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*