COMPETENCE, TRAINING AND COLLABORATION OF UNIVERSITIES WITH NATIONAL STATISTICAL OFFICES

1. Introduction

In 2015, I published an article on “Methodological issues and challenges in the production of official statistics” (Pfeffermann, 2015). In Section 10 of that article, I raised the following question: Are universities training students to work at National Statistical Offices (NSOs)? My starting point was that the importance of NSOs and other similar organizations is without a doubt, and that NSOs are among the largest employers of statisticians and economists. I believe that this is still the case.

To answer this question, I browsed the curricula of undergraduate and postgraduate courses in statistics and economics at the top twenty-five universities in the world, as ranked by the Shanghai Academic Ranking of World Universities. My findings at that time indicated a big shortage of specialized courses relevant to official statistics, such as survey sampling, time series analysis and national accounts. These topics underlie the work of NSOs. Only 40% of the top 25 universities offered specialized courses on survey sampling. Many universities offered courses on time series analysis, but these courses have not focused on topics such as seasonal adjustment or trend estimation, which are produced and published routinely by NSOs. Courses on themes related to national accounts were absent in most leading universities.

We are now 10 years later and for writing this paper, I revisited the courses offered by the top 25 universities as ranked in 2024.
https://www.shanghairanking.com/rankings/gras/2024/RS0502

The list of the top universities has somewhat changed, but the findings are similar. The availability and content of core courses vital for the work of NSO’s have not changed in any significant way. However, several significant developments have occurred, mostly inspired by the huge technological advances in the last decade.

Acknowledgment: I am deeply grateful to Mr. Yoel Finkel, Deputy Director of Israel’s CBS, for many thoughtful comments and suggestions.

2. Changes in programs of statistics departments at universities

The last decade witnessed the revolution of what is known as data science, which had a fast impact on the courses offered by departments of statistics all over the world. In fact, many of the departments changed their title from “department of statistics” to “department of statistics and data science”, a small change but with a huge impact. At the Hebrew University of Jerusalem, the number of students studying statistics has doubled along the years since the change of their title and curricula. I surmise big increases in the numbers of students in departments of statistics in other universities, worldwide.

Although universities still do not dedicate courses solely to topics like survey sampling, seasonal adjustment or national accounts, many universities now offer programs in data science. The addition of courses on machine learning, big data analysis, data visualization, new advanced computer software and alike to the traditional statistics program is a big change. In addition, programs like the European Master in Official Statistics (EMOS), the University of Maryland Joint Program in Survey Methodology (JPSM) and the University of Southampton Master program in official statistics, (see Appendix 1) have expanded, to better bridge the gap between the academia and real-world statistical production. The integration of data science into the traditional curricula of department of statistics in universities is becoming standard practice. Increased collaboration between statistics, computer science, engineering, and social sciences reflects the interdisciplinary nature of data science. New programs combine traditional statistics with computational and programming skills.

With the rise of data-driven policymaking, university programs now place more focus on data ethics, privacy concerns and the effective communication of statistical results. Courses in data visualization and statistical storytelling are being emphasized, ensuring that students are prepared to present data in an understandable and responsible way. The high demand for data scientists in industry has influenced academic changes, with programs placing greater emphasis on skills like data visualization, predictive modeling, and ethical data use. I should add to this that Massive Open Online Courses (MOOCs) and certifications in data science and official statistics are supplementing traditional university education. These platforms make

specialized training accessible to students and professionals around the world, providing a global reach for official statistics education. See Appendix 2 for examples.

3. Changes in the work of National Statistical Offices

Parallel to the big changes in the curricula of university statistics departments, NSOs have also expanded quite dramatically their methods of work, with more and more emphasis on data science, and increasing use of nonprobability samples (voluntary web surveys and possibly big data), although the latter is still primarily in a research stage. There is increasing reliance on administrative records, web scraping, and sensor data, and the use of machine learning (ML) methods for the production of official statistics. Classical examples for the use of ML include the choice of the correct address of residents, when there exist different addresses in different files or registers, the use of text recognition (natural language processing) to convert written occupation of respondents into given categories and in Israel, categoration of the degree of religiosity. Many NSOs have now separate departments of Research and Development (R&D), with the aim of incorporating data science techniques for the production of official statistics.

4. So where are we now?

Reading the last 2 sections seems to lead to the conclusion that nowadays, universities are well preparing students to work in NSOs or other organizations producing official statistics. There is growing emphasis in NSOs on the use of data science methodology, and universities teach data science courses as part of their curricula. However, as I discuss below, the situation is far from being satisfactory.

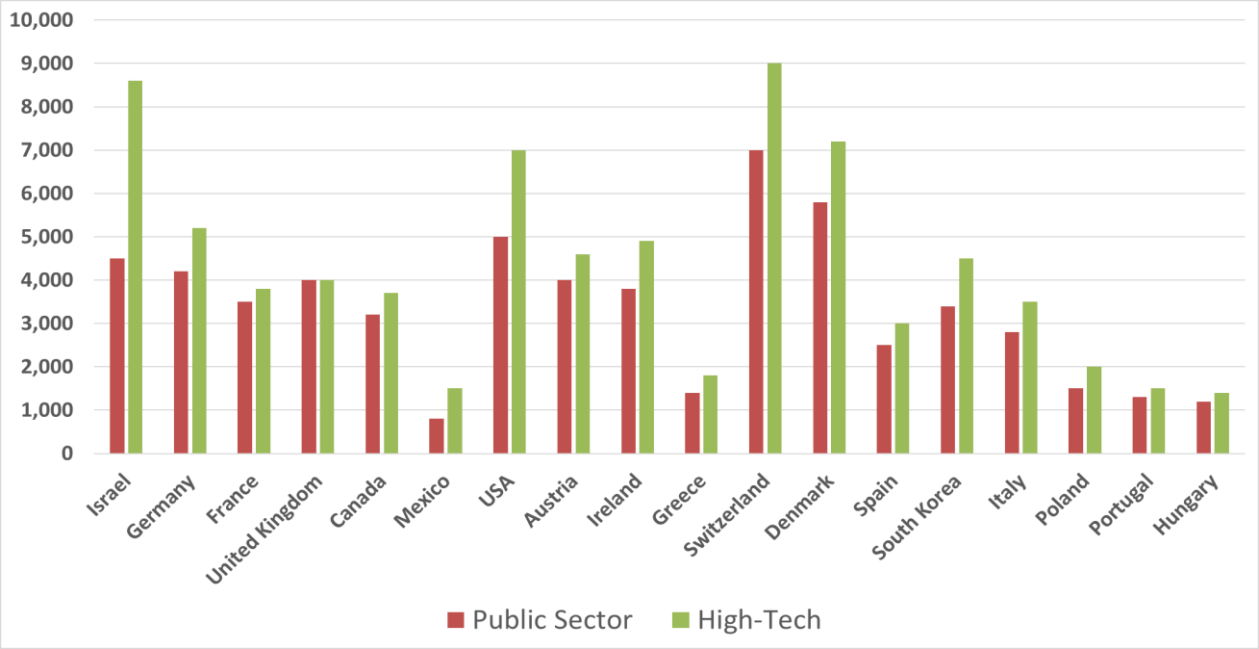
To begin with, at least in my view, classical sample survey data such as Labor Force surveys, Household expenditure surveys, establishment surveys and alike will continue to provide the core source of data for the production of official statistics in many years to come. (There are many statisticians claiming that in the future, official statistics will be based mostly or solely on administrative files, web scrapping, sensor data and alike, but we are not there as yet.) NSOs carry out more than 60 surveys annually. Students with expertise in data science but without knowledge about traditional surveys, how they are selected and how the official statistics are produced from them will not be of great help in the running of traditional surveys. As I mentioned before, there is growing use of data science technology in NSOs so clearly, the

recruitment of experts in data science at NSOs is well in place. However, this will not resolve the problem I just mentioned regarding the need for students with sufficient knowledge in methods underlying the daily work of NSOs.

In fact, data science can also gain a lot from formal courses in survey methodology. In a recent article, Eckman et al. (2024) make the following statement. “Whether future AI models are fair, trustworthy, and aligned with the public’s interests rests in part on our ability to collect accurate data about what we want the models to do. However, collecting high-quality data is difficult, and few AI/ML researchers are trained in data collection methods. Recent research in data-centric AI has shown that higher quality training data leads to better performing models, making this the right moment to introduce AI/ML researchers to the field of survey methodology, the science of data collection…. lessons from survey methodology can improve the quality and efficiency of training data and thus improve models trained on those data.” More generally, in the recent conference IOS-ISI in Mexico, Silva (2024) mentioned that the "Academia can learn from NSOs how best to explore and analyse data”. In the same conference, Lundy (2024) mentions a long time collaboration between Statistics Canada and Carleton university in Ottawa, which includes a data science graduate course on data science involving nowcasting of Canadian Labor market indicators.

A second, related issue is whether NSOs can actually recruit students with data science expertise. In Israel, and I suspect in many other countries as well, this is a major problem. The simple reason for this is that the salaries in NSOs are considerably lower than in High-Tech companies and other organizations applying High-Tech techniques. Figure 1 compares the average salaries of employees in the public sector (to which NSOs ordinarily belong), with the average salaries in Hi-Tech, in selected member countries of the OECD. As can be seen, except for the United Kingdom, the average High-Tech salaries are always higher. (In Israel, there is a 100% difference.) Adding to this that there is high demand for students with degree in data science in High-Tech companies, it is not surprising that students with such a degree will much prefer to apply for a job at a High-Tech company.

Figure 1. Average monthly salaries of employees in the public sector and HiTech in selected member countries of the OECD (In US dollars).



- **The public sector includes all government offices and institutions, but there are some differences in definitions between countries.**

5. Training of employees in data science within NSO's

As mentioned above, it is quite difficult for NSOs to hire students with data science expertise on long-term contracts, and the number of new employees that they can hire is limited anyway. On the other hand, there is great interest, both by managers and other employees, to enhance their knowledge in this area and apply it in their work.

Below are actions already taken, or that should be taken by NSOs to at least partly solve this problem.

1- Training and development programs: Implementing comprehensive training programs that cover data science, available software, and digital tools. This will help staff to acquire the necessary skills to analyze and interpret data in a digital world.

2- Collaboration with Tech experts: Partnering with technology companies and academic institutions to gain insights into the latest digital tools and methodologies.

3- Emphasizing data literacy: Staff are trained not just in technical skills, but also in data literacy—understanding data sources, data quality, and the implications of data use. This is crucial for making informed decisions based on statistical outputs.

4- Adopting innovative technologies: NSOs are exploring, and more and more integrating technologies such as big data analytics, AI and ML into their production

processes. Staff training is aligned with these technologies to ensure effective implementation.

5- Promoting a culture of continuous education: To keep pace with rapid technological changes, NSOs are fostering an environment that encourages continuous professional development and learning. This includes online courses, webinars, and access to digital resources .

6- Enhancing Inter-departmental collaboration: Encouraging collaboration between different departments within NSOs ensures that staff can share knowledge and best practices related to digital tools and data management.

7- Utilizing open data platforms: NSOs are increasingly making use of open data availability, which provides staff with access to a wider range of data sources and enhance their ability to conduct comprehensive analyses and improve production.

8- Certification Programs: Some organizations offer certification programs in statistical methods, data science, and related fields to formally recognize the skills acquired by staff. The certifications, often involve professional promotion and salary increases, and are therefore in high demand. These programs are usually tailored to the specific needs of the organization and may involve partnerships with universities, professional associations, or international statistical organizations to ensure that they are up-todate with the latest trends and technologies in the field.

The approaches to digital transformation among NSOs vary between countries, influenced by factors such as technological infrastructure, government policies, budget allocations, and staff capabilitis. By adopting the kind of actions I listed above, NSOs can be better equipped to handle the challenges posed by the digital transformation and improve the range, timeliness, visualization, quality and relevance of official statistics .

6. SUMMARY REMARKS

The incorporation of data science in the curricula of departments of statistics is undoubtedly a very welcome change. NSOs all over the world need (many) employees with expertise in data science. However, at least in the near future, NSOs will continue to run many surveys for their core production of official statistics and they need employees with sufficient knowledge in areas such as classical survey sampling

theory, seasonal adjustment and trend estimation, and national accounts. Universities, do not generally offer designated courses in such areas. More cooperation between NSOs and Universities on this matter is much needed.

References

Eckman, S., Plank, B. and Kreuter, S. (2024). Position: Insights from Survey Methodology can Improve Training Data. *Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. Position: Insights from Survey Methodology can Improve Training Data

Lundy, E. (2024). Collaboration and Engagement with Academia: The Statistics Canada Experience. IOS-ISI conference, Mexico.

Pfeffermann, D. (2015). Methodological issues and challenges in the production of official statistics. *Journal of Survey Statistics and Methodology*, 425–483.

Silva, P. (2024). Win-Win Partnerships – Public Statistics Meets Academia. IOS-ISI conference, Mexico.

Appendix 1: University Programs and Specializations

1. **European Master in Official Statistics (EMOS)**
2. Focus: Advance postgraduate education in official statistics and data science at universities across Europe.
    - Participating Universities: University of Trier (Germany), University of Southampton (UK), KU Leuven (Belgium).
    - Website: EMOS
3. **University of Maryland – Joint Program in Survey Methodology (JPSM)** ○ Focus: Master's and Ph.D. programs in survey sampling, statistical modeling, and practical applications in government and research.
    - Collaboration: Works closely with the U.S. Census Bureau and other federal statistical agencies.
    - Website: JPSM
4. **University of Southampton – MSc in Official Statistics** ○ Focus: Specialized program covering survey design, time series analysis, demographic statistics and data science.
    - in partnership with the UK Office for National Statistics (ONS).

    Website: https://datasciencecampus.ons.gov.uk/capability/msc-in-data-analytics-for-government/

5. **KU Leuven – Master's in Statistics with a Track in Official Statistics**
    - Focus: Combines traditional statistical techniques with data science methods, including time series and large-scale survey design, tailored for public policy applications.
    - Website: KU Leuven Statistics

6. **Carnegie Mellon University – MS in Statistics and Data Science** ○ Focus: A blend of statistical methods, computational techniques, and data science applications, including big data analysis and official statistics. ○ Website: CMU Statistics

7. National School of Statistical Sciences at IBGE, Brazil (ENCE)
○ Focus: preparing professionals to carry out research and apply available approaches for the analysis of socio-spatial processes and for the production of public information. Master and PhD programs. ○ Collaboration: fully funded with strong collaboration of professionals from IBGE ○ Website: https://ence.ibge.gov.br/index.php/mestrado-e-doutorado/apresentacao.

Appendix 2: UN Innovative Curricula and Initiatives:

1. **Big Data and Official Statistics Training (UNSIAP)** ○ Focus: Training programs on integrating big data and modern technologies into official statistical processes.
    - Target Audience: Professionals in national statistical offices.
2. **UNESCO and Eurostat Collaboration** ○ Initiative: workshops and online courses on modern statistical practices and data ethics for official statistics professionals. ○ Examples: Courses on using machine learning and administrative data sources in official statistics.