&ind_y=false&rdim=country_group&idim=country_group:eu:non-eu&idim=country:no:is:cy:uk:hu:cz:es:si:sk:se:ro:pt:pl:at:nl:mt:lu:lt:lv:it:ie:el:fr:fi:ee:de:dk:bg:be&ifdim=country_group&tstart=665190000000&tend=1472508000000&ind=false



and explore the interactive visualization

Questions:

- **Q1:** What trends can you see in the quote of unemployment in Germany from 1991 to 2015? Compare this with the trend of the EU and another country of the EU, chosen by yourself. Inform yourself about the definition of unemployment in the country.
- **Q2:** Think of a further question with regard to other variables (e.g. Age, Gender).
- **Q3:** What do you see? What could be causes for this?

Figure 4. Mini project on the rate of unemployment in the European Union

Mini projects had the goal to analyze a specific topic with open data and free visualizations from the internet to enhance a whole group discussion about the topic in the last session.

In the mini project of Figure 4, our participants should use Google Public Data to compare the development of unemployment rates in different countries of the European Union. In particular, our participants were asked to compare the unemployment rate of Germany to the unemployment rates of other European countries they could choose on their own. The last session covered the presentation of findings of the mini projects and whole group discussions about the context of the presented findings, for example about unemployment rates in different European countries.

CONCLUSION & FURTHER PLANS

We can state that our participants worked statistically on many civic statistics contexts. Especially in the project work of the Gender Pay Gap, our participants have been really engaged. The evaluation, which has been done in form of filling out evaluation sheets at the end of each

session (for details see Biehler, Frischemeier, & Podworny 2017), shows that our participants liked the exploration of German income structure data and the presentation of their findings via Power Point very much.

In the next winter term 2017/2018, we plan to teach a redesigned course. In this course, we will keep the general structure of the previous one, but we plan to support our students in a more concrete way, especially when designing the students´ sessions and to implement more project sessions, since our participants have worked on these activities very engaged, liked them very much and see potential in them to implement these activities in their further teaching.

REFERENCES

Bauer, T., Gigerenzer, G., & Krämer, W. (2014). *Warum dick nicht doof macht und Genmais nicht tötet: Über Risiken und Nebenwirkungen der Unstatistik*: Campus Verlag.

Biehler, R., Frischemeier, D., & Podworny, S. (2017). *Design, realization and evaluation of a university course for preservice teachers on developing statistical reasoning and literacy with a focus on civic statistics*. Paper presented at the World Statistics Congress 61, Marrakech, Marocco.

Biehler, R., Hofmann, T., Maxara, C., & Prömmel, A. (2011). *Daten und Zufall mit Fathom: Unterrichtsideen für die SI mit Software-Einführung*. Braunschweig: Schroedel.

Biehler, R., Kombrink, K., & Schweynoch, S. (2003). MUFFINS: Statistik mit komplexen Datensätzen - Freizeitgestaltung und Mediennutzung von Jugendlichen. *Stochastik in der Schule, 23*(1), 11-25.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9-13.

Engel, J. (2017). Statistical Literacy for active Citizenship: A Call for Date Science Education. *Statistics Education Research Journal, 16*(1), 44-49.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education, 32*(2), 124-158.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning. Connecting Research and Teaching Practice*. The Netherlands: Springer.

Krämer, W. (2007). *So lügt man mit Statistik*. München: Piper.

Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education. *International Statistical Review, 84*(3), 528-549.

Schiller, A. (2017). *The Importance of Statistical Literacy for Democracy – Civic-Education by Statistics*. Paper presented at the Challenges and Innovations in Statistics Education, Szeged.

Wassong, T., & Biehler, R. (2010). *A Model for Teacher Knowledge as a Basis for Online Courses for Professional Development of Statistics Teacher*. Paper presented at the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

INCREASING INTERACTIVITY IN CLASS

Dániel Kehl, Ferenc Kruzslicz and Roland Baczur
University of Pecs - Faculty of Business and Economics
kehld@ktk.pte.hu

Increasing interactivity in large size classes is a major issue in our teaching experience. One possible tool is an Audience Response System (ARS). This short paper aims to put the use of this tool in a theoretical framework, gives an overview of different products to consider, introduces a pilot-study from the University of Pécs and describes our first experiences.

GENERAL OVERVIEW

Problem description

Any person who deals with any form of teaching has already encountered one of the following problems:

- large classes;
- passive students;
- lost attention.

Different problems may work both as cause and effect. Cost effective management of high number of applicants result in large size classes/groups. It is inevitable that people with different habits should be included in these groups, which can lead to loss of attention and the passivity of certain students. Lecturers cannot interact with all of them individually, which can also lead to the loss of attention

However, these problems can be remedied in a number of ways, and such a solution may be the Flipped Classroom model.

Theoretical framework (Jackie Gerstein, 2012)

Briefly, the Flipped Classroom is a strategy to reverse the traditional learning environment usually by delivering instructional content (often online), outside of the classroom. The model is basically a cycle of learning model. The cycle often begins with an experiential exercise. This is an authentic, often hands-on learning activity that fully engages the students: they become interested in the topic because of the experience, they have a desire to learn more. The next step is the What phase: students are exposed to and learn concepts touched upon during the Experience, they explore what the experts have to say about the topic (the information is presented via online/videos/websites/etc.). The third phase is the So What step: students reflect on their understanding of what was learnt during the previous steps. The final step of the cycle is the Now What: during this phase, the students will show what they have learned and try to apply based on what they mean to them. This goes beyond reflection and personal understanding in that students must create something that is individualized and extends beyond the lesson with applicability to the students' everyday lives.

In this framework, the so-called knowledge transfer is not executed in the usual form and new tools, new opportunities appear. Although, this leads to another problem: how could these new tools be integrated in teaching?

Possible solutions

We believe that a possible solution to the problem is the use of ARS (Audience Response Systems). It helps to create a two-tiered relationship between the presenter and her/his audience. It is needed to have all the member of the audience the possibility to be able to vote online in the simplest possible way. The use of ARS involves many options: for instance, the listeners can join in with a simple vote during the presentation (they can participate in the course, *Figure 1*.), to indicate if they have lost the thread somewhere. The presenter will be able to check presence (authentication is possible) and the opportunity of embedded questions related to the presentation is given. There is the opportunity for the audience to influence the lecture flow and to get extra credits. The advantage of the system is that it can be seamlessly integrated with different LMS systems. With the help of ARS people can learn more efficiently.

Figure 1. Learning pyramid. (Infographic published by Neil Beyersdorf on Linked in 2015)

Available solutions

This system is already available from the technological side in the market, and many providers offer similar opportunities. The main features are:

- Hardver-, software-based or hybrid
- Embedding into presentation tools
- Question and answer types
- Additional features

The main products are:

- Optivote

It seemed a promising solution: it offered a hardware-based solution to which a special device was needed. The price of the device was relatively cheap, but the service was not. The project was discontinued after 2010. The moral of a fable: expensive, extra hardware is unnecessary.

- Adobe Connect

A software-based solution that allows full video/audio integration, image and application sharing. There is no need to install it, you just have to pay for it. There are plenty of extensions available for it.

- Poll Everywhere

This product can be used to access various online polls from various devices such as browsers, Twitter, and mobile. Web and PowerPoint embedding is available. The results can be displayed in several ways, including a geolocation mapping.

- Turning Technologies

Hybrid Solution: combines hardware and software-based technologies.

- SOPRESO

Primarily a presentation sharing system (PDF, Keynote, Prezi, PPT, etc.). The audience can also initiate various issues, such as questions, problems, etc. Social network authentication is also possible. Hungarian innovation and development.

- Learning Catalyst

Specifically linked to educational and eLearning/Presentation environments. It is innovative and contains many types of questions. It operates based on the principle of Flipped Classroom and Student-Student Interaction. Independent of the operating system, only a browser needed.

- Moodle survey
- kahoot.it
- etc.

PILOT AT THE UP

First of all, it is necessary to find out from the student side, whether there is a need for an ARS service. For testing purposes, a group of 70 students was tried to determine if they are susceptible to voting. The response rate for the open method questionnaire was not influenced by the fact that the scores were compiled by an instructor or by a student. Behind this phenomenon, the students cannot assume their answers and opinions before each other. Afterwards, it was possible to look at the different products more closely.

Time schedule

In the organization of our Faculty, we tried to try out several products to learn its potential benefits and disadvantages and to see that given product in a crisp situation. Within this opportunity, we have participated with our colleagues in the following programs:

- Pearson Learning Catalyst demo (Nov. 2015)
- SOPRESO demo (Dec. 2015)
- Turning Technologies demo (Feb. 2016)
- kahoot.it demo (Mar. 2016)

Basically, all demos had their own advantages and disadvantages. We also received positive feedback from our colleagues. The charm of novelty and the welcome of the students will be the question.

Students' expectations

With a survey, we also tried to assess students' needs. A total of 235 fills were generated, and their aggregated results are shown in the following 3 figures.

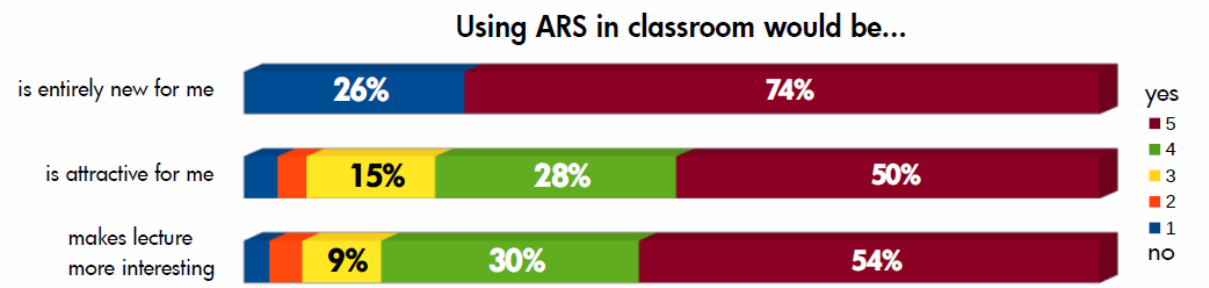


Figure 2. Using ARS in classroom would be…

Figure 2. shows that the introduction of ARS into a given course would be a whole new world for most of the students, but it would be an attractive value for the majority and could make the presentation more interesting. Involving new technologies in education may seem like a novelty, but with the advancement of technology both sides need to develop.

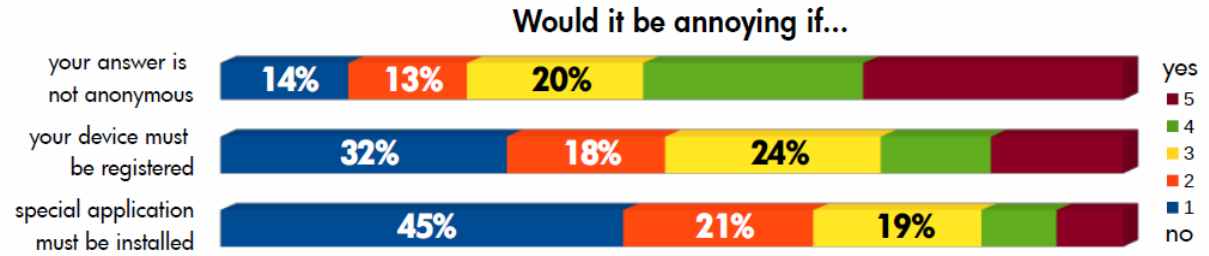


Figure 3. Would it be annoying if…?

Based on Figure 3., anonymity plays an important role (we got similar results from our classroom experiment). However, it may be interesting to note that installing a third-party application would not be annoying (neither the registration of that device).

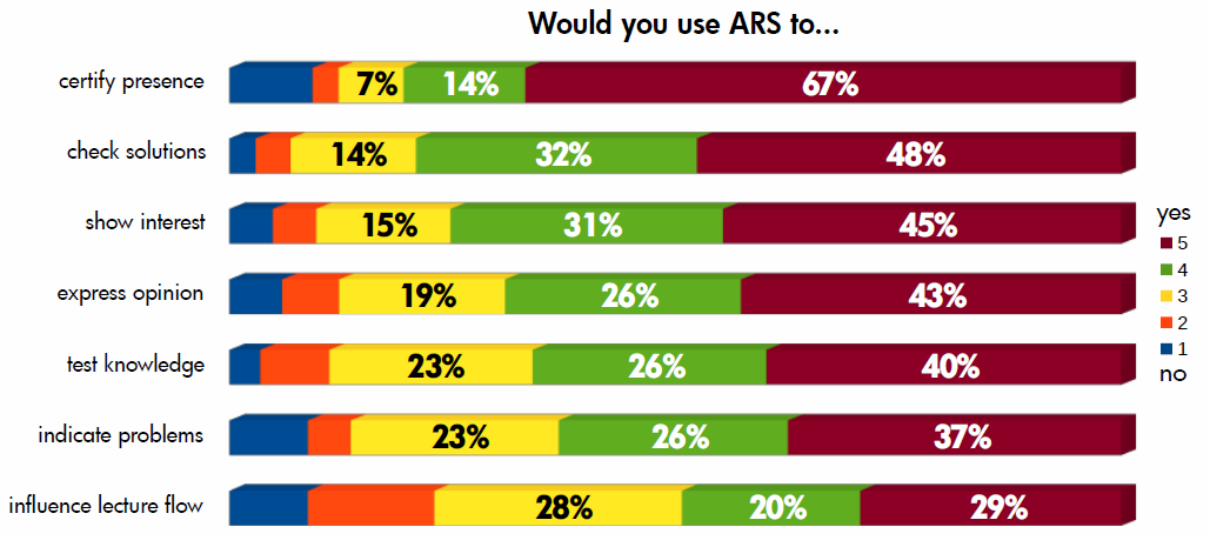


Figure 4. Would you use ARS to…?

Also from Figure 4. we can read the results that confirming our tentative experiment. Students are less likely to use ARS technology to change the flow of the lecture, rather they consider it as an administrative tool (confirmation of presence). What is more of a matter of interfering with questions in the lecture flow, the more the number of insecure responses and the proportion of yes replies decreased.

Most importantly, the survey found out: anonymity for students is an important thing and they do not really want to take the responsibility to have a say in the class' flow. This is a very interesting mentality.

Lecturers' expectations, opinions

After the students, we were curious about the colleagues' opinions, so different possibilities were discussed. Some respectable (older) colleagues have showed resistance. Fear was basically because of the need to allow mobile devices to be used in such systems, and in some cases their use should be encouraged. If we did not want to use mobiles, then in case of „clicker"-based solutions the ownership and distribution of devices is a question. It is another problem with the building such system: is it used by every single colleague or just the majority? For further uses, how should the data have obtained from the in-class polls? At the same time, it would also be important to have an opportunity to switch between anonymous and unanonymous use – once the students experience the unanonymus mode, then everything else is just a matter of trust.

Turning Technologies „clickers"

We have been able to use this technology for a free trial period: we received a receiver and a few clickers. It was basically a pleasure for the students, it was easy to manage, they considered the tool interesting. Rather, it was considered useful at BA level than the master, where interactivity is not (such) a problem. The question really is how much losing its interest in continuous use – we are not sure about long-term usage. Interactivity is limited but supported, it can be used for presence-checking, quick tests. The price of 45 euro/device plus receivers. The students argued that they would rather support smart devices (though they also know that they do not have 100% coverage). There is a chance of cheating (of course not in the investigated groups), especially if the tests were counting on half-yearly performance as well).

Kahoot

A free, web and mobile device-based voting system. It's easy to use: participants need to answer simple choice questions only at the moment. No pre-registration required, e.g. Neptune code can also be used. More and more colleagues start using it in different ways. One semester was tested with Probability and Statistics „readiness tests" for extra credits – it's also a proxy for presence. The results were promising (Figure 5.).

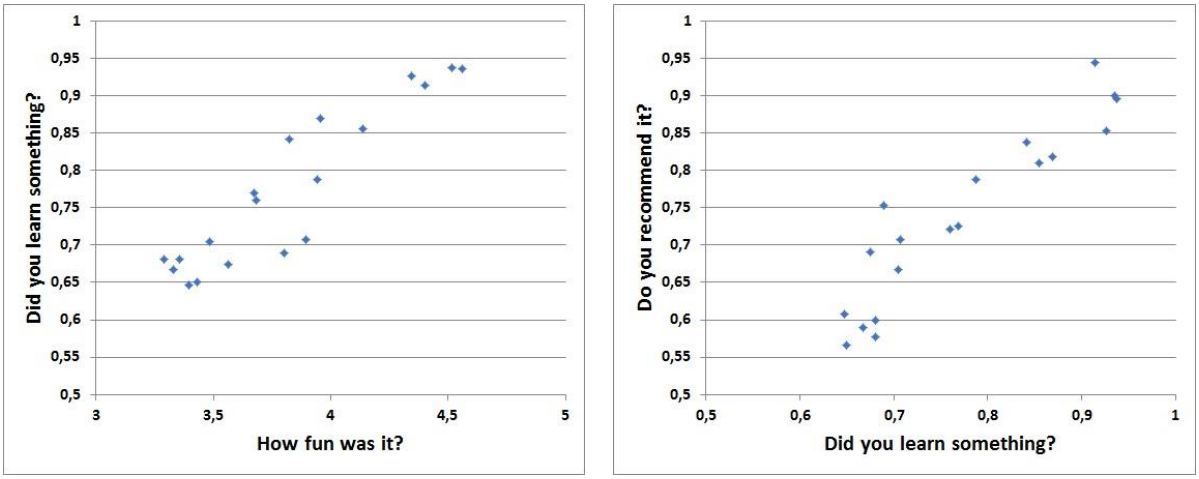


Figure 5. Correlation results from different KAHOOT Quizzes.

Each point in the figure represents a lecture with a KAHOOT test. Each test ended with a survey where students were able to express their opinion about the questions: they marked on a 1 to 5 scale that how they liked the test (quiz) or not, did they learned something, or did they recommend it for somebody else. There is a clear correlation between the fun-learn and learn-recommend relations.

CONCLUSION

Basically, our suggestion is to start using some free solution first. The participation from the lecturer's side should not to be forced, instead it should be voluntary. After a while, when the "critical mass" is reached, one might consider a flexible paid service. It should also be noted that different importance/techniques are needed for smaller and larger groups. The students seem to like it if it is not overused, but using them at the right time is a great tool to break the long lecture and shake the audience. Last but not least, it is a great possibility for students to show their excellence.

As a result of our testing, we found that bandwidth proved to be the most important bottleneck after all: without the expansion and development of the Wi-Fi network, only mobile solutions could be considered – in connection with this fact, a project for the development of our Wi-Fi network was also launched.

REFERENCES

Gerstein, Jackie (2012). *The Flipped Classroom: The Full Picture*. Amazon Digital Services LLC.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

FLIPPED CLASSROOM TEACHING IN STATISTICS

Mónika Galambosné Tiszberger
tiszbergerm@ktk.pte.hu

Classical lectures are rather inefficient for the members of the Z-generation. They do not feel any motivation in it and it is usually not helpful in the learning procedure. They need much more chance to interact in the classroom and this interaction must be forced. We also need to make them realize that during their university studies basically they will not be taught but they have to learn on their own to reach deep, advanced level knowledge and skills by the end. The balance of the responsibility of their studies must be pushed toward them, so they will be able to handle their future life responsibilities as well. If the students are motivated to study before the "lectures", there is a chance in the classroom for a real discussion of the topics. The aim of this presentation is to introduce some techniques, methods that can be used in a "flipped classroom", together with the experiences of the previous semester in a 2+2 type statistics course.

INTRODUCTION

Classical lectures in the universities are rather inefficient for the members of the Z-generation. It can be seen in every field of higher education that students do not feel any motivation in it and this attitude is usually not helpful in the learning procedure. These young adults need much more chance to interact in the classroom and this interaction must be forced. We also need to make them realize that during their university studies basically they will not be taught but they have to learn on their own to reach deep, advanced level knowledge and skills by the end. The balance of the responsibility of their studies must be pushed toward them, so they will be able to handle their future life responsibilities as well. If the students are motivated to study before the "lectures", there is a chance in the classroom for a real discussion of the topics.

This paper presents a "good practice" from University of Pécs, Faculty of Business and Economics. First the idea of the flipped classroom teaching methodology is introduced. In the second part the reader can find a detailed description of the flipped classroom tools that are applied in a course called Business Statistics (within our English programme). The results and feedbacks of the students and the feeling and experience of the teacher are discussed in the third section. Finally some conclusions are formulated together with advises for those who would like to have flipped classroom.

ABOUT THE IDEA OF FLIPPED CLASSROOM

The idea of the flipped class started with lecture being done at home with the help of video and/or audio files, and what was once considered homework is done in class. So, the order of the lecture and homework components of the class are flipped. In other words, learning is done at home and deeper understanding, practicing is done at school, with the help or guidance of the teacher. So the main aim is to provide more time in the classroom for real learning and understanding.

This method is transferring the responsibility of learning from the teacher to the students. If the student has the freedom to decide about the timing and way of learning, it will more belong to him/her. The role of the teacher will be to guide the understanding, and students become active learners rather than just sitting in a classroom without doing anything. (Alvarez, 2011)

I had the feeling for a few semesters now, that lectures make less and less sense. The students hear the terminology, the idea and methodology of certain statistical problems for the first time during the lecture. Even if they pay attention, only a few of them would catch the main ideas. We all know, that simply asking them to read the book before the lecture is not working. I have read about the flipped classroom methodology, but I had to see it in practice to really understand it. I had a chance to visit Capilano University in Vancouver, Canada to see more types of courses applying this technique in a certain way. This real life experience made me start to think of "my own" flipped classroom.

The basic tools for a flipped classroom, as I see:

- very well defined learning material (book, book chapters, articles, etc.), which is available for students before the beginning of the course;
- pre-prepared, short videos about the most important topics or about the steps of exercise solving (also available for students before the beginning of the course);
- assessment of preparation even before class
- interactive atmosphere supported by
- group work (random group members is the best),
- name tags to be able to call anybody,
- little tasks to involve more and more students;
- group exercises which aim deeper understanding of the problems and solutions or higher order of critical thinking;
- diverse assessment scheme;
- trust in the method and trust in yourself.

As it can be seen the basic idea has several tools and solutions. Obviously a teacher in a certain course might not use all of the tools, but only those that fit in the topic, nature, type of the course, and most importantly to the personality of the teacher. Based on my own experience, I have to say that this latter point is indeed very important. Any type of teaching methodology will only work if the teacher believes in it, and feels comfortable in the classroom. So everybody should find a personalized tool set for his/her temper.

FLIPPED CLASSROOM IN THE COURSE CALLED BUSINESS STATISTICS

I have introduced this method in a course called Business Statistics at University of Pécs, Faculty of Business and Economics, which is for second year BA students in our English programme. We have 2+2 class format (2×45 minutes of lectures and 2×45 minutes of seminars each week, for 13 weeks). The group size was about 50.

The idea of preparation from the students' side before they actually come to the class was the basic of the main concept from my point of view. In the course outline they received a detailed schedule about what and when they should read. As a motivation, they had to fill in a quiz of the given topic before the lecture (it was closed 1 hour before the lecture). Beside the motivating function I also had the chance to check their answers and see, which parts were clear, or which parts were unclear to most of the students. It helped me to focus on the unclear parts during the lecture period. They had another quiz from the same topic after we have discussed and practiced the given material together, so we could see, how they developed. (Altogether they had 20 quizzes during the semester, each worth 0.5 points.) I did not prepare short videos, however I clearly declared the book chapters/pages that need to be read before the class. It was never longer than 25 pages (with pictures, exercises, etc.).

At the beginning of the first class everybody created his/her own nametag. One side it contained the first/nickname of the student, with huge letters, and on the other side the full name. The name tags were taken from a box at the beginning of the class, it was hanging in the students' neck during the class, and it was replaced to the box at the end. The nametags have several useful roles:

- the creation was fun with colorful papers;
- the teacher and the classmates can call everybody by name;
- those nametags, that remain in the box indicates the missing students, so no additional check of presence is necessary.

In the lecture period all the students were together. Before each class a random group schedule was introduced, so from week to week different four students had to work together. Randomness makes it a bit more complicated for the students to sit down in the given structure, however, it has many advantages:

- they have to get to know each other;
- Hungarian speaking and non-Hungarian speaking students were mixed up, so they were forced to discuss everything in English;
- they experienced how to work in a group, they had the pressure on them to prepare for the classes, otherwise they could not participate in the work, which might have been a an inconvenient situation that they want to avoid later;

- the distribution of "good" and "not so good" students were random, so they had the chance to learn from each other, or to realize that the material can be understood.

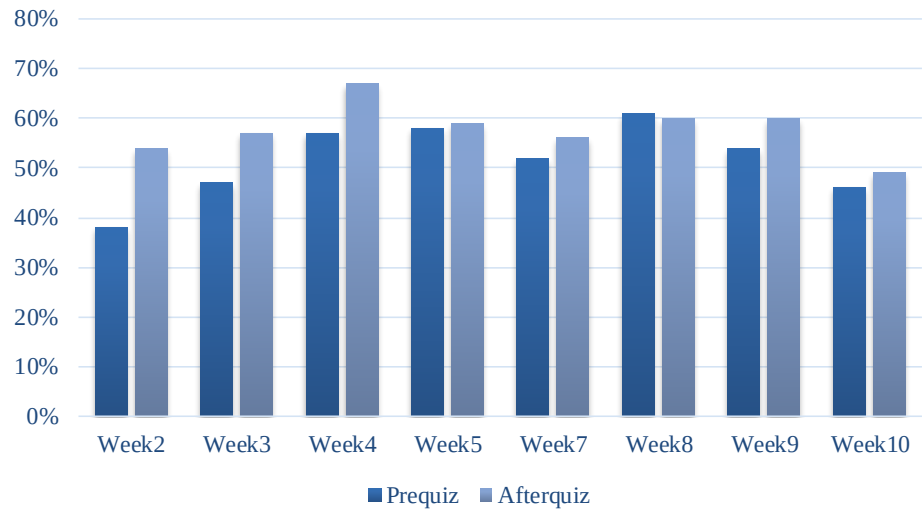
During the lectures there were no lectures. A question or small subtopic was introduced, then the groups had to discuss it in 5-8 minutes. After it their conclusions were discussed together. The teacher cleared up or summarized the information. There were mainly 3-4 sets of these rounds. There was an additional exercise for the groups in the last 10-15 minutes, which they had to submit in paper before they have left. The solutions were always discussed at the beginning of next week's class. The group works were evaluated each week, and every member received the same amount of points. Altogether the group exercises worth 10% in the overall grade.
The tutorials or seminars were held in a computer lab in two groups. They were more like regular seminars: solving exercises with the help of Microsoft Excel using different datasets.
The assessment scheme is as follows:

- 10% quiz results;
- 10% group exercises;
- 15-15% midterm tests;
- 50% final exam (additional condition: 50+% has to be gained in the final exam in itself).

RESULTS AND FEEDBACKS

The final results of the course were better in average as in the previous years, and there were much less grade 2 (pass) and much more grade 5 (excellent).
The participation in the classes is not really a good indicator, as in our English programme students must attend at least 75 percent of the classes to have permission to participate in the final examination. So teaching methodology basically makes no difference, because they are forced to be there anyway.
The participation rate of the quizzes was between 70 and 85 percent. This means that students felt the responsibility of fulfilling the requirements. The detailed results of the "prequiz" and "afterquiz" can be seen in figure 1. The average of the "prequiz" results is 52 percent, which means for me that students really took some effort in the preparation for the classes. In each week the "afterquiz" results were higher compared to the "prequiz", so students could learn more during the courses. And it is also clear, that there were some remaining tasks in studying as the mean of the "afterquiz" results is not more than 60 percent.

Figure 1: Results of the quizzes by weeks



Source: Own edition from Neptun results

Altogether the performance of the students was better compared to the pervious experiences (I teach the course now for 9 years). As the structure, the material, the exercises and the examination

is generally the same, we can conclude that flipped classroom techniques inspired the students for more learning and understanding.

There is an official evaluation form that students have to fill in for every course they have at the end of each semester. According to the overall average results I gained my personal best. The students have the chance to give comments in two questions. For the question "What was the worst aspect of the module?" nobody had put anything. For the question "What was the best aspect of the module?" I received the following comments (all comments are quoted word by word, and only those are mentioned, that are in connection with the teaching methodology):

- "Because we have quizzes, so we can review knowledge well."
- "We have the group work, we can share our idea with each other."
- "Regular quizzes, two midterms. It required continuous work (to do well)."
- "That we had quizzes and got group works every lecture to achieve points."
- "The number of midterm and quizzes helped to describe materials."
- "We have the group work, we can share our idea with each other."

It might sound strange, but I think that students do not really pay attention to the weight of certain evaluation elements, but the nature of the exercise, experience is what matters. They really enjoyed the group works and worked hard to have them solved in the last ten minutes of the classes. Most of them filled in the quizzes week by week, even if they were not so successful. Most of them studied hard for the midterm tests. It seems they did not have the thought that they should consider the final as "most important" just because it is 50 percent of the points. It means that is worth to give them "little" challenges, because they will work on them, and – perhaps sometimes unaware of the fact that they are learning – they gain the knowledge in the meanwhile.

CONCLUSION

My own, and it seems that also most of my students' experience is very positive of the elements of the flipped classroom teaching methodology. It gently forces them to catch up with the speed of the learning from week to week. The classes are more personal and interactive, so the topics get closer to them. Each student had a chance to talk aloud at least once during the semester. I, as a teacher, also enjoyed the classes because we had much more chance to discuss things, and many good questions came up, which was unimaginable in the way of "regular" lectures.

I continued to use this model in this semester as well, and I try to apply some of the elements in other courses as well. The success seems to be similar, which verifies my own feelings. *(We have a so called "Board of Studies" meeting in the middle of each semester in the English programme, where the student representatives have the chance to present mainly their problems of the different modules. This autumn the second year students – to whom I teach Business Statistics now – highlighted that they are really like the way of teaching, and enjoy the classes. This was a real honour.)*

I can only recommend this methodology to every teacher within a certain group size. It can be a good starting point, to visit some classes where you can see the techniques in practice. You can simply personalize the tools by selecting those that you like, or those that you can manage. And from semester to semester you have the chance to expand or modify the tools to fit the students' and your needs in the best way.

REFERENCES

Alvarez, B. (2011). Flipping the Classroom: Homework in Class, Lessons at Home. *Education Digest*: *Essential Readings Condensed For Quick Review*, 77, 18-21.

"Emberi Erőforrások Minisztériuma Supported by the ÚNKP-17-4-III New National Excellence Program of the Ministry of Human Capacities"

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

PERSONALIZED EXAMS IN PROBABILITY AND STATISTICS

Gallusz Abaligeti and Dániel Kehl
80 Rákóczi str, Pécs, 7622, Hungary
kehld@ktk.pte.hu

In this paper we introduce an example of using free open-source software solutions to create and grade personalized exams in probability and statistics. The statistical computing software R and the exams package is used together with Moodle, leading open-source learning management system. At the moment we use spreadsheets to store the text of the exercises and choose randomly amongst those. As a result our students face different stories (text), different numbers and potentially different questions but very similar structures. Results of the midterms on average are very similar to the previous paper and pencil results and students like the immediate feedback.

INTRODUCTON

At the University of Pécs, Faculty of Business and Economics BA students have to take two semesters of introductory statistics. Covered materials include basic probability calculations, descriptive statistics, inferential statistics, measures of association, correlation and regression analysis. From year to year the number of enrolled students is higher and higher, going over 400 recently. The computer lab can sit approximately 50-60 students both for practicing, midterms and finals. The two midterm tests focus on the ability of computing different statistical measures, applying methods and hence can be automated. The final focuses more on explanation of results and more comprehensive tasks so it is a mixture: paper and pencil answers based on computer aided calculations, but emphasis is on explaining and interpreting results. The high number of students and low number of seats at the computer lab result in a great number of groups in case of a midterm which means a lot of different versions of the same exercises. In line with the effort of our University to create a student-friendly learning environment we also offer midterm retakes, meaning even more exercises are needed. One year ago we decided to personalize the midterms so all students have their own tasks, leaving less chance of cheating or copying each other's solutions.

Preparing and especially grading midterms took a significant amount of working hours before using R (R core team, 2016) and Moodle (Moodle HQ, 2017). Learning the features of the exams package (Zeileis et al., 2014) and developing our solution clearly was a great initial effort but after having a decent number of reusable midterm problems creating, assigning and grading midterms happens almost automatically. Most of the students want to see their papers to understand which exercise was correct and which was incorrect. In case of over 400 students this procedure also took a lot of time. Using the automatic grading system of Moodle students see their solutions and the correction immediately after submitting their answers. They are given some time to go through these after the midterm and they can flag those questions where they see some problems with the correction of their answers (there are always some typos etc.). From the teachers' side it is a lot easier and quicker to go through those flagged questions in Moodle which saves a lot of time. According to our experiences students like the immediate results of their midterms.

In the next section we introduce a detailed example of creating an exercise and creating midterms using the above mentioned open-source software. We do not want to go into details regarding the exams package, there is a documentation (Gruen – Zeileis, 2009) and great tutorials (http://www.r-exams.org/tutorials)

A DETAILED EXAMPLE OF CREATING MIDTERMS

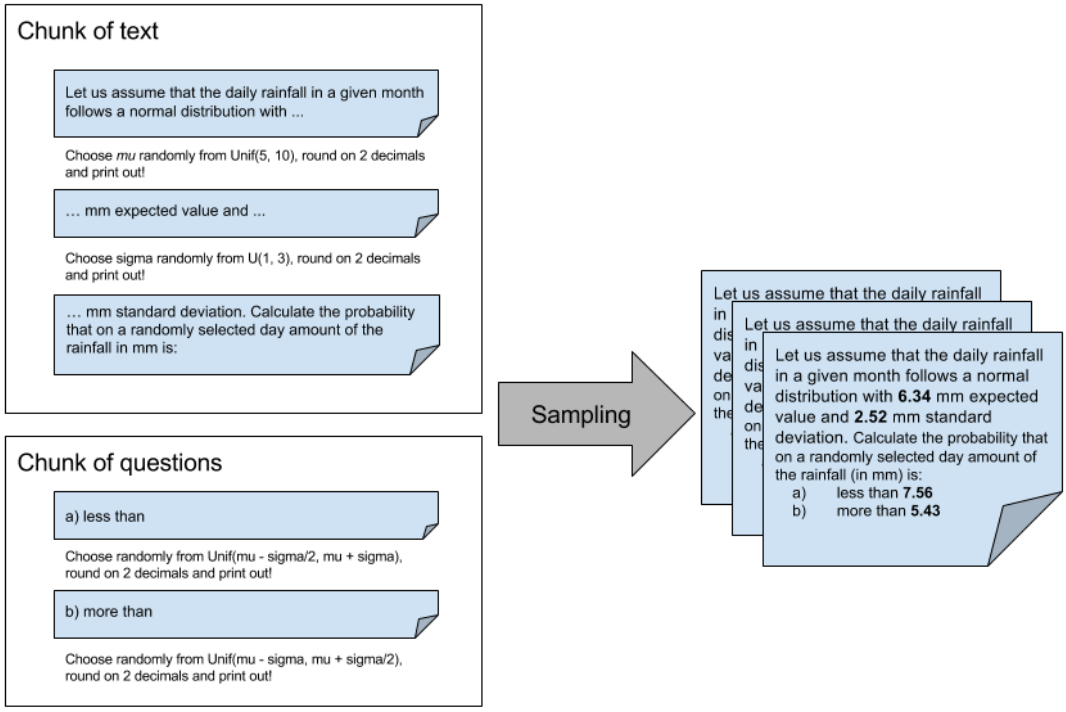
Next we introduce the exams package and how we use it to create midterm questions through an example of tasks related to the normal distribution (introduction to probability is a part of the two semesters). The exams package is working with Rnw files (sweave files) containing a mix of text and R code, allowing the user to execute and embed the results of R computations and graphics within a document. Using built-in functions one can create a variety of outputs, including pdf files and most importantly from our point of view Moodle xml files. As we did not want to use the same text in case of all our students we were looking for a possible way to change the "story"

behind the same exercise. At the moment we store the stories in csv files. As different stories need different random numbers to generate, we also store those together with the stories. To make the procedure more clear we show an example of creating as we call it a meta-exercise (related to the normal distribution). After that we outline a general scheme and workflow.

Creating a meta-exercise

As the normal distribution plays an important role both in probability theory and inferential statistics we chose this topic to introduce how we store the stories and their parameters and create exercises based on these. We want to create exercises where the students have to calculate probabilities of intervals and/or do the inverse cdf calculations based on different stories. For this purpose we have to identify those chunks of text that have to be replaced for each story. One also has to take into account that the solutions should be reasonable, that is we want to avoid result that are too small or too close to one. Also questions themselves should make sense, we want to avoid questions like what is the probability that the IQ of a randomly chosen student is between 101 and 101.5. Avoiding these type of problems needs cautious planning of generating random numbers.

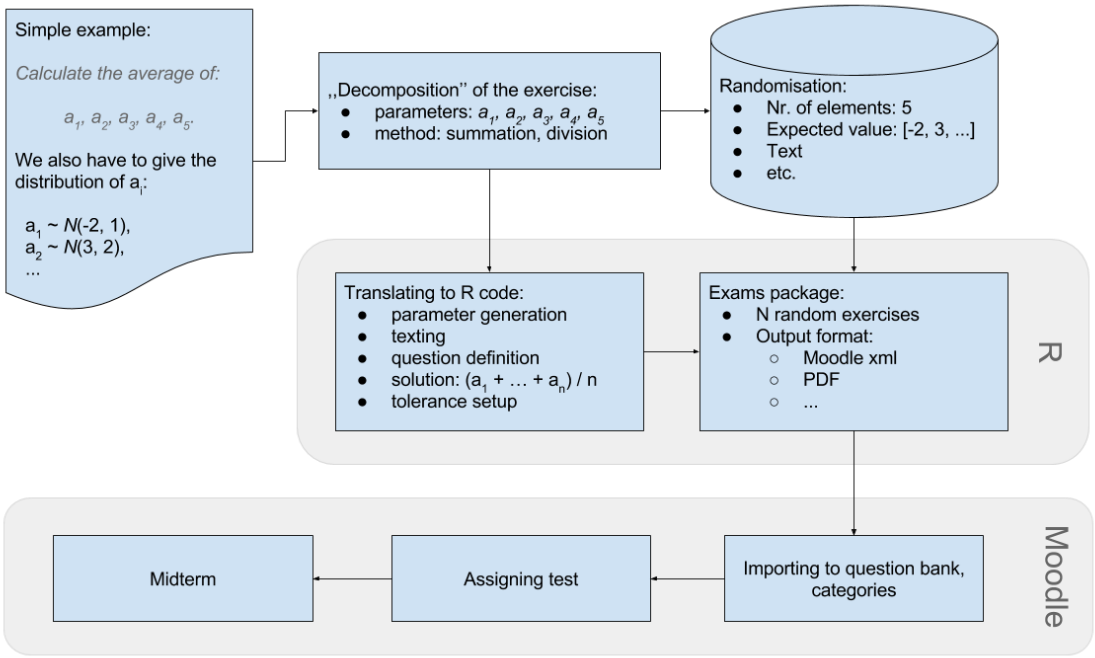
In case of the normal distribution meta-exercise we use 4 text chunks, one is containing the actual story of the exercise (e.g. Let us assume that the daily rainfall in a given month follows a normal distribution with *mumumu* mm expected value and *sigma* mm standard deviation. Calculate the probability, that on a randomly selected day the rainfall is…). Chunks 2-4 contain text in connection with questions for the inverse cdf (e.g. What is the amount of rainfall where the probability is *ppp* that on a given day there is more rain?). In the previous examples the italic values denote parameters that are random, hence all generated problems are going to be different. These parameters depend on the story. Expected value of daily rainfall might vary somewhere let's say between 5-10 mms, but other stories require different expected values. To generate different problems for students we use parameters *mu_from*, *mu_to* and *mu_by* for example for the expected value (e,g. *mu_from* would be 5, *mu_to* 10 and *mu_by* 0.5) In case of employee wages, parameters (expected value and standard deviation) can be in the range of hundreds of thousands (in forint). Another parameter is hence the number of significant digits that is what rounding should we use to make the example nice and meaningful. The last parameters for this meta-exercise is a probability, related to the inverse question, also using the "from, to, by" notation.



Creating a midterm

After creating .Rnw files or meta-exercises for the different topics that we want to include in a midterm we generate as many different tasks as needed and then set up the midterm in Moodle. Setting up the midterm consists of putting all the needed Rnw files in a vector and calling the *exams2moodle* function in R. There are many other supported formats (first of all pdf where there is an option to use custom LaTeX files as templates for the desired outlook of the result) in the package as well. The function creates xml files ready to read them into the Moodle question bank of the given course. Assigning each of the problems to students happens in Moodle. As the solutions (and possibly explanation of the solution) are embedded in the xml file, Moodle is able to grade the exercises automatically and also give a (textual) feedback.

As a summary of creating examples in R and the midterms in Moodle see the following general process:



CONCLUSION

As a result our students have to answer personalized exercises on the midterms which leaves a lot less opportunity for working together during the exam. Results after introducing this system are very similar to previous years. Although the population of student is clearly not the same it makes sense to compare the students as they are similar. Students like the immediate feedback of the system so they don't have to wait days until they get their results. In theory this system gives an opportunity to create unlimited number of practice problems but we did not implement such a system yet. The preparation, setting up the package and clearly creating general enough questions was a big effort but on the other hand now lecturers are able to save a lot of time and develop teaching material etc. instead of grading. We recommend our colleagues who are teaching similar courses to experiment with the exams package and automated creation of exams at least as part of the assessment.

REFERENCES

Gruen, B. & Zeileis A. (2009). Automatic Generation of Exams in R. *Journal of Statistical Software*, *29*(10), 1-14.

Moodle HQ (2017). *The Moodle project.* *https://moodle.org*.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. *https://www.R-project.org/*.

R Tutorials. http://www.r-exams.org/tutorials/

Zeileis, A., Umlauf N. & Leisch, F. (2014). Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond. *Journal of Statistical Software, 58*(1), 1-36.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

USING VISUALIZATION TOOLS IN CLASSES

Péter Kovács, Klára Kazár, Éva Kuruczleki
University of Szeged
kovacs.peter@eco.u-szeged.hu
kazar.klara@eco.u-szeged.hu
kuruczleki.eva@eco.u-szeged.hu

Not only statistical method's usage but also their applications on real data have key importance in statistics classes, which can help to understand social- and economic phenomenon. IT tools, real data and visualization tools can mean the link among the statistical methods (for instance correlation, descriptive statistics, regression models) and application fields (for instance ageing society, poverty, income inequalities). Besides applications, statistical literacy should be also improved in the classes.

The mail goal of the paper is to show statistical course materials which are using real statistical data and newer interactive visualization tools (for instance Gapminder or tools for visualizing social and economic networks). The ProCivicStat project is the framework of our work.

INTRODUCTION

The implication of interesting and lifelike topics is highly important in statistics classes. Students should discover the connection between data and real life, and teachers should make the students part of the data set; in this way, their engagement and activity could be increased (Brown, 2016). This paper offers two examples from real life: first, as economic and social networks are deeply embedded in our everyday life, network analysis could be an interesting topic in statistics classes. Second, understanding a social phenomenon such as poverty has a key importance for students in the field of education, hence examining poverty and income inequalities could also be an important topic in statistics classes.

Besides the above mentioned current topics, innovative solutions should also be included in statistics education. University students are the members of Generation Y, who share common characteristics. They are considered the most technologically literate generation, they cannot imagine life without IT tools or Internet. Generation Y cannot be motivated/encouraged by the traditional way of teaching anymore, but they are eager to use technology during classes instead (Eckleberry-Hunt & Tucciarone 2004, Reilly, 2012, Sox, et al. 2014). Using technology, IT and visualization tools in statistics classes provide an appropriate solution for teaching the members of this generation (Chance, et al. 2007, Ridgeway, 2016). More specifically, the usage of personal response systems (PRS or Electronic Voting Systems) is also recommended; they can motivate students to answer questions and they can also confirm whether students have understood the topic covered in the lesson (Lancaster & Titman, 2014).

The ProCivicStat project funded by the European Commission is a strategic partnership among six universities which create resources (theoretical framework, datasets, visualization tools and teaching materials) for statistics education (Engel 2017). One of the goals of this project is to offer citizens topics in statistics education which are relevant for their future life (e.g. economic networks or financial literacy); and another important part of the project is the application of modern solutions (e.g. visualization tools or evaluation, voting systems) in classes to grab the attention of our students. Based on these, the goal of this paper is to introduce two specific teaching methods developed in the framework of ProCivicStat in the topic of network analysis and in the topic of poverty and income inequalities with the help of modern IT solutions.

APPLICATION OF SOCIAL AND BUSINESS NETWORKS

Network studies is an emerging research topic these days as academics in more and more research areas start to realize how deeply embedded networks are in every field of natural and social sciences and even in economics and how well the tools provided by graph theory can be applied to visualize and examine a set of variables describing any phenomena (Barabási, 2016, Roverato, 2017).

Visualizing networks can help the recipient students -either on lower or higher level of education- to better understand the structure of any social or economic phenomena, and to help them discovering the underlying relations between the constituents of a network. Network analysis and graph theory can be introduced in statistical education through various methods. For instance, one way is through showing an adjacency matrix containing correlation coefficients or interdependencies of different variables of a problem, and translating this adjacency matrix into a graph that better visualizes these coefficients, as graphs are great tools for visualizing interdependencies and relations between variables. This way, our aim of enhancing complex thinking of students to better understand complex data and systems can be achieved.

As part of the ProCivicStat project, introductory and advanced level lesson plans are developed. Introductory level lesson plans teach students the basics of graph theory. The lessons start at the very beginning explaining how networks are embedded in our everyday life, just to mention a few examples, transportation network, the internet or even our brain is a network. A network of metro stations, webpages or neurons. Introducing networks through such examples help students to realize that network theory itself is not as distant as they think, and these examples also help them understand faster the notion of nodes, edges, the weight of edges and other properties of a network. After becoming familiar with the building blocks of networks, students can get an insight into creating a graph. We chose Gephi as the main visualization tool in networks and graphs lesson plans, as it is a software solution that contains all the essential features that we need, but simple enough to be quickly understood by the students. Apart from that, Gephi is an open source program that can be downloaded and used entirely for free by any individuals and institutions. Of course, numerous other network analysis and visualization tools are available, our choice is just a guideline, the task lists and lesson plans can be adjusted to any tool the teacher is familiar with or is willing to use. Tasks aiming at the use of visualization tools teach the students how to create simple graphs by first providing them with an input dataset with just a few details –e.g. nodes, edges between them, weight of the edges and whether the edges are directed or not- and then showing them the main steps on how to create the graph itself. Introductory lectures therefore contain exercises that focus mainly on discovering the details of already created graphs and only include basic tasks on creating graphs.

Advanced level lesson plans are when statistics meet graph theory and network science. One might expect that after completing the introductory level lessons, students can instantly continue with the advanced level. That is partially true for the network analysis part, however even though introductory level can be easily understood by even high school students, the statistical part of these lesson plans have higher prerequisites to be obtained previously. These prerequisites include familiarity with basic statistical terms or basic notions of causality and as well some methods of statistical modeling, such as calculating correlation coefficients or applying regression models. Tasks at advanced level begin with a short overview of the main properties of graphs through a visual illustration of a much more complex graph. One example for this task is a graph illustrating correlation coefficients of S&P 500 stock prices through a period of a few years' time: such graph can visualize which stocks moved together in the examined period and highlight to what degree the volatility of one stock can influence others. Apart from learning how to illustrate more complex systems and how to customize it, students can also learn how to cluster nodes of a system with the help of modularity classes, as a result of which they become able to discover certain sub-systems within a network composed of agents that have stronger relations between each other and weaker with other members of the initial system. By completing the advanced level tasks, students obtain the ability to join statistics and network science to

create graphs that provide a deeper insight to a certain phenomenon and learn advanced customization of graphs.

After assessing network studies lesson plans, one might pose the question: why should network studies be included in civic statistics and what benefits students can infer from learning such high-level network studies? The answer is that it is not a must, but an opportunity to include these materials. The main aim of teaching network studies is to develop and enhance complex thinking of recipients and to teach them how to better understand complex systems. It is true that advanced level lesson plans require such statistical knowledge that some target groups might not have, but we felt that it is important to develop advanced level too, as some recipients might want to implement that knowledge in their own research field. It is certainly useful to learn the introductory levels, as even from that lectures, people can realize how deeply embedded networks are in our life and how many phenomena can be understood as a system of nodes connected by edges, and if one might feel that they want to get a better understanding of networks, that is when advanced level tasks can help, in an easy to understand and entertaining way. Network studies therefore even though do not constitute an essential part of civic statistics, certainly provide an additional visualization method and a way of thinking about social and economic phenomena.

APPLICATION FOR EXAMINING POVERTY AND INCOME INEQUALITIES

The aim of this lesson plan is to understand poverty, income inequality and statistical concepts (e.g. correlation) better with the help of Gapminder. Why is it important to create a lesson plan for those topics? First, the examination of poverty and income inequality is important, due to their impacts on our lives. Poverty and income inequality can lead to a widening gap between the rich and the poor, may reduce people's opportunities to study and can have negative effects on people's health (Keeley 2015). Therefore, those topics are considered as actual questions from the point of view of education. Second, Gapminder is a very spectacular dynamic visualization tool, so it is easy to find delightful and stimulating tasks for students. It is possible to download data and to reach the definitions of the indicators. There are some drawbacks: the number of available indicators in the topic of poverty and income inequality is lower compared to Eurostat Databases or to OECD Stat, and there is no option for uploading data to Gapminder. But still, it is important to use Gapminder, because this visualization tool fulfils the needs of Generations Y students.

The target group of this lesson could be university students; because some previous knowledge is needed. Statistical knowledge about concept of correlation coefficient and the logarithmic-linear transformation are needed; and some context knowledge about poverty and income inequalities is necessary. Concerning the structure, the lesson plan consists of two parts. The first part is an introduction of the usage of Gapminder, and in second part students can explore Gapminder themselves as a kind of independent work.

The first part of the lesson focuses on the usage of Gapminder. The lecturer shows

- built-in visualizations in the topic of poverty and income inequality (e.g. Yes, most billionaires live in the US or Wealth & Health of Nations), which may rise students' interest in the Gapminder and can be an enjoyable part in the lesson. At this point, the results seen in visualization (e.g. Which countries are rich? Where is higher income inequality?) are discussed.
- options of visualization (axes/indicators, play button, speed of visualization, chart/map view, country selection, other options)
- data in Gapminder World, where indicators can be searched by topic and the definitions of indicators are also available
- creation of a not built-in visualization e.g.: examining the relationship between poverty (ratio of people below 2$ a day) and food supply.

The second part of the class includes independent work in teams. Each team receives a question or should raise a question themselves, e.g.: What kind of indicators can be connected to poverty (ratio of people below 2$ a day) in Africa? What kind of indicators can be connected to income inequality index (Gini) in South-America? What kind of indicators can be connected to the income share of the poorest 10% of the population? How did poverty (ratio of people below 2$ a day) change in Europe? The questions have to be answered with the help of Gapminder. Finally, the teams present their results at the end of the lesson, when the other teams and the instructor evaluate and give feedback about the work of each team.

As an assessment we can conclude that this lesson plan meet the needs of Generation Y (technology orientation, teamwork, feedbacks, entertainment, openness for visualization tools). At the same time the lesson provides to the students an introduction into the main concepts of poverty and income inequality and into the usage of Gapminder too. However, it could be important for students being able to create independent analysis by using other data which are not found in Gapminder or by using other statistical methods which are not built in Gapminder. The usage of other statistical database or methods could be a next level in the topic of poverty and income inequality education, but this lesson plan offers a basis for the education of university students.

CONCLUSION

Using visual tools in education certainly has the benefit of maintaining the attention of the audience. Apart from that, in our previously introduced lesson plans and teaching materials visuals also serve the purpose of making understanding complex social and economic phenomena easier.

Network studies might seem to be a distant topic for students, however by highlighting the resemblance of everyday structures, such as the transportation system, to a network, the ability of students to understand complex systems can be enhanced greatly. Visual representation of networks helps students to discover connections between the constituents and to explore sub-systems, i.e. clusters, within a bigger network, acquiring skills that can be implemented in other research fields as well. While introductory materials aim more at developing general knowledge of the students of all ages, advanced level network materials are more suitable for researchers or university students.

University students are the target group of developed materials in the topics of poverty and income inequality visualized with the help of Gapminder. Students can learn about serious issues in an entertaining way with the help of visualization tools of Gapminder, while they also develop their digital competences. Gapminder seems to have some limitations regarding the available data and statistical tools, however lessons with Gapminder provide a strong base for further statistical education of university students.

To conclude, in the above chapters we introduced two types of our developed teaching materials within the framework of the ProCivicStat project. Both topics aim at teaching students about given issues, while at the same time improving their general understanding of the world. Software solutions of both types of lesson plans have their limitations, however they seem to be a good starting base for further education in civic statistics.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Barabási, A. L. (2016). *Network Science.* Northeastern University, Boston.

Brown, M. (2016). Engaging Students in Quantitative Methods: Real Questions, Real Data. In Engel, J. (ed.): *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE)*, Berlin: ISI/IASE.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics. Technology Innovations in Statistics Education, 1(1) Article 2.

Eckleberry-Hunt, J. & Tucciarone, J. (2011). The Challenges and Opportunities of Teaching "Generation Y". *Journal of Graduate Medical Education, 3*(4), 458-461.

Engel, J. (2017). Statistical Literacy for Active Citizenship: A Call for Data Science Education. *Statistical Education Research Journal, 16*(1), 44-49.

Keeley, B. (2015). *Income Inequality: The Gap between Rich and Poor*. Paris: OECD Insights, OECD Publishing.

Lancaster, G. A. & Titman, A. C. (2014). Personal Response Systems as a Learning Aid in an Epidemology Course for Postgraduate Statistics Students. In Makar, K., de Sousa, B. & Gould, R. (eds.): *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, Flagstaff: ISI/IASE.

Reilly, P. (2012). Understanding and Teaching Generation Y, *English Teaching Forum, 50*(1), 2-11.

Ridgway, J (2016). Implications of the Data Revolution for Statistics Education, *International Statistical Review, 84*(3), 528-549.

Roverato, A. (2017). *Graphical Models for Categorical Data*. Cambridge University Press.

Sox, C. B., Kline, S. F. & Crews, T. B. (2014). Identifying best practices, opportunities and barriers in meeting planning for Generation Y, *International Journal of Hospitality Management, 36*(January), 244-254

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

EXPLORING CIVIC STATISTICS WITH CODAP

Joachim Engel
Ludwigsburg University of Education, Ludwigsburg, Germany
engel@ph-ludwigsburg.de

Data are abundant, quantitative information about the state of society and the wider world is around us more than ever. In order to root the public debate based on facts instead of emotions and to promote evidence-based policy decisions we as statistics educators are challenged to promote understanding of statistics about society. This report summarizes a hand-on workshop in which participants explored the potential of the freely available web-based data platform CODAP (http://codap.concord.org) and its usefulness for investigating civic statistics data. Materials of the workshop were developed as part of the EU-funded ProCivicStat Project and are freely available (www.procivicstat.org).

INTRODUCTION

Data on important societal topics are becoming increasingly accessible to the general public and to individual citizens or social action groups, on a huge range of topics such as migration, employment, social (in-)equality, demographic changes, crime, poverty, access to services, energy usage, living conditions, health and nutrition, education, human rights, and many others. In order to ground public debate to be based on facts instead of emotions and to promote evidence-based policy decisions, statistics education needs to embrace two areas widely neglected in secondary and tertiary education: understanding of multivariate phenomena and the thinking with and learning from complex data (Engel 2016, Ridgway 2015). The project ProCivicStat, a strategic partnership of six universities funded through the Erasmus+ program of the European Union, explores a subfield we call Civic Statistics which focuses on understanding quantitative and statistical information about society as provided by the media, statistics offices and other statistics providers. Understanding Civic Statistics is required for participation in democratic societies, but involves data that often are open, official, multivariate in nature, and/or dynamic, that is not normally taught in regular mathematics and statistics education, let alone in politics or social studies.
The challenge is multi-facetted. Data literacy for civic engagement involves, among many other aspects, specific statistical knowledge, ICT skills, "data habits of mind" (Finzer 2013), critical thinking, and much more (Engel 2017). In addition to conceptual blueprints for understanding multivariate phenomena in a data-rich world, the EU project ProCivicStat provides authentic and relevant data sets and develops and tests teaching and learning materials for innovative teaching to a wide range of target groups. The ultimate goal of the project is to strengthen civil society, empowering informed citizens to evidence-based decision-making and civil society engagement. Teaching materials, extensive datasets, conceptual representations of civic statistics are available through the website www.procivicstat.org.

COMMON ONLINE DATA ANALYSIS PLATFORM

ProCivicStats addresses the needs of the civil society and aims to empower students to become informed and active citizens. Its target group are not professional statisticians, computer scientists or empirical researchers. To make data about society openly accessible and data visualisation tools manageable for the general public, appropriate easy-to-use digital tools are essential. Over the last decade many innovative data visualization tools have been developed that permit even novice users to engage with data in ways that are impossible using static displays as in textbooks or print media. The Common Online Data Analysis Platform CODAP (http://codap.concord.org) is a recently developed tool for data exploration and visualization that allows the user to do a whole range of own data analysis steps. CODAP is a freely available, web-based environment for data management and visualization of complex data that supports the many desired transformations and restructurings of data. CODAP is designed as an educational tool for novices. There is no need to

install software. Provided with a link to a CODAP document, an internet connection and a browser suffices to get started.

Most data in traditional textbooks are “flat”, i.e. arranged rectangularly in tables with a manageable number of lines (“the cases”) and some columns (“the variables”). The textbooks usually provide only the columns necessary to solve a given problem. In contrast, authentic data from the net, e.g. about the state of society, are often hierarchical (ordered according to countries, regions, continents, by years, etc.) and usually have a complex multivariate structure of correlated variables that are often non-linearly related - unlike the data of current mathematics education in which students learn statistics. Sometimes, data from the internet must be aggregated or disaggregated before analysis, variables must be re-encoded or transformed to allow appropriate visualizations.

Figure 1 shows a screenshot of the top rows of a data table with 96061 records obtained from the United Nation data base on worldwide refugee movements (http://data.un.org/Data.aspx?d=UNHCR&f=indID%3AType-Ref). The first row refers to the number of refugees (n=1) who left Iraq to go to Afghanistan in 2013, the second row lists Iranians coming to Afghanistan etc. Depending on the question of interest, the data have to be rearranged by country of origin or by country of residence before being graphed. When trends over time are of interest, the same data have to be restructured again in a timely order. In other situations data may have to be transformed or aggregated in order to be useful for an illustrative representation or the desired analysis. Cleaning, transforming, and structuring data are necessary skills, but these skills are not taught in the traditional classroom with its focus on inference based statistics with tidy data.

Download Explore Select columns Select sort order Link to this page

80428 records | Page 1 of 1609 |

Country or territory of asylum or residence	Country or territory of origin	Year	Refugees*	Refugees assisted by UNHCR	Total refugees and people in refugee-like situations**	Total refugees and people in refugee-like situations assisted by UNHCR
Afghanistan	Iraq	2013	1	1	1	1
Afghanistan	Islamic Rep. of Iran	2013	36	36	36	36
Afghanistan	Pakistan	2013	34	34	16,825	16,825
Afghanistan	State of Palestine	2013	1	1	1	1
Albania	Algeria	2013	0	0	0	0
Albania	China	2013	12	12	12	12
Albania	Dem. Rep. of the Congo	2013	5	5	5	5
Albania	Egypt	2013	3	3	3	3
Albania	Iraq	2013	5	5	5	5
Albania	Montenegro	2013	2	2	2	2
Albania	Peru	2013	1	1	1	1
	Serbia (and Kosovo:					

Fig. 1: First rows of a data table retrieved from the UN Data Base on worldwide refugee movement

CODAP supports the required transformation and restructuring of the data. Figure 2 shows a display of the first rows of the refugee data, ordered by year (upper level) aggregated at the level of country of residence. Aggregation and restructuring of the data table were made possible in CODAP by simple drag-and-drop data moves. Now, the highlighted data of 2013 can be graphed displaying the number of refugees for each residence country in 2013.

UNdata_Refugees

Years (13)	Residence (233)		Country_o	UNdata_Refuge
Year	Country... residence	totalRefugees	Country... _of_origin	Refugees
2013	Austria	24058	Afghani...	11906
2012	France	21426	Chad	1
2011	Germany	114227	Congo	122
2010	Greece	2582	Eritrea	101
2009	Hungary	1723	Ethiopia	136
2008	Italy	44322	Iraq	2966
2007	Japan	117	Islamic ...	3188
2006	Jordan	641794	Mexico	1
2005	Netherl...	53823	Nigeria	346
2004	Poland	461	Somalia	2278
2003	Portugal	133	Sudan	265

Fig. 2: First rows refugee data restructured with CODAP to explore the yearly distribution of refugees across various host countries

HANDS ON ACTIVITIES

In the following we present and discuss some of the activities participants of the workshop performed. Participants received worksheets (available through the ProCivicStat website www.procivicstat.org) with background information about the context, a short technical description of the data including its source and several closed and open questions to guide the exploration to be done individually or in pairs. A link provided on the worksheet leads to a CODAP document containing the required data.

1. Some so rich others so poor – Income distribution in Europe


Promoting Civic Engagement via Exploration of Evidence: Challenges for Statistics Education

ProCivicStat © - Students' Worksheet, S.101

A few so rich, others so poor - income inequalities in Europe

Joachim Engel — Achim Schiller

Ludwigsburg University of Education

Ludwigsburg, Germany

What is inequality?

Inequality is typically viewed as different people or households having different degrees of living standards. Thus, inequality is concerned with the relative position of different individuals (or households) within a distribution.

Why does it matter?

On July 14, 1789, the French stormed the Bastille, a medieval fortress-prison in Paris in one of the key moments of the French Revolution. The average people were fed up with the ~~don. ccc Bégum~~ — aka the monarchy — and they were protesting the vast inequality between themselves and the upper echelon.

~~Geo-Gerard~~ 268 years: inequality is still a major issue.

Why are there in some countries large discrepancies between the rich and the poor while in other countries the income distribution is more equal? What have countries in common that have a large discrepancy between the rich and the poor? Figure 3 shows the head of the worksheet.

The data from EuroStat include the following variables: Country, Year, Population size, Mean Income, Median Income, the Gini-Cofficient for the income data as well decils, quantils and the income share of the 5% lowest and 5% highest earners per country and year.

Fig 3: Head of the worksheet on income inequalities in Europe

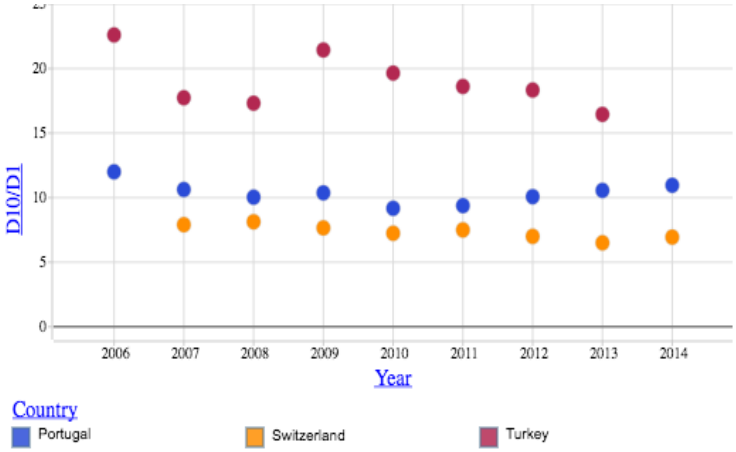


Fig 4: Dotplot of the ratio 10th decile and 1st decile between 2006 and 2014 for Portugal, Switzerland and

One question on the worksheet asked:

To investigate the trend of a newly defined variable D10/D1, defined as ratio of the 10th to the 1st decile, over the years for Portugal, Switzerland and Turkey, what do these countries have in common? How do they differ with respect to D10/D1?

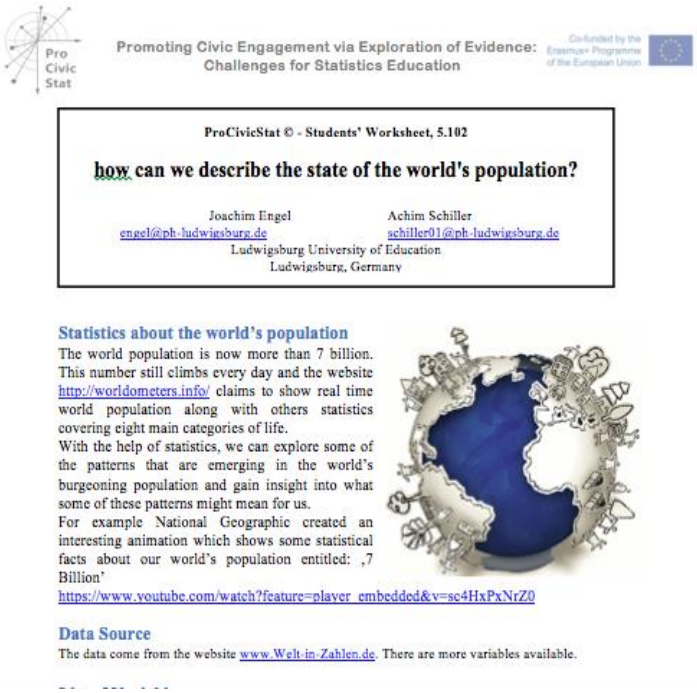
To address this question required the following steps: (1) define a new variable D10/D1 for the quotient of the 10th and 1st decile; (2) plot D10/D1 versus year; (3) select the three countries in question and hide all the unselected cases, resulting in the visualisation of Figure 4.

A possible answer to above questions could be:

The D10/D1 ratio in Switzerland has a minimum of 6.5; the maximum is 8.13. In Portugal the minimum is 9.17 and the maximum is 10.96. Turkey has as minimum 16.45 and as Maximum 21.44. In Switzerland, no real trend can be seen over the years 2007 to 2014, and the rise and fall in value is changing irregularly. In Portugal, the ratio has risen again since 2010, which means that the uneven wealth distribution is increasing. In Turkey, a downward trend can be seen after the financial crisis in 2009, which means that inequality is decreasing. However, there are still clear differences between these countries.

2. How can we describe the state of the world's population

The world population is now about 7.6 billion. With the help of statistics, we can explore some of the patterns that are emerging in the world's burgeoning population and gain insight into what some of these patterns might mean for us (see Figure 5).


Promoting Civic Engagement via Exploration of Evidence: Challenges for Statistics Education

ProCivicStat © - Students' Worksheet, 5.102

how can we describe the state of the world's population?

Joachim Engel engel@ph-ludwigsburg.de
Achim Schiller schiller01@ph-ludwigsburg.de
Ludwigsburg University of Education
Ludwigsburg, Germany

Statistics about the world's population

The world population is now more than 7 billion. This number still climbs every day and the website http://worldometers.info/ claims to show real time world population along with others statistics covering eight main categories of life.
With the help of statistics, we can explore some of the patterns that are emerging in the world's burgeoning population and gain insight into what some of these patterns might mean for us.
For example National Geographic created an interesting animation which shows some statistical facts about our world's population entitled: ‚7 Billion'
https://www.youtube.com/watch?feature=player_embedded&v=sc4HxPxNrZ0

Data Source

The data come from the website www.Welt-in-Zahlen.de. There are more variables available.

Fig. 5: Head of worksheet on world's population

The data are from the website www.Welt-in-Zahlen.de.
17 different variables, representing average values for 222 countries of this planet such as migration, nutrition (in kilocalories per person), fertility, GNP per person, internet access, number of physicians per 1000 and many more.

To investigate the distribution of nutritional intake per person across continents, CODAP allows to draw boxplots for each continent, thus revealing that people in some countries of Africa and Asia live on very low calories. But observe the large spread (Figure 6). Figure 7 shows scatterplots plus least square line, separate for each continent, for GNP per Person versus number of physicians per 1000 people

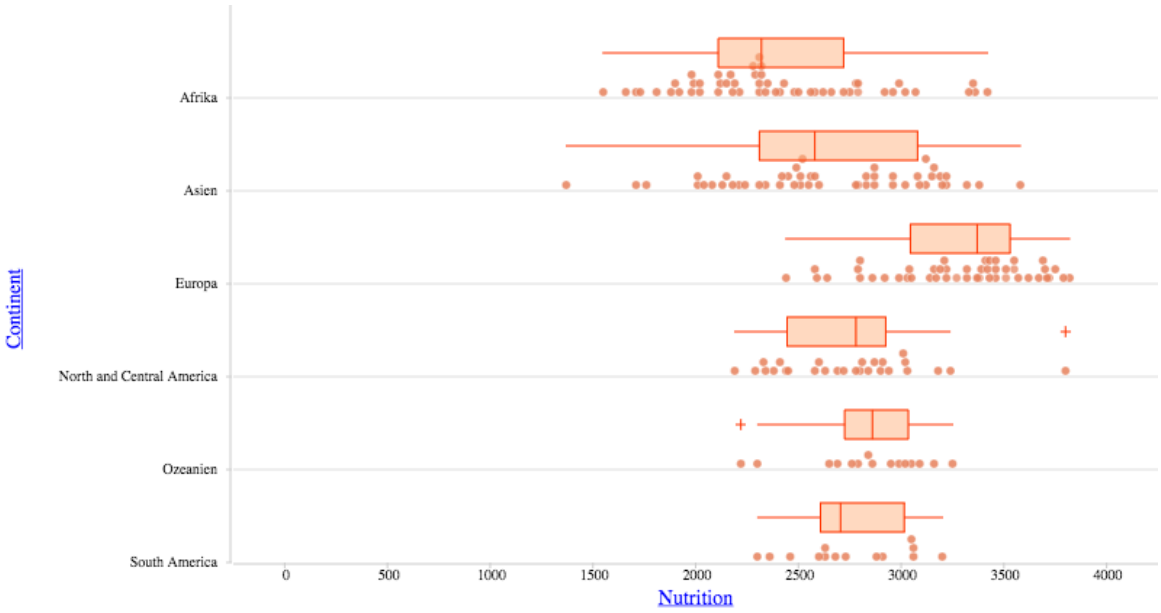
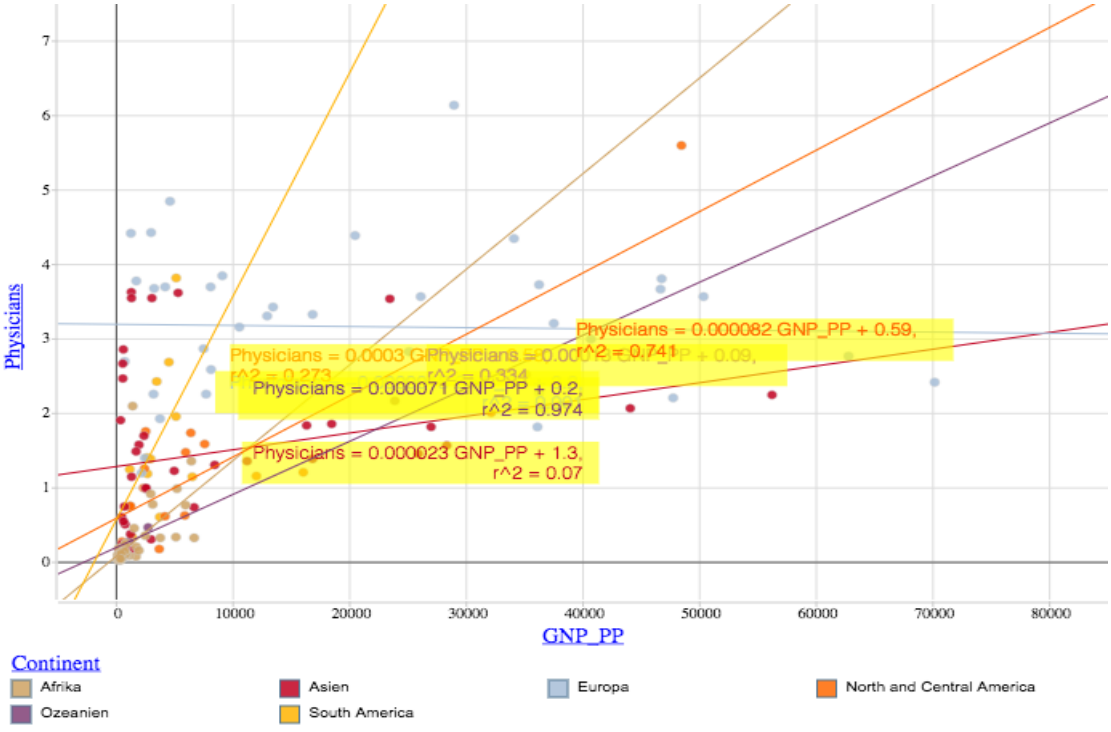


Fig 6: Boxplots representing mean human calorie intake, separated by continent. Each dot represents a different country.



-

Fig. 7: Scatterplot of “number of physicians per 1000 inhabitants” and “GNP per person” in 222 countries with least squares lines. Every dot represents a country, continents are coded by dot color.

3. Are referees in European football racially biased?

Fig. 8: Head of worksheet on investigating possible referee bias

Are players with dark skin tone more likely than light skinned players to receive red cards from referees in European football (see Figure 8)? The data set comprises information on 1419 football players in four professional European football leagues with 19 variables such as number of red, yellow-red and yellow cards received during a player's career, position played, height, weight and a rating of skin color (1=very light, 5 =very dark). While it is easy to create boxplots of the variable RedCard or the newly defined variable RedCardsRate (= RedCards per Game), the tricks question is the search for possible confounding variables, i.e. third explanatory variables that may account for an observed relationship between two variables. Therefore, before looking at boxplots separate for of, say, RedCardsRate at each level of the variable Skintone, one may ask: Are red cards and skin color equally distributed across the four countries (England, France, Germany and Spain)? What about the distribution of the position across players of different skin color? Are players of color more often represented in some positions than in others?

Figure 9 displays the distribution of Position versus Skintone and Country League versus Skintone. The figures reveal that dark skinned players play more often as attackers or midfielders than as defenders or goalkeepers. Also, the French and British League has a higher percentage of coloured players than Germany or France.

These observations matter to our guiding question for potential racial bias of referees. For red cards tend to be given more often to defenders than to midfielders or attackers. Also, referees in England, Spain and France tend to brandish red cards more often than in Germany (Figure 10). All of this calls for caution in jumping to quick conclusions because Country League and Position may well be confounding variables that may have a strong impact on referees red card giving, thus masking the skin tone variable.

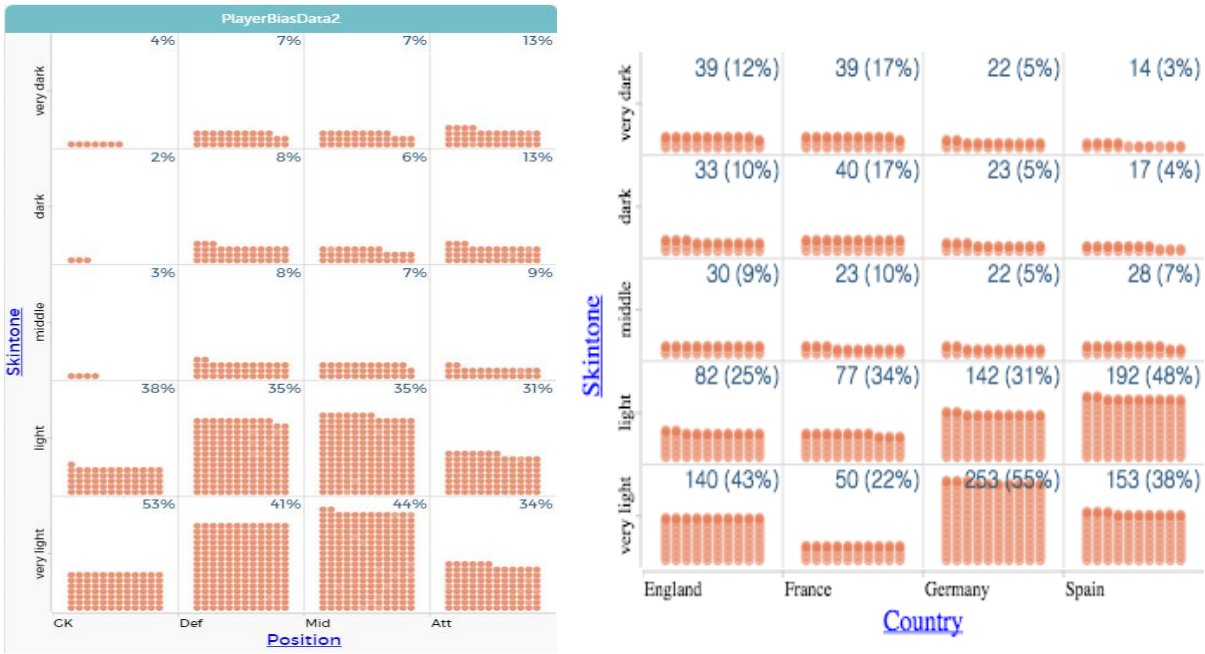


Fig. 9: Bivariate distribution of the nominal variables Position and Skintone (left) and Country League and Skintone (right)

In summary, we conclude by means of the diagrams that colored players are increasingly attackers. The position of midfielder and defender have nearly the same proportion of all skin tones while coloured goalkeeper are underrepresented.
Likewise, the variable RedCardRate is not evenly distributed across the four.

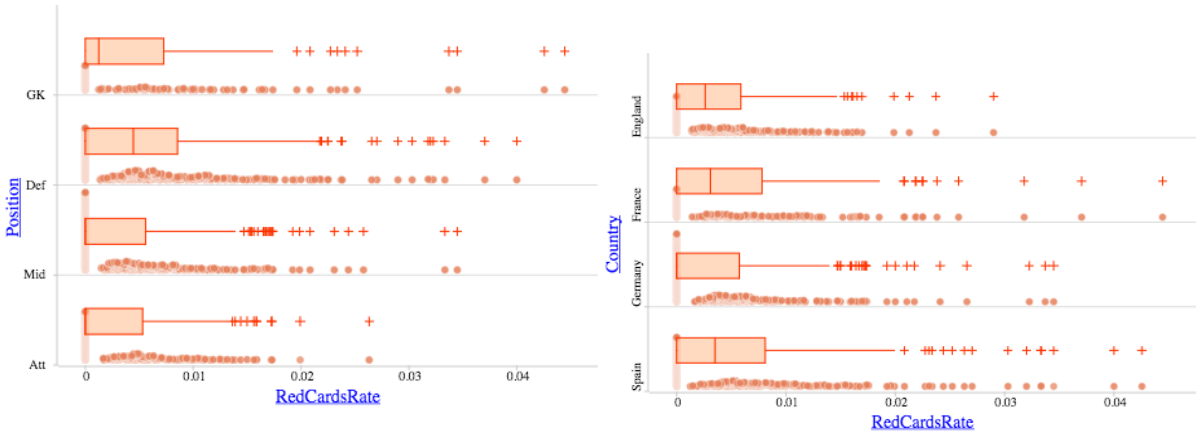


Fig. 10: Distribution of RedCardRate across the position of a player (left(and across country leagues (right)

CONCLUSION

CODAP proved to be a tool, most users manage to work with even without much instructions. Recurring topics for the data analysis based on CODAP were: (1) Comparing distributions, (2) Aggregating data, (3) Restructuring data, (4) Investigating and comparing subgroups, (5) Search for explanatory third variables, and (6) Model functional relationships.

Introducing new digital tools for data analysis to support learning with and from data raises important questions. What is the relative value of curated versus self-selected data sets? In ProCivicStat we provide curated data sets on a whole range of topics related to Civic Statistics. We initiate the analysis with a few closed questions, but also ask for more open exploration which may require from the learner to immerse deeper into the context. On a technical level, new questions may imply the need to define new variables and to restructure the data.
An alternative could be to let learners choose their own topic and have them search for appropriate data sets. It may be more powerful and motivating to learn with the data you have chosen yourself. We tried this path in a seminar with students at our university. However, it is hard to find suitable data you are looking for; if you found them, it is often not at all trivial to import them into the software you are using (even though CODAP allows under some circumstances for web scraping) and the freely chosen data set may not teach what the instructor wants to teach.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Engel, J. (2017). Statistical Literacy for active citizenship: a call for data science education. *Statistics Education Research Journal* 16(1), 44-49.

Engel, J. (Ed.) (2016). Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE. http://iase-web.org/Conference_proceedings.php.

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, 7(2). www.escholarship.org/uc/uclastat_cts_tise.

Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review, 84*(3) 528–549.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

DIDACTICAL ISSUES OF DATA MODELING WITH INTERACTIVE CHARTS

Djordje M. Kadijevich
Institute for Educational Research, Belgrade, Serbia
djkadijevic@ipi.ac.rs

This workshop, based on both theoretical considerations and practical work, deals with simple data modeling using pivot charts and dashboards, and the didactical issues of this kind of modeling. The workshop is primarily based upon the author's IASE 2016 Roundtable paper, as well as his recent experience in teacher professional development concerning this topic.

TOPIC RELEVANCE AND ITS EDUCATIONAL VALUE

The science of data has in some ways been present in computing for more than a half a century (please recall that the term *datalogy* was introduced by the Danish scientist Peter Naur (1928–2016) almost sixty years ago; see https://en.wikipedia.org/wiki/Data_science), and has been a hot topic in recent years. Having in mind widespread demands for modeling huge amounts of data for business, scientific or other purposes (to obtain from them potentially useful information that may lead to new knowledge), it is not surprising that expertise regarding the study of data (i.e. data scientist post) has been increasingly in demand. Although statisticians usually say that the study of data is in fact the statistical analysis of data (i.e. data science "=" data analysis), this is not strictly true. This is because the science of data requires not only solid knowledge and skills in computing (e.g. programming and databases) and mathematics (the application of a variety of mathematical and statistical models), but also, among other things, a high degree of creative thinking and communication skills. It is thus not surprising that the job of data scientist is highly paid (frequently over $100,000 annually). As some kind of data modeling is likely to be present in the future professional work of most students, there is a growing call to include basic data modeling at earlier educational levels (e.g., Engel, 2017). Please recall that *Data* has been a content area in a series of international assessments in school mathematics and science known under the acronym TIMSS (http://timssandpirls.bc.edu/). TIMSS tasks in this domain usually deal with tabular and graphical representations of data, drawing conclusions from such representations, which are indeed important aspects of elementary data modeling.

In the last ten years, the educational values of data modeling by using interactive charts (with the application of elementary mathematical or statistical models) have been the focus of study among a number of researchers engaged in improving statistics education, particularly those working at the Smart Centre, Durham University, UK (https://www.dur.ac.uk/smart.centre/). The approach of these British researchers was summarized in a recent paper titled *Visualise then Conceptualise* (Nicholson, Ridgway, & McCusker, 2011), for example. Their research evidences that this kind of modeling, even when arranged in an informal way, supports the understanding of important statistical notions, such as size of effect, trend, interaction, and confounding variable. Bearing in mind that interactive objects of the dashboard type consists of two or more pivot charts, such interactive objects enable better data modeling than simply relying on single pivot charts. This is, among other things, because the data modeler can compare the outcomes on two or more charts.

Researchers in statistics education are just beginning to study how to use dashboards successfully in teaching. Bearing in mind the educational value of this kind of modeling, students should be supported in the creation and use of sets of pivot charts whose structural complexity increases gradually. In doing so, a number of challenges should be addressed by the teacher. These challenges, relating to 1) data to analyze, 2) dashboards to create, and 3) modeling to implement, have not been examined so far (cf. Kadijevich, 2016).

WORKSHOP DETAILS

Main prerequisites

- basic knowledge of mathematics and statistics (e.g., absolute and relative frequency, percentage, chart types, the average value);

- basic knowledge related to spreadsheets and relational databases (field i.e. attribute, record-entity, making queries);
- basics of visual programming using the drag & drop approach.

Practice at the workshop

Each workshop participant works on his/her personal computer with a recent version of Microsoft Excel installed. The participant knows how to use this program (at least on an introductory level). In addition (for optimal practice at the workshop), this personal computer has access to the Internet, and the participant has already registered at https://www.zoho.com/reports/dashboard.html to be able to use the ZOHO environment for creating dashboards.

Main outcomes (benefits)

The participants improve their knowledge and skills in the following domains:

- the preparation of data for the visualization of relations in those data;
- the visualization of those relations by using one or more charts;
- understanding important statistical concepts, such as size of effect, trend, interaction, and confounding variable;
- applying a methodical approach to the practice of data modeling in mathematics, informatics, and statistics education.

EXAMPLES OF DATA MODELING

- *Pivot chart*

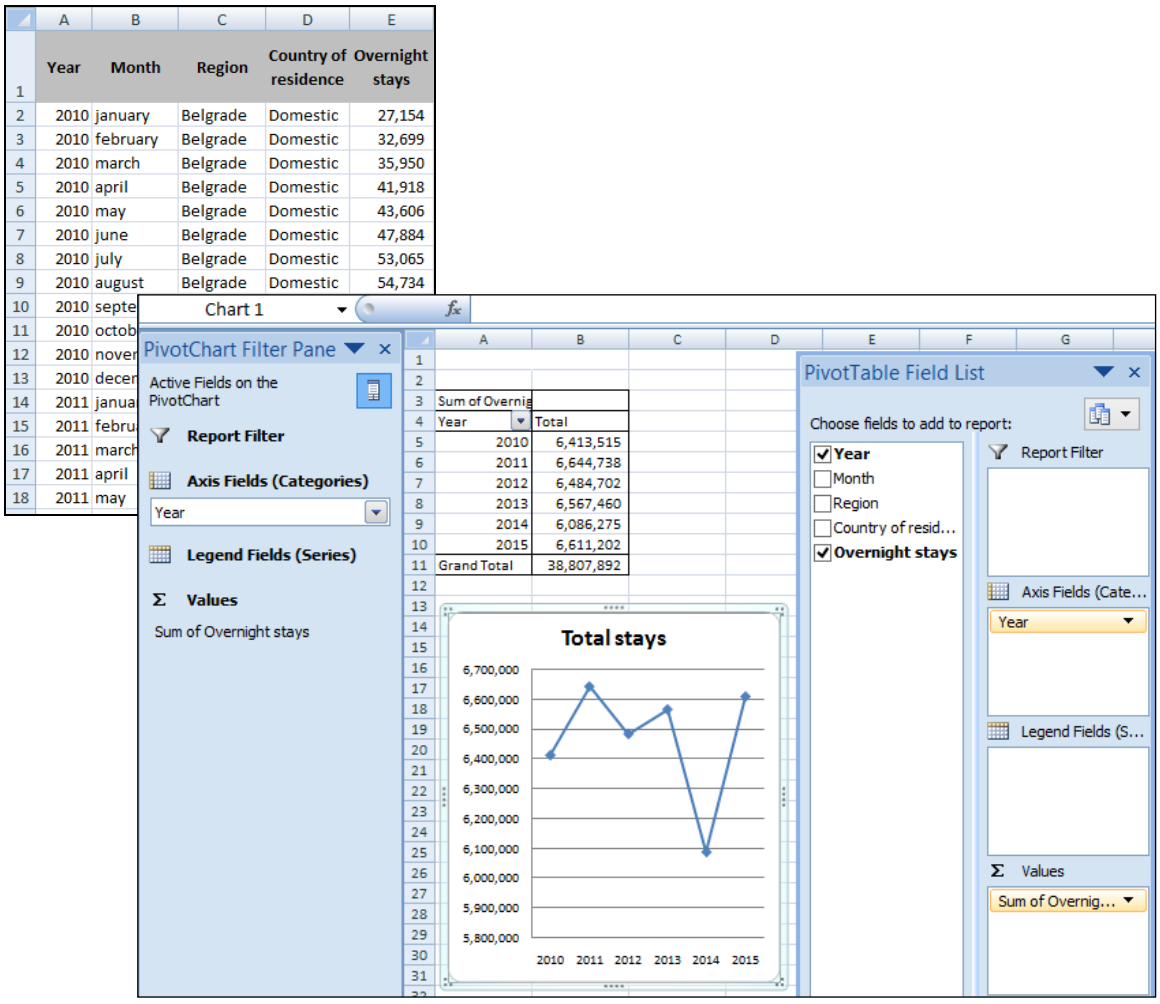


Figure 1. Initial data, pivot table, and pivot chart (source: www.mi.sanu.ac.rs/~djkadij/Seminar17.xls)

- *Dashboard*

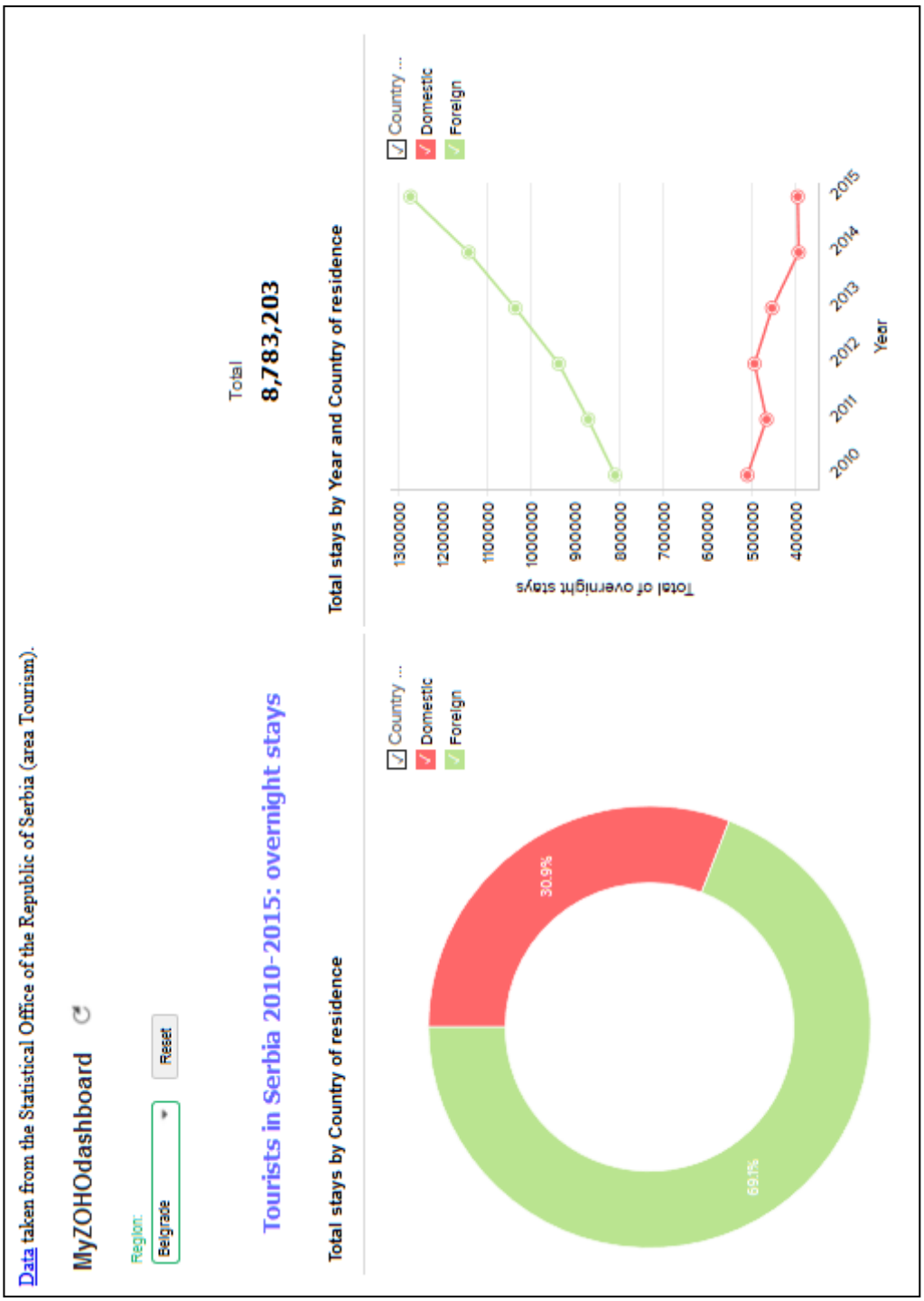


Figure 2. Screenshot of a dashboard available at www.mi.sanu.ac.rs/~djkadij/Dashboard.htm

PIVOT CHART AFFORDANCES

Pivot tables and pivot charts are simple tools of business intelligence, which provide summary reports in the form of interactive tables and graphs. These tools allow some summary parameters (e.g., the sum or the average of the values considered) to be shown in relation to different levels of details. The user selects these details with respect to his/her requirements, and, if necessary, further collapses or breaks them down (the so-called roll-up or drill-down views, respectively).

Consider, for example, the following chart regarding the sale of some items.

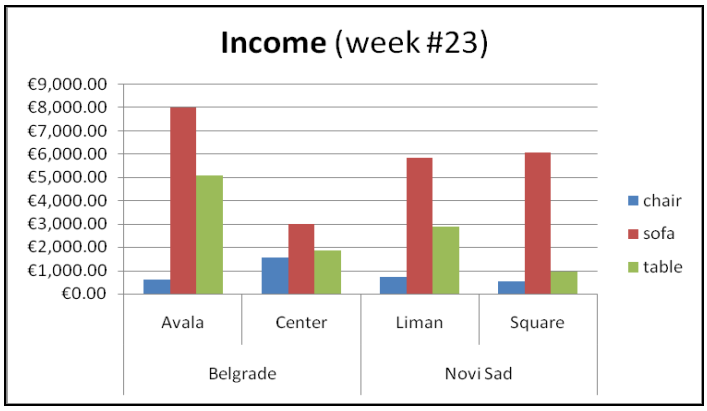


Figure 3. A sales chart

Is easy to realize that this chart is related to the sales of a number of items, that these items are sold in different stores, and that these stores are located in two cities. Therefore, in addition to the finding of the total income (Level 0), the use of the pivot approach can help its user to analyze the following:

- the total income by city (Level 1 – the first level of details with respect to the categories of one independent variable; also called attribute or indicator), to find out in which city the total income was the highest;
- the total income by city and store (Level 2), to determine in which store and in what city this income was the lowest; and
- the total income by city, store, and item (Level 3), to find out in which store and in what city the total income for a particular item was the lowest.

WORKSHOP TASKS

Tasks for self-preparation

Participant's self-preparation for the workshop (before it!) includes the following two tasks:

- By using a file with some sales data (e.g., http://www.mi.sanu.ac.rs/~djkadij/LaudonASE1.xls), generate graphs that enable a business analysis like the one given above. The main question is how to model the realization of this analysis didactically in order to attain successful students' work on at least two levels of complexity (with simple learning requirements *vs.* some more complex ones).
- By using the visualization titled *Cardiovascular Disease Risk Factors* (available at https://www.dur.ac.uk/smart.centre/freeware/), summarize several findings regarding some, to you important relations in the data analyzed. Through play with this visualization, try to illustrate some of the important statistical notions mentioned above (size of effect, trend, interaction, and confounding variable). To this end, you may use the content of the Appendix.

Main task at the workshop

It is planned that each workshop participant, in accordance with a theme chosen by himself/herself, will carry out data modeling by using interactive charts. (If the *ZOHO* environment is used, helpful instructions for data modeling with dashboards are given in a video available at https://www.youtube.com/watch?feature=player_embedded&v=aPLg4dp-f28). This data modeling should be based on a real data set provided by a relevant agency or institution, such as *Eurostat*, or the *World Bank*. Particular attention should be paid to the didactical issues of this modeling at a specific educational level (e.g. in a grade 8 mathematics class), to enable successful students' work on at least two levels of complexity (with simple learning requirements *vs.* more complex ones). This work is based upon asking questions and examining answers (results) in context, which are usually very challenging parts of data analysis (Gal & Trostianitser, 2016; Gould et al, 2016).

Acknowledgement. This contribution resulted from the author's work on project 47008 funded by the Serbian Ministry of Education, Science, and Technological Development. The author dedicates the contribution to his son, Aleksandar.

REFERENCES

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16* (1), 44–49.

Gal, I., & Trostianitser, A. (2016). Understanding basic demographic trends: Connecting table reading, task design, and context. In Engel, J. (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany*. The Haag, the Netherlands: ISI/IASE. *http://iase-web.org/Conference_Proceedings.php*.

Gould, R. et al. (2016). Teaching data science to secondary students: The mobilize introduction to data science curriculum. In Engel, J. (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany*. The Haag, the Netherlands: ISI/IASE. *http://iase-web.org/Conference_Proceedings.php*.

Kadijevich, Dj. M. (2016). Data modeling with dashboards: Opportunities and challenges. In Engel, J. (Ed.), Promoting understanding of statistics about society. *Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany*. The Haag, the Netherlands: ISI/IASE. *http://iase-web.org/Conference_Proceedings.php*.

Nicholson, J., Ridgway, J., & McCusker, S. (2011). Visualise then conceptualise. *Social Science Teacher*, *40*(3), 8–12.

APPENDIX – SMALL STATISTICAL REMINDER

The following paragraphs contain informal explanations of four key statistical notions.

- The *size of effect* describes which variable is more related to the target variable. For example, graphs may show that while smoking increases the risk of an infarct about two times, insufficient physical activity increases that risk even 3–4 times.
- *Trend* tells us in what way the values of one variables change over time. A chart may signal that these values increase or decrease over time, and if a trend is present, it may be linear or not.
- The *interaction* of independent variables exists when the effect of one of them on the dependent variable differs for different values of other variable. Let us suppose that one study found that contrary to Faculty of Mathematics, there were gender differences in the average duration of graduate study at Faculty of Medicine. In that case, there was an interaction between variables gender and faculty. Interaction would be easily recognized in graphs because the distance between the average values of the dependent variable for the values of one independent variable (e.g., gender) changes by changing the value of other independent variable (e.g., faculty).
- *Confounding variable* is a variable that could affect the relationship between variables in question. For example, students in science departments might differ from students in social science departments with respect to their average weight, but this difference might be caused by the fact that these two groups of students differed in their average age. If this applies, variable age is a confounding variable, and its consideration may reduce or annul these weight differences.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

GAMESTAT-I., AVAGY A FEJEKBE SURRANÓ TUDÁS

Dr. Kolnhofer-Derecskei Anita, Dr. Nagy Viktor, Kovácsné Bukucs Erzsébet
Óbudai Egyetem Keleti Károly Gazdasági Kar (Budapest)
Derecskei.Anita@kgk.uni-obuda.hu
Nagy.Viktor@kgk.uni-obuda.hu
Bukucs.Erzsebet@kgk.uni-obuda.hu

Az Óbudai Egyetem Keleti Károly Gazdasági Karán a névadó halálának 125. évfordulójára időzítve egy új módszert teszteltünk a 2017-es év tavaszi félévében. Az alapképzéseinkben (BSC) résztvevő hallgatók részéről a tárgy elsajátítása során fellépő nehézségeket oktatói oldalról kihívásként értékelve, a célcsoportra szabott új módszerek alkalmazására törekedtünk. A Statisztika I. tárgy anyagának (főleg leíró statisztika) elsajátításához a hagyományos eszköztár alkalmazása mellett meghatározott játékszabályok betartásával hallgatóink konkrét példákon keresztül tesztelhetik tudásukat. A workshopunk célja volt, hogy bemutassuk ezt az új technikát, mely során a „száraz" statisztika online formában, játékos keretek között (gamification) lopózik be a hallgatói fejekbe. Több forduló során gyűjtött pontjaikat a játékosok saját döntésüknek megfelelően számíttathatták be a tárgy hagyományos számonkéréseinek eredményeibe. Mivel a program a tavaszi szemeszterben került megvalósításra, az őszi workshop időpontjára már annak eredményességéről és a tapasztalatokról is részletesen beszámolhattunk. Reméljük, hogy ezzel egy új hagyomány alapjait teremtjük meg, modernizálva ezen kimondottan hasznos tárgy oktatását és a hallgatóinkban tudatosítva a tárgy valódi gyakorlati jelentőségét.

GAMIFIKÁCIÓ AZ OKTATÁSBAN

A játékosítás (gamifikáció) megjelenése az oktatás területén nem újkeletű, ameddig a marketing és a humánmenedzsment szakemberei rácsodálkoztak a gamifikáció lehetőségeire, addig az oktatók számára mindez önmagában még nem jelent újdonságot. Az oktatás, mint tevékenység megkíván egy bizonyos fokú kreativitást, mely motiválja a hallgatókat és egyben kihívást jelent az oktatónak, megvédve a berögzött, idővel megunt mechanikus sémáktól. Magyarországon egyre több pedagógus alkalmazza és növekvő számú kutató tanulmányozza a gamifikációt. Ugyanakkor, ezen cikkünknek nem célja a hazai és nemzetközi szakirodalmak felsorakoztatása, sokkal inkább egy gyakorlati megvalósítás bemutatására törekszünk. Véleményünk szerint a gamifikáció közel sem azonos videójátékok alkalmazásával az oktatásban, sokkal inkább jelent egy szemléletmódot, a játék elemeinek és jellemzőinek beépítését az oktatásba. Ezek az elemek a következők lehetnek: (1) természetes tevékenység, egyfajta tanulási forma, de megjelenésében játékhoz hasonló, látványra hangolt (2) a játékosok önkéntesen kapcsolódnak be (3) előre rögzített szabályrendszerhez kötött, melyet a játékosok ismernek és elfogadnak (4) történetbe ágyazott (5) különböző szinteken, ehhez igazított célok mentén zajlik (6) biztosított benne a visszajelzés, és a jutalom. Így az általunk elfogadott munkadefiníció szerint, a gamifikáció szabad cselekvés (önkéntes), de kötelező szabályok szerint zajlik egy meghatározott cél érdekében. Sikert ígér (kihívást hordoz magában), önmagáért való, de meghatározott keretei vannak. Játékosítás (gamification) alatt azt a folyamatot értjük, melynek során a játékelemek, játékmechanika és játékos gondolkodás felhasználásával egy alapjában véve nem játékos közeget olyan vonzóvá, szórakoztatóvá és motiválóvá szerkesztünk, mint amilyen a játékok világa.

Használata nem kötelezően kötött az IKT eszközök használatához, szemléletünkben sokkal inkább jelent egy új irányú pedagógiai módszertant. Ugyanakkor az egyetemi képzés nappali tagozatain már megjelentek a Z generáció ifjú tagjai is, a zömében Y generációs Millenárisok között. Márpedig mindkét nemzedék elszakíthatatlan ezen eszközeitől, kihívások elé állítva az oktatókat. Összességében egyetértünk Damsa Andrei gondolataival, mely szerint „ A módszer tulajdonképpen egy átkeretezést fed, mellyel olyan elemeket iktatunk be egy adott programba, mint például a pontrendszer és a szintlépési lehetőség, az önkifejezést elősegítő avatar megszerkesztése vagy akár a versenyzési lehetőség. … Fontos megemlíteni, hogy a játékosítás során nem konkrét játékokat használunk fel, hanem a játékokban fellelhető motivációs elemeket alkalmazzuk." (Damsa 2014) Ehhez mérten terveztük meg első gamifikációs alapokon nyugvó Gamestat 1. játékunkat.

GAMESTAT1.

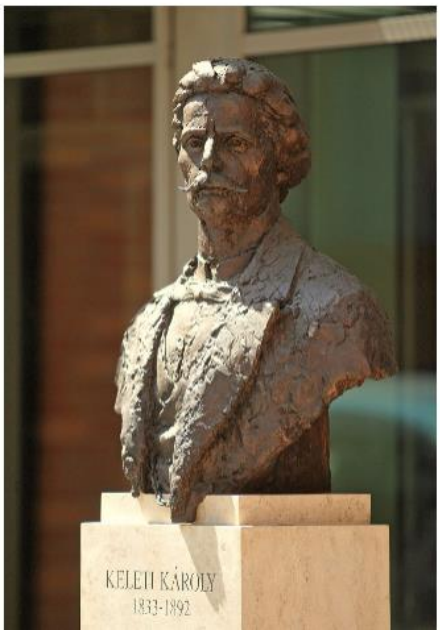
Az Óbudai Egyetem Keleti Károly Gazdasági Karán a statisztika tantárgy oktatásáért felelős összeszokott csapatunk, folyamatosan törekedik az állandó megújulásra, megőrizve a tárgy alapozó jellegét, hiszen a későbbiekben az ezen oktatáson szerzett tudás mintegy kiszolgálja a szaktantárgyakat. Ugyanakkor hasonló problémákkal küzdünk, mint a partneregyetemek oktatói, melyek közül legnagyobb gondot a tömegoktatás okoz. A tantárgy előadásokból valamint gyakorlati szemináriumokból épül fel, de a magas óraszám ellenére is az egyik „mumus" tárgy, igen magas bukási aránnyal. A hallgatók kérése, hogy hogyan szerezhetnének plusz pontokat, valamint az a cél, hogy „hallgatóbarátabbá" tegyük ezt a tantárgyat ösztönzött bennünket arra, hogy 2017. januárjában életre hívjuk a GameStat 1. játékunkat.

Statisztika oktatás az ÓE KGK-n

A játék a Statisztika I. tantárgyhoz kötötten indult el, BSC és FOSZK szinten nappali, valamint levelező tagozaton. Ezen tantárgy hangsúlyosan a leíró statisztikai alapokra fókuszál, viszonylag kevés matematikai, valószínűségszámítási ismeret előzetes meglétét igényli. Ennek ellenére a tárgy az elmúlt évek tapasztalatai alapján az egyik legnehezebben teljesíthető volt a hallgatók számára. A tárgyfelelősök által készített önálló, gyakorlatorientált jegyzetek alkalmazása ugyan javított a helyzeten, de szükségesnek láttuk olyan módszerek alkalmazását, amik tovább javítják a tantárgyból elért eredményeket, és talán fel is keltik az érdeklődést a statisztika iránt.

Maga a játék

A játék elindításához első körben meghatároztuk a hallgatói célcsoportot, majd a pontozási rendszert, felosztottuk a tananyagot ehhez hangoltan készült idő és feladatbeosztás alapján haladtunk. Minden adott részért a kihívást kiíró oktató felelt, a munkánkat hallgatói demonstrátor segítette, aki kezelte a beérkező válaszokat. A hallgatók számára mindezeket a játékszabályban rögzítettük (1. ábra). A játékfelületét mindenképpen online elérhetőséggel szerettük volna megoldani, teret adva a gyors visszacsatolásnak és a könnyebb kezelhetőségnek, nem beszélve a beküldött anyagok archiválásáról, közös eléréséről. Hosszas próbálgatások után a feladatkiírások a tárgy Moodle alapú e-learning oldalán jelentek meg, a beküldés pedig emailben történt, melyhez a Gmail levelező rendszeren hoztunk létre egy levelezőfiókot.



A GameSTAT1 verseny JÁTÉKSZABÁLYA

1. A játékban részt vehetnek: a 2016/17/2 félévben a Statisztika I. (GVMST12GND, GVMST12GNC, GVMST12KNC, GVMST12KND vagy GVMST12KLC, GVMST12KLD, GVMST1GMLD, GVMST1GMLC vagy GVMST1ATNO) tárgyat felvett hallgatók. A nappalis, levelezős és foszk hallgatók külön kerülnek értékelésre.
2. A játék 6 fordulóból áll, fordulónként nappali tagozaton legfeljebb 2-2, levelezőn 1-1, foszk képzéseken 2-2 pont szerezhető a feladat típusától függően a nappali tagozaton 20, levelezőn 5, foszk képzésen 9 leggyorsabb vagy legjobb megoldást benyújtóknak – erről a feladatot kiadó oktató saját hatáskörében, illetve a feladat típusától függően dönt. Az egyes feladatok a Moodle rendszerben kerülnek meghirdetésre, a megoldásokat a gamestat1.2017@gmail.com címre várjuk.
3. A szerzett pontok (SZP) beszámítása úgy történik, hogy a soron következő zárthelyi dolgozat pontszámához automatikusan hozzáadjuk. Ha valamelyik zárthelyi dolgozaton a játékos nem vesz részt, akkor az összesen szerzett pontok fele a pótzárthelyi pontszámához adódik. Az adott félév szorgalmi időszak végével a pontok érvényüket vesztik. A levelező tagozaton ez arányosan értendő.
4. Az egyes fordulók mindig páratlan hét péntek éjfélig kerülnek meghirdetésre. Az egyes fordulók beküldési határideje a meghirdetést követő hét péntek éjfél. A határidőn túl beérkezett megfejtéseket nem áll módunkban elfogadni. Az eredményeket Neptun kóddal tesszük közzé a Moodle rendszerben.
5. A beküldött anyagokat a szervezők későbbiekben felhasználhatják.

1. ábra Gamestat 1. játékszabálya

A 14 hétig tartó szorgalmi időszak páratlan heteiben péntekig került közzétételre a feladat, amit az azt követő (páros) hét végéig lehetett beküldeni értékelésre. Összességében – tagozattól függetlenül – a zárthelyi dolgozatokkal megszerezhető pontok 12%-át lehetett így elérni. Az elégséges

eléréséhez képest ez 23,5%-ot jelent, ami nagyvonalú ajánlatnak tűnik – ennek ellenére a résztvevők szám(arány)át tekintve nem lehet tömeges részvételről beszélni.
A feladatkiírások (kihívások) 14 hétre bontottan 6 körben zajlottak, igazodva az aktuálisan oktatott tananyaghoz. Az egyes kihívásokat a következő fejezetben részleteztük. Amennyiben részletesebben is tájékozódna a játékról, ezt a http://tinyurl.hu/Ao9o/ linken teheti meg.

Az egyes kihívások

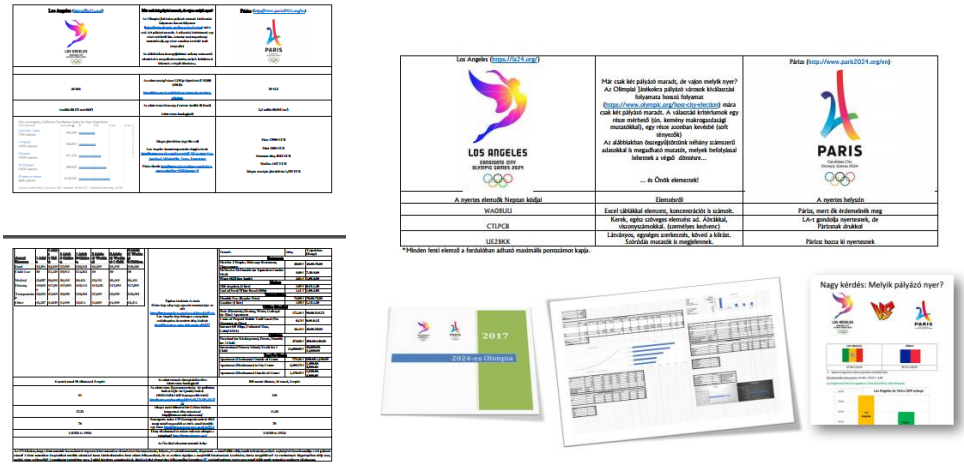
Első kihívás

Karunk névadója, Keleti Károly életéről, munkásságáról, koráról készített A/3 plakát a statisztika nyelvén. Cél: az ismérvek, statisztikai mérőszámok és skálák beazonosítása. Adatkeresés, feldolgozás szabadon, kreativitás, szerkesztés. A Kar névadójának megismerése.
Eredmény: 6 plakát született különböző stílusban és az oktatott tananyag szempontjából eltérő minőségben. Volt olyan plakát, amelyen már megjelent a különböző ismérvek beazonosítása, és értékelhető statisztikai ábrázolás, és volt, ahol csak a díszítő elemek között tűnt fel néhány (téma szempontjából nem is releváns) képlet.

2. ábra Első kihívás feladatkiírása és eredményhirdető plakátjai

Második kihívás:

A 2024 Olimpiai helyszínre pályázó városok, Los Angeles és Párizs esélyeinek összehasonlítása. (aktualitás: Budapest visszalépett) Cél: Leíró statisztikai elemzések, viszonyszámok számítása, mértékegységek, skálák, helyzet-, és szóródásmutatók alkalmazása, diagramok készítése, összevetése. Az Olimpiai Játékokra pályázó városok kiválasztási folyamatának megismerése.
Eredmény: Három pályamunka érkezett, tartalmazták a levezetéseket. További adatokat is kereshettek. Erre a kihívásra érkezett a legkevesebb pályamunka, melynek feltételezhető oka az időbeli ütemezés (4-5. szemeszteri hét) valamint a feladatkiadás mérete (nem a feladat mértéke, hanem a mérete volt nagyobb 2 darab A/3 méretű lapon szerepelt).



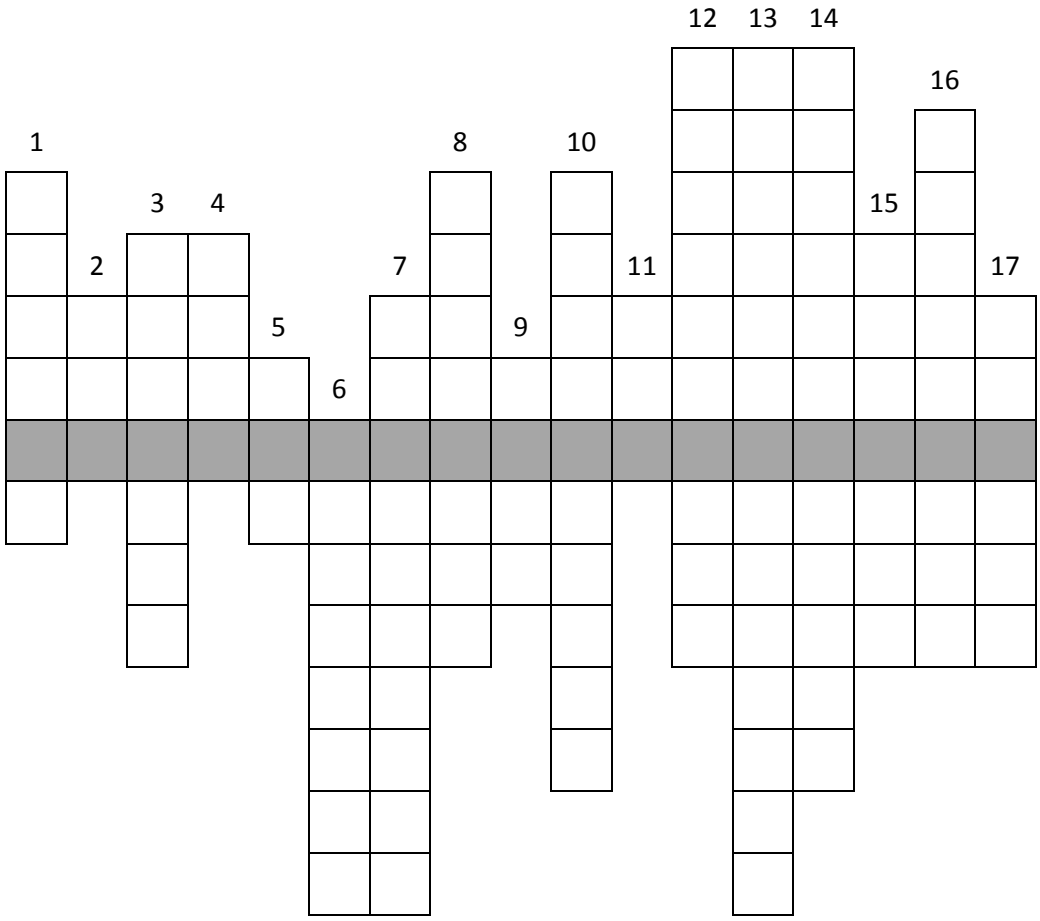
3. ábra Második kihívás feladatkiírása és eredményhirdető plakátjai

Harmadik kihívás:
A sokszínű Magyarország. Egyszerű területi összehasonlítás viszonyszámok segítségével. Ez a 3. forduló feladata volt. Cél: megismertetni a KSH honlapját, társadalmi problémák számokkal jellemzése. Eredmény: A hallgatók véleménye szerint, „túl sok munka nagyon kevés pontért, hozzá sem kezdek". Viszont, aki teljesítette ezt a kihívást ettől eltérően gondolkodott „nem is volt nagy munka, minden adat megtalálható a honlapon. 'Tök érdekes' lett az eredmény, várom a következő feladatot".
Három megoldást született. A beérkezett munkákban összehasonlításra került:

- Pest megye és Bács-Kiskun megye
- Pest megye és Vas megye
- Budapest és Baranya megye

A pályázóktól megkérdezte a kihívás kiírója, hogy mi alapján választották ki a megyéket. Bács és Vas megye esetében ez a születési hely megyéje. Baranya: 'sok érdekes dolgot hallott a hallgató erről a megyéről, kíváncsi voltam mit mutatnak a számok' válasz érkezett.
Alacsony részvétel okai között talán az is szerepel, hogy a feladat kiírása az első zh-időszakra esett. Erre az időre legközelebb egy egyszerűbbnek látszó, gyorsabban megoldható feladatot kell ütemeznünk.

Negyedik kihívás:
Rejtvényt fejteni mindenki szeret. Legalább is feltételeztük. Egy mini keresztrejtvénnyel találkoztak a hallgatók ebben a fordulóban, ahol nem egyszerűen a megoldás megtalálása volt a cél. Hogy ne lehessen erre könnyen rájönni, a megfejtés a szórás angol megfelelője volt, amit definiálni kellett.



1. A legtipikusabb érték a sokaságban, de az is lehet, hogy nincs ilyen.
2. Ezzel az ábrával a diszkrét mennyiségi ismérvek előfordulási gyakoriságát szemléltetjük.
3. Az idősorok egyik fajtája.
4. Szereti ön a statisztika tárgyat?
5. Bruttó hazai termék.
6. Index, mely olyan összehasonlítás eredménye, ami összegzett adatok felhasználásával történik.
7. Az egyik számított középérték, átlag.
8. Olyan normális eloszlás, melynek várható értéke 0, szórása pedig 1.
9. Hordozójától független, tényszerű szám, szöveg, jel.
10. Két mennyiségi ismérv közötti kapcsolat szorosságát vizsgálja.
11. A 2016. évi törvény sorszáma, mely a hivatalos statisztikáról szól.
12. Az *f* és a *g* is ezt mutatja, csak kicsit másképp.
13. Valószínűségi.
14. Mérésének egyik módja a Herfindahl-Hirschman index.
15. Amit hisztogrammal ábrázolhatunk, azt ezzel is lehet grafikusan.
16. A mezőgazdasági termelőiár-index és a mezőgazdasági ráfordításiár-index hányadosa.
17. Magyar statisztikus, közgazdasági statisztikai és földrajzi író, a honismereti szemlélet és munkálkodás hazai megteremtője, a magyarországi közgazdasági statisztika első jelentős képviselője. (Elek)

4. ábra A negyedik kihívás keresztrejtvény feladata

Ötös kihívás:
Nézzünk a tükörbe! Kérdőíves felmérés készítése a társai körében 4 előre megadott + egy szabadon választott kérdéssel. Cél: kérdőív értékelés során a sztochasztikus kapcsolatok erősségének mérése (a tananyaghoz igazítva) a szabadon választott kérdés kapcsán: mire vagyok kíváncsi? Mit is szeretnék megtudni? Az erre vonatkozó kérdés pontos megfogalmazásának gyakorlása. Eredmény: 15 megoldás érkezett, ebből csak néhány tartalmazott sztochasztikus kapcsolatra vonatkozó vizsgálatot. A legtöbb pályázó a kérdőívek kiértékelésével nem jutott túl a megoszlások vizsgálatán.
A legérdekesebb szabadon választható kérdések némelyike komoly kérdést feszeget, míg néhány inkább csak formális, mások inkább „lazára veszem a figurát – lássuk mit szól hozzá a tanár" típusú.
Komoly kérdések például:
- Összesen hány napot nyaralt a múlt évben?
- Tanulmányai mellett rendszeresen dolgozik-e?
- Hetente átlagosan hányszor jutott ideje nyugodt étkezésre?

Nem túl érdekes kérdések például:
- Az elmúlt 1 évben hányszor volt kirándulni?
- Az elmúlt 1 évben hányszor volt moziban?
- Az elmúlt évben hány koncerten járt?

A tanárpukkasztónak szánt kérdések például:
- Szereti-e a kakaót?
- Az elmúlt egy évben hány új ismeretséget kötött?
- Havonta átlagosan hány forintot költ alkoholra?

A leginkább semmitmondó szabadon választható kérdések, annak figyelembe vételével, hogy a társai között kellett a felmérést végezni:
- Életkor?
- Milyen nyelven beszél/tanul?
- Legmagasabb iskolai végzettség?

A magasabb részvétel okai között a következőket látjuk. A feladat kiírása a félév végé esett. Az első zh eredményei alapján felértékelődött a megoldásért kapható 2 pont. Ugyanakkor, kérdőíves felméréssel gyakran találkoznak a hallgatók, kevésbé érezték nagyon sok munkának egy ilyen felmérést megcsinálni. Ez a beérkezett válaszok értékelésének minőségén is látható volt, igaz mély elemzést csak néhány pályázó végzett.

Hatodik kihívás

A feladat célja az volt, hogy néhány, a mindennapokban is használt, alapvető mutató számszerű értékeit megismerjék, azok alakulásáról képet kapjanak, illetve egy esetben egy gyakran előforduló fogalmi háttérrel megismerkedjenek. Ennek érdekében a KSH weboldalain való kis keresgélésre buzdítottunk a következők szerint:

Megnevezés	Érték vagy szöveges válasz
Élveszületések száma 2016-ban	(fő)
Szárazkolbász (kg) országos fogyasztói átlagára 2017. márciusában	(Ft)
Fogyasztóiár-index (2017. március)	(%)
Budapest Liszt Ferenc Nemzetközi Repülőtér forgalma (2016. január-december, menetrend szerinti járatok), induló utasok száma	(fő)
Bruttó átlagkereset (2017. február)	(Ft)
Munkanélküliségi ráta (2017 január-március)	(%)
Mért nem számítja a KSH a korábban használt létminimum mutatót? Hogyan mérik helyette a szegénységet (melyik nemzetközileg elfogadott mérőszámrendszert használjuk jelenleg)?	
Szegénységi küszöb (2 felnőtt, 2 gyermekes háztartás) 2016-ban, éves szinten	(Ft)

5. ábra Hatodik kihívás feladatkiírása

Eredmények és tapasztalatok

A tantermi oktatás során két kolléga hetente találkozott a hallgatókkal, akik személyes visszajelzéseit a következőkben lehet összefoglalni: A személyes ismeretség okán megállapítható, hogy azok várták a feladatok megjelenését, akik az előző fordulókban részt vettek. Ilyenkor érkezhetett visszacsatolás arról is, hogy egy adott típusú feladat mennyiben jelentett tényleges motivációt.

Az oktatói tapasztalatok, fejlesztési javaslatok az egyes kihívásokhoz mérten:

Első kihívás: A feladat bevezetéséhez, kihasználva az újdonság hatását, megfelelő marketingtevékenység szükséges. A közvetítőeszköz kiválasztása is fontos.

Második kihívás: Színesebb, látványosabb feladatkiírás szükséges. A pontokat gyorsan megszerezhető, instant feladatokat preferálják a hallgatók. A feladatkiadás mérete nehezen kezelhető a hallgatók számára.

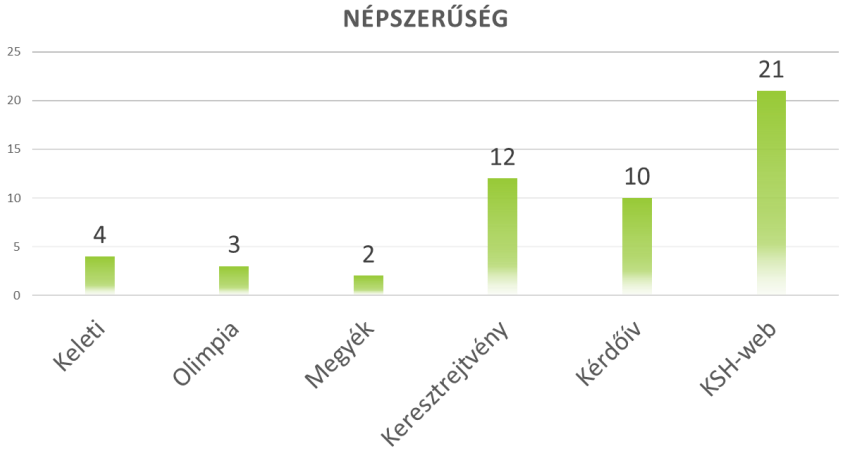
Harmadik kihívás: kevesebb vizsgálati szempontot kellett volna kérni, egy konkrét kérdéskörre kellett volna jobban fókuszálni a feladat kiírást. Majd egy újabb körben, második feladatként lehetett volna tágítani a vizsgálati szempontrendszert.

Negyedik kihívás: a rejtvényfejtés gondolata sikeresnek bizonyult, úgy tűnik, ezt a későbbiekben is alkalmazva hatékony eszköz lehet az egyes fogalmak rögzítésére.

Ötödik kihívás feladat: a kérdőív típusú feladatoknál lehet, hogy a feladat megfogalmazásban kicsivel több instrukciót kellene adni a kiértékelésre vonatkozóan.

Hatodik kihívás: Ugyan a negyedik feladathoz hasonlóan ez is olyan volt, ahol „sablon-megoldást” vártunk, a beérkezett megoldások számából következtethetünk arra, hogy egy jó megoldás elterjedt.

Az egyes feladatok népszerűségét mutatja a következő diagram, ahol az egyes feladatokra beérkezett válaszok számát hasonlítottuk össze a nappali tagozatos alapképzéses hallgatók körében.



7. ábra Az egyes kihívások eredményeinek összesítése

Érdekes lehet még megvizsgálni, hogy az egyes hallgatók hány feladatra jelentkeztek (szintén nappali tagozat, alapképzés):

Lehetőségek (db)	Létszám (fő)	g
0	113	0,82
1	14	0,10
2	4	0,03
3	3	0,02
4	1	0,01
5	1	0,01
6	2	0,01
Összesen	138	1,00

8. ábra Az egyes kihívások népszerűségének összesítése

Sajnos azonban a kihívásokkal szerezhető pontszámok valamint a két zárthelyi írásbeli dolgozaton szerzett pontszámok között nem mutatkozott korreláció. Sőt kifejezetten volt két olyan hallgató is, akik a gamestat1 pontszámaikkal érték csak el a sikeres teljesítéshez szükséges minimális pontszámot.

ÖSSZEGZÉS ÉS A JÖVŐ (GAMESTAT2.)

Célunk, hogy élvezetesebbé tegyük a statisztika oktatását megvalósulni látszott, mi oktatók legalábbis élveztük. De egyet kell értenünk a magyar középiskolai oktatásban elismert gamifikációs szakértőkkel, mely szerint „Nem könnyű eltérni a megszokott rendszertől. Nem várjuk azt, hogy a diákok azonnal belevetik magukat, elkezdenek pontokat gyűjteni és élvezik mindezt.” (Nádori & Prievara 2011)

Ebben a cikkben a GameStat 1. tapasztalatait mutattuk be, azonban a játék népszerűsége valamint a tantárgyi struktúrák tették indokolttá, hogy az aktuális 2017/18. őszi félévben a GameStat2. játék elinduljon, melybe igyekeztünk a GameStat 1. tapasztalatait is beépíteni. Ebben a félévben két nagyobb változtatás történt 1. a játék felülete teljes egészében a Moodle keretű e-learning rendszerben zajlik 2. változtatások történtek a játékszabályban, azonban ezzel nem éltek a hallgatók eddig, ezek a következők: lehet csoportosan is pályázni, a szerzett pontok elajándékozhatók. Kíváncsian állunk mi, oktatók is az új kihívások elébe, melyek eredményeiről a jövőben kívánunk visszajelzést adni.

KÖSZÖNETNYILVÁNÍTÁS

Az Emberi Erőforrások Minisztériuma ÚNKP-17-4. I és III. kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült.

FELHASZNÁLT IRODALOM

Damsa, A. (2014). Szabályok közt, szabadon! *Géniusz Műhely Kiadványok*. *http://tehetseg.hu/konyv/szabalyok-kozt-szabadon*

Nádori G. – Prievara T. (2011). *Kis nagy IKT könyv*. *Tanárblog.hu*

Rigóczky Cs. (2016). Gamifikáció (játékosítás) és pedagógia. *Új Pedagógiai Szemle*. 3-4.

Challenges and Innovations in Statistics Education
Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

MOST AKKOR EZ MIT IS JELENT? - SZÁMOK ÉS ÉRTELEM

Regina Zsuzsánna Reicher
Senior Lecturer, Óbuda University, Keleti Faculty of Business and Management, Hungary
reicher.regina@kgk.uni-obuda.hu

A gyakorlati statisztika oktatás mindig komoly kihívásokkal küzdött. Az elméleti tudás gyakorlatba való átültetése a hallgatók számára nehéz feladatnak bizonyul függetlenül attól, hogy statisztikusként vagy csak érdeklőd hallgatóként szembesülnek a problémával. Bár számos statisztikai elemző program segíti a kutatók, oktatók és hallgatók munkáját, mégis azt tapasztalhatjuk, hogy a szoftver nem helyettesíti az emberi elemzőkészséget. Nem tárja fel helyettünk az összefüggéseket, nem értelmezi az adatokat.
A statisztikai elemzésekre alkalmas szoftverek ugyan segítséget adnak a számítások elvégezésében, de emellett két nagyon fontos kérdéssel kell megküzdenie annak, aki az eredményeket később használni szeretné. Előre meg kell határozni a „Mit szeretnék megtudni?" kérdéskört, hiszen ennek ismerete nélkül a számítások nem sokat érnek majd. Emellett pedig az elemzések lefuttatása után meg kell válaszolnia a „Mit is jelentenek ezek a számok?" kérdést.

A STATISZTIKA OKTATÁS MÚLTJA ÉS JELENE

Oktatási rendszerünk még mindig a hagyományos poroszos módszeren alapszik. Az egyetemi oktatás is csak ritkán tér el az elmúlt évtizedek vagy akár évszázadok megszokott módszerétől. A régi módszerek hatékonysága azonban az új típusú hallgatók között nem mindig működik zökkenőmentesen. Több kutatás is igazolta már, hogy a bevett tanulási stratégia csak közepes teljesítményre predesztinál (Csullog et al. 2014). Sokan sokféle véleményt formáltak már meg az egyetemi szintű oktatásról, közte a statisztika oktatásról. Abban mindannyian egyetértettek, hogy gyakorlatias, aktuális és életre nevelő tananyag lenne a cél. Azonban mit nevezhetünk aktuálisnak? Mit tartunk életre nevelőnek? A tudományos világ is változik. Talán nem olyan gyorsan évül el a tartalma, mint például egy informatikai ismeretnek, de kétségtelenül frissül. Új módszerek, új irányzatok jelennek meg. A statisztikára is igaz ez az egyszerű tény. Aktuálisat tanítani nagyon nehéz, mert könnyen lehet, hogy mire a kurzus végére ér a hallgató, az tudása már nem a legfrissebb, új tudományos eredmény, új irányzat jelent meg. Így talán nem is az a legfontosabb, hogy a legfrissebb tudást adjuk át a hallgatóknak. Sokkal fontosabb, hogy a megszerezett tudást képesek legyenek adaptálni a mindennapjaikban, a munkájukban, a jövőbeli életükben.

Régebben talán könnyebb volt ennek a tárgynak az oktatása, mert nem párosult hozzá az IT állandó változó környezete. És talán éppen ezért nehezebb is volt, hiszen a számításokat, eredményeket hosszabb ideig tartott kinyerni az adatsorokból, látványosan megjeleníteni a kívánt célcsoportnak. A mai diákok a statisztika elsajátítása közben a speciális számítógépes programok ismeretébe is betekintést kapnak.

Az elmúlt évtizedek oktatási tapasztalatai alapján elmondható, hogy a leggyakoribb problémát nem a különböző statisztikai alkalmazások használatának nehézsége okozza, hanem sokkal inkább az előkészítés és az értelmezés problémája. Ennélfogva hamar rájöttünk, hogy a múltbeli problémák a technológia fejlődése mellett is érvényesek és aktuálisak maradtak. Bármilyen szoftvert is hívjunk segítségül az oktatás során, az nem helyettesítheti az értelmezést, az adtok tisztán látását, a „statisztikus gondolkodásmódot". Oktatásunk során kiemel cél, hogy hallgatóink értő gondolkodás mellett legyenek képesek a statisztikai elemzéseket elvégezni. Ne akarjanak"értelmetlen" összefüggéseket kiszámoltatni az alkalmazással.

Az új módszer nem mai. Már a 80-as években Barrows és szerzőtársa kifejlesztett egy problémamegoldáson alapuló oktatási módszert, melynek lényege, hogy a tanulók egy adott probléma megoldása során sajátítanak el valamilyen tananyagot. A problémaalapú oktatás projekt alapú tanulás tantervet és folyamatot is jelent.

A PBL és más tanulási stratégiák alapvető jellemzőit és eltéréseit a hagyományos módszerektől jól összefoglalja Savin (Savin-Baden, 2000).

Tanulási stratégia	Leírás
Előadás	A témát a tanár fejti ki.
Esetalapú módszer	Írott esettörténet és előadás alapján tartalmi és koncepcionális megbeszélés az osztályban.
Esettanulmány	Írott esettörténet, amit az osztály előzetesen tanulmányozhat, majd megbeszélhet (jellemzően kisebb csoportokban).
Módosított esettanulmány	Részleges, írott információk tanulmányozása az osztályban. A hiányzó információk meghatározása csoportmunkában. Esetleg többlet információk adhatók az osztálynak.
Problémaközpontú	A tanulók egy szimulált problémán/forgatókönyvön dolgozhatnak.
Problémaalapú	Részleges, írott információk adása és tanulmányozása az osztályban. A probléma megoldásához szükséges tanulási feladat középpontba állítása. A tartalmak és az értelmezések a kulcselemek megtanulását szolgálják.

1. táblázat Tanulási stratégiák jellemzői

Tény azonban, hogy a statisztika oktatás alap problémája a matematika oktatásunk problémáin alapszik. Már Pintér is megfogalmazza, hogy a matematika oktatás alapját a problémamegoldás kellene, hogy képezze (Pintér, 2012). Hasonló vizsgálat alapján közel azonos eredményre jutott Molnár Gyöngyvér, aki 9–17 évesek komplex problémamegoldó képességét és iskolai tudását hasonlította össze. Eredményei azt mutatták, hogy a matematikai gondolkodás összetett folyamat, több más tantárgyhoz, kompetenciához és készséghez erősen kötődik, melyek fejlesztése nélkül, csak önmagában, a matematikai problémára fókuszálva nem lehetséges.

A statisztika jövője

Ezeket, az egyébként közoktatásra kifejlesztett módszereket kell átültetni az egyetemi oktatás gyakorlatába. Azonban az egyetemi oktatók óriási kihívással állnak szemben, amikor a tömegoktatás keretében egyénre szabott módszereket, kiscsoportos technikákat akarnak alkalmazni. 200-250 fős nagyelőadások során nem tudunk lehetőséget találni erre. Gyakran marad tehát az előadás, a visszajelzések nélküli „tudásátadás”. Az egyetlen lehetőség talán a gyakorlatokon lehet, ahol kisebb csoportokban, 25-30 fővel megteremthető a problémaalapú oktatás, az értő közönség bevonása.

Statisztikai elemzésekre számos tudományterületen szükség van. Ez a tudomány nem önmagáért, hanem a gyakorlati alkalmazásért fejlődött és fejlődik mind a mai napig. Éppen ezért a gyakorlati problémák megtalálása nem okozhat gondot. Minden szakon, minden hallgatót megérinthetünk valami aktuális, a jövőbeli szakmájához szorosan kapcsolódó kérdéssel, melyre a választ egy jól elkészített statisztikai elemzésben találhatják meg.

STATISZTIKAOKTATÁS A GYAKORLATBAN

Szembesülve a felsorolt problémákkal és nehézségekkel sokat gondolkodtunk a lehetőségeken. Szembe kellett néznünk azzal a ténnyel is, hogy a hallgatók egy igen jelentős része kötelező rossznak tartja a statisztikát, nem érdeklődik, nem motivált. Ezért egy olyan lehetősége mellett döntöttünk, mely azokat a hallgatókat célozza meg, akik nem zárkóznak el teljesen a statisztika kötelező elsajátítása feletti, hasznosítható tudás elől.

A statisztika I és statisztika II előadás és papír alapú gyakorlat teljesítése után – ahol a gyakorlatot is már a hagyománytól eltérő módon, gamification módszeren alapuló formában teljesíthetik a hallgatók – egy kifejezetten problémamegoldáson alapuló és statisztikai elemzésre alkalmas szoftveres megoldásra építő szabadon választható egy féléves kurzusra jelentkezhetnek az érdeklődők. A kurzus kiscsoportos, a labor méreteinek következtében 20-23 fős lehet. a tárgy feltétele az alapstatisztikai ismeretek elsajátítása.

A félév menete

Alapvetően két részre bontható a félév. Az első részben együtt próbálunk meg egy piaci vagy vállalati problémát megoldani. Minden órát egy problémafelvetéssel kezdünk. A témaválasztás attól függ, hogy az adott kurzuson milyen képzésről érkezett hallgatók vesznek részt. A BSc hallgatók a Microsoft Excel programot használják. A probléma felvetése és a „brainstroming" során eljutunk oda, hogy szükséges lenne az adott terület átvizsgálása, idősoros adatok gyűjtése. Közösen határozzuk el, hogy milyen adatokra van szükség, rávilágítva arra, hogy az adatok formája befolyásolhatja a későbbi kiértékelés folyamatát. Igyekszünk szem előtt tartani a kutatás célját. Tapasztalatom szerint ez komoly nehézség a hallgatók számára.

A lekérdezett, összeállított adattáblát készen kapják, azonban ezek nyers állományok, minden, a lekérdezés és kódolás közben keletkezett hibát tartalmaznak. Így az első feladatuk az adatbázis adatainak tisztítása, az adatok elemzésre alkalmas formába öntése. Ekkor ellenőrizzük azt is, hogy mely adatokon milyen típusú vizsgálatokat hajthatunk majd végre. Ekkor gyűjtjük az ötleteket, mely kérdésre, milyen statisztikai elemzésekkel tudunk majd választ találni.

Az adatbázis elemzését közösen végezzük, lépésről lépésre haladva, megismerve az adatokat és a kiértékelésben segítséget adó Excel lehetőségeit. Először egyszerű leíró statisztikákat készítünk. Minden esetben megbeszéljük, hogy melyik eredmény mire enged következtetni, milyen stratégiai lépéseket tehetünk az eredmények láttán. Az egyszerű leíró statisztika után hamar elkövetkezik annak a problémának a felmerülése, hogy két adat együttes ismeretében szeretnénk információt kinyerni. Ennek megoldását a kereszttáblás elemzések adják. A kereszttáblák értelmezése sok további elemzés alapját képezik, ezért hosszú időt engedünk a hallgatóknak az ebben való jártasság megszerzésére. A sorszázalékok és oszlopszázalékok értelmezése, használhatósága után újabb változók bevonásával többszintű kereszttáblák vizsgálatára is lehetőségük van. Ezen elemzések hatására minden esetben következtetéseket próbálnak meg megfogalmazni, hogy az eredmények láttán ilyen vállalati stratégia kialakítása várható, milyen piaci változásokra lehet számítani, milyen marketing vagy termelésmenedzsment döntés lenne célszerű.

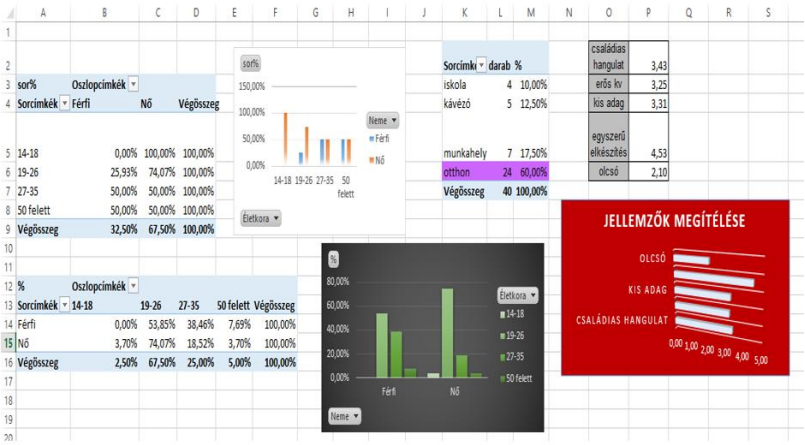
A következő lépésekben az összefüggés vizsgálatok következnek. Ennek kivitelezése az Excelbe gyakran csak több segédszámítással kivitelezhető, de jó alkalom ez arra, hogy a kereszttábla ételeivel kapcsolatos ismereteket elmélyítsük. A korreláció és a chinégyzet próba értelmezése fontos szerepet játszik a kurzus céljaiban. az ok-okozati összefüggések feltárásának feltételeit minden esetben átbeszéljük, vizsgáljuk és tisztázzuk értelmezési korlátait.

A feladat végén szót ejtünk az Excelben elvégezhető egyéb vizsgálati módszerek lehetőségéről, de csak abban az esetben számolunk vele, ha a hallgatók fogékonynak látszanak erre, érdeklődőek, képességeik megengedik. A fő cél, hogy a megszerzett tudásuk stabil legyen, az érdeklődésüket a statisztika iránt felkeltsük, fenntartsuk.

Az önálló feladat

A félév második felében egy önálló kutatást kell elvégezniük. A félév elejétől kezdve közösen gondolkodnak egy olyan témán, amelyben mindenki találhat magának érdekes kihívást, kérdést. A téma kiválasztása után a hallgatók közösen meghatározzák a kutatás célját, célcsoportját, megfogalmazzák a hipotéziseiket a témával kapcsolatosan, majd összeállítanak egy rövid, lényegre törő kérdőívet. A kérdőív összeállítása után megtervezik, hogyan és hol tudják legkönnyebben elérni a választott célcsoportot. Természetesen nem törekszünk reprezentatív mintavételre – bár a mintavételi eljárásokat tisztázzuk a halhatókkal. A kiértékeléskor minden esetben ki kell térniük arra a tényre, hogy az eredmények csak az adott adatbázisban szereplő válaszadókra érvényesek. A 20-22 fő összesen kb. 120 kérdőívet töltet ki. Mindenki papír alapon dolgozik, hogy a kódolás feladata is megjelenjen. Az így elkészült adatbázisrészleteket összefésülik majd tisztítják a bennmaradt hibákat. Eddig dolgoznak közösen. Az ezután következő elemző feladatok már önálló munka formájában zajlanak.

Minden hallgató azt vizsgálja az adatbázisban, ami az ő érdeklődését leginkább felkeltette. Ehhez kereshetnek szekunder statisztikai adatokat és összehasonlíthatják a saját adataikkal. Minden hallgató olyan nehézségű elemzést végez, ami a számára teljesíthető. Nincs kötelező minimum. A kutatásból származó számosított eredményeket egy kutatási jelentésben kell összefoglalniuk, mely tartalmazza az egyes eredmények magyarázatát is. Minden kutatási jelentés magában foglal néhány gyakorlati javaslatot az adott területre vonatkozóan.



1. ábra Egy hallgatói megoldás

Az értékelés szempontjai az órai aktív részvétel, az önálló ötletek, a számítások helyes elvégzése, a kapott eredmények megfelelő értelmezése. Nem cél a hallgatók erős nyomás alá helyezése, viszont feltett szándékunk a statisztika megkedveltetése.

Távlati célunk hasonló kurzus bevezetése A MSc képzésünkön, ahol SPSS vagy más statisztikai szoftver alkalmazása mellett kifejezetten vállalati problémák megoldását keresnénk a statisztika segítségével.

CONCLUSION

A statisztikus és nem statisztikus hallgatók oktatásában egyaránt célravezetőnek bizonyult a valós adatok, gazdasági problémák elemzése, önálló kutatási munkára és a tanultak minél szélesebb körben való alkalmazására való ösztönzés, mely során a hallgatók olyan, a mindennapokban fellelhető problémákra keresik a választ, mely akár közvetve vagy közvetlenül az ő életükre is hatással van.

Oktatási célunk, hogy hallgatóink korszerű ismeretekre tegyenek szert a statisztika gyakorlati ismereteinek elsajátítása során, amely alkalmassá teszi őket a piaci kihívások gyors és hatékony kezelésére, piacképes tudásuk előnyt jelenthessen a munkaerőpiacon. Fontosnak tartjuk, hogy olyan tudás birtokába kerüljenek, melyet egy kisvállalkozás vagy egy multinacionális cégnél való elhelyezkedés esetén egyaránt tudnak hasznosítani. Kiemelt figyelmet fordítunk arra, hogy segítsük az önálló gondolkodást, kutatási problémák feltárását, összefüggések feltárását, ok-okozati viszony vizsgálatát, hogy gondolkodó, statisztikát kedvelő, ismerő és alkalmazni tudó nemzedéket neveljünk.

REFERENCES

Barrows, H. S. & Tamblyn, R. M. (1980). *Problem-Based Learning an Approach to Medical Education.* New York: Springer Publishing Company.

Csullog K., D. Molnár É., Herczeg B., Lannert J., Nahalka I., & Zempléni A. (2014). *Hatások és különbségek. Másodelemzések a hazai és nemzetközi tanulói képességmérések eredményei alapján*. Budapest: Oktatási Hivatal.

Molnár Gy. (2004). Problémamegoldás és probléma-alapú tanítás. *Iskolakultúra, 2004*(2) 12-19.

Pintér K. (2012). *A matematikai problémamegoldás és problémaalkotás tanításáról, doktori értekezés*. Szeged: Szegedi Tudományegyetem Természettudományi és Informatikai Kar Bolyai Intézet.

Savin-Baden, M. S. (2000). *Problem-based Learning in Higher Education: Untold Stories. Buckingham:* The Society for Research into Higher Education & Open University Press.