**Collaboration with universities**

*– a way to promote innovation and experimental culture at statistical offices*

Faiz Alsuhail, Statistics Finland, faiz.alsuhail@stat.fi

**Abstract**

*Statistics Finland has had fruitful collaboration with the University of Helsinki and the University of Jyväskylä for several decades. Thesis traineeships are an important part of this collaboration. Each year, a handful of students are recruited for a period of six months. The main task of the student is to write a thesis on a given topic. Statistics Finland provides the student with a research problem, a data set, facilities and day-to-day guidance. The academic supervisor is from the student’s university. This paper presents how a research problem is drafted in collaboration with the academia, how the students are recruited and supported during the traineeship as well as what the trainee program has produced. We illustrate with concrete examples how the thesis traineeship can promote innovation and experimental culture in a statistical office. These examples involve testing new tools and methods. We argue that the thesis traineeship is a good way to involve the scientific community in the activities of a statistical authority. In particular, we show that the thesis traineeship is not only a mutually beneficial practice for the academia and statistical offices but also for the students.*

**Keywords:** Academia, traineeship

**1. Introduction**

Statistics Finland has had fruitful collaboration with the academia for more than two decades. Thesis traineeships are an essential part of this collaboration. In this paper, we argue that the trainee program is a good way to involve the scientific community in the activities of a statistical office. In particular, it has promoted an innovative and experimental mind-set within Statistics Finland.

This paper is organised as follows. First we describe how innovative ideas are formalised into academic research abstracts. We argue that this process itself facilitates innovative brainstorming within a statistical office. Then, in chapter three, we describe the way thesis trainees are recruited and supported during their traineeship. We list some concrete results of the collaboration in chapter four. Chapter five concludes.

**2. Drafting research problems – a way to promote experimental culture**

Statistics Finland organises regular brainstorming exercises where potential themes for master’s theses are discussed. These usually start with identifying challenges, interesting datasets, new tools, methods and practices as well as ideas that can be tested.

The aim of these discussion is two-fold. Obviously, the primary desired output is a list of potential research topics that can later be discussed and finalised with the academia. These brainstorming exercises, however, also promote innovative thinking at the statistical office. Usually, a brainstorming exercise encourages staff members to think up problems that are interesting but – due to time or resource constraints – out of the scope of daily development activities.

These innovative and experimental ideas are usually related to interesting datasets, phenomena where further analysis is needed or areas where the thesis could complement ongoing projects and development. Sometimes, staff members come up with new experimental ideas that are fully unexplored. In these cases a master’s thesis project can be a good way to get proofs-of-concepts.

Later in this paper, we will provide the reader with some concrete examples on master’s theses that have been completed by Statistics Finland’s trainees.

Once the round of brainstorming has been completed there is a list of ideas on which a master’s thesis could be written. This list is then discussed with professors from the universities of Helsinki and Jyväskylä. The aim is to transform preliminary research ideas into more mature, academic research abstracts.

The dialogue with the academia serves three purposes.

First, it is a way to ensure that an academic thesis can be written on the problem at hand. It must be ensured that the scope of the problem is suitable for a thesis and that the – usually quite practical – challenges can be discussed within a theoretical framework. Also, this discussion ensures that the problem is formulated in a way that there will be an academic supervisor available.

Second, the dialogue is a good way to get first-hand information on relevant methods and scientific tools to tackle to day-to-day challenges of Statistics Finland. This information can be used in recruitments and drafting the profile of a future statistician.

Last, but not least, the dialogue ensures that the finalised research problem is interesting and relevant from Statistics Finland’s point of view. Hence, the dialogue allows the problems to be formulated in a way that they meet the academic requirements and respond to Statistics Finland’s interest.

A typical formulation of a research problem includes a working title, brief description of the problem at hand as well as a few relevant literature references. For each topic, relevant datasets have been identified. Statistics Finland maintains a repository of thesis topics and research problems. The repository is updated on a regular basis. Currently the list contains around 40 potential research topics in the fields of economics, statistics, data science and quantitative social sciences.

**3. How the students are recruited and supported?**

*3.1. Recruitment*

The contact professors get the final list of potential research problems to circulate to their students. The professors act as contact points between Statistics Finland and the students: e.g. it is the professors who do the grassroots marketing of the opportunity to write a thesis on Statistics Finland’s topic.

The professors not only promote the opportunity to work on a topic from Statistics Finland but they also target the most talented students with their message. The professors present potential candidates to Statistics Finland, which then recruits these students as full-time trainees. This head-hunting has resulted in Statistics Finland getting talented and motivated trainees.

The traineeship is full-time work that lasts six months. During this time the trainee writes his thesis on a given topic and dataset. The trainee is entitled to the same benefits as anyone else working at Statistics Finland, including working facilities and access to relevant data. They are paid a monthly salary, which equals around 1600 euros.

Statistics Finland recruits around 5-7 thesis trainees each year.

*3.2. Support*

The students can fully concentrate on their research problem while working at Statistics Finland. They are supported by an informal steering group, which consists of experts interested in the theme of the thesis or who have been collecting/ working on the data used in the research.

This steering group provides assistance in issues such as SAS-macros, the data used (including metadata) and provides insights on the subject matter area. The trainee and his steering group meet on a regular basis to discuss the progress of the work, give feedback, advice and to toss ideas around for additional views.

Also, all the trainee’s colleagues are available for day-to-day support.

The academic supervisor takes part in the regular meetings with the trainee and the rest of the group. The professor supports the trainee in all scientific questions and provides assistance in the methods used. In addition, the trainees tend to get additional peer support from a seminar organised by the faculty of the university. Also, having the trainee, the professor and people from the statistical office around the same table is an excellent way to promote innovative thinking.

Statistics Finland has had thesis trainees for more than two decades. Up to this day all of them have been able to complete their task of completing their thesis on the given timetable. We believe this is due to the fact that the students recruited are motivated and talented people who are actively supported by Statistics Finland’s staff members and their supervisor.

**4. What has the trainee programme produced?**

Here we give some examples of master’s theses that have been written at Statistics Finland. We claim that they have contributed to the implementation of new methods and promoted an innovative mind-set among the staff. Many of the old thesis trainees have been recruited to Statistics Finland after their graduation.

*4.1. New methods*

Multiple theses have been written on methods and statistical techniques that have later become an integral part of Statistics Finland’s methodological tool pack. Examples of such include seasonal adjustment (Turunen, 2005), statistical disclosure control (Konnu, 2006, Nissinen, 2011), imputation (e.g. Piela, 1999, Vertanen 2004, Pasanen, 2005) and statistical editing (Ikäheimo, 2001).

In all these cases the trainee, together with the guidance of his supervisor, has been able to demonstrate the use of the given method. Looking back, having the thesis trainee, the professor and staff members around the same table has greatly contributed to the implementation of these statistical techniques.

The latest new methods that have been tested include web-scraping (Tzu-Jou, 2017) and machine-learning algorithms for data classification (Gellagos, 2017).

*4.2. Innovation and experimental culture*

Nowcasting is an excellent example of how collaboration with universities has promoted innovative thinking and experimental culture. The idea of improving timeliness by using continuous cumulative (micro) data sets was proposed by Statistics Finland’s staff members and formalised as an academic research project by a professor familiar with suitable time series techniques.

This work has resulted in an active dialogue between staff members and academia, interesting theses (Beniard 2010, Fornaro 2011 and Mustonen 2017) and plenty of know-how. Currently, Statistics Finland is an active partner in the Big Data ESSnet on Early Estimates and one of the leading NSI’s on nowcasting.

*4.3. Mixed mode data collection*

Engaging the academia in a major development initiative is a good way to gain insights. Having the opportunity to recruit trainees to test relevant hypotheses has encouraged people to come up with topics for research. This, on the other hand, has promoted an innovative and experimental mind-set within the statistical office.

For example, the current work on implementing mixed mode data collection has benefited from collaboration with universities. In particular, it has allowed Statistics Finland to gain information on the effects of multi-mode data collection. Both Nikula (2010) and Toivola (2014) studied how to weight a data collected on different modes and how this effects the quality of final estimates. Vartiainen (2011) studied the representativity in samples collected by using different modes.

*4.4. Data driven research questions*

Despite the demand for more thorough analysis, Statistics Finland staff has limited time to study all these questions. The collaboration with universities, however, encourages people to think up such research questions and – once they are assigned to a trainee – support the research as a steering group member.

Some innovative research projects include a study on the differences in salaries paid by large and small enterprises (Pehkonen, 2011), the effects of measurement errors in GDP announcements to output fluctuations (Kaukoranta, 2009) as well as the probability of a divorce (Wallenius, 2013).

Estimating new and emerging phenomena has been done by (among others) Peltola (2017), who estimated the volume and value of digital music and Andersson (2016) who estimated the value of shadow economy in the construction industry. Korhonen (2008) and Luukko (2016) both studied immigrants’ employment status from different perspectives.

**5. Conclusions**

As a result of our collaboration we have been able to develop, learn, fail, and recruit good people.

Thesis traineeships are a cost-efficient way of utilizing the know-how of professors and their students. Many of these trainees have continued working at Statistics Finland after graduation. We see the trainee program as an important recruitment channel for new talents.

The collaboration stimulates creative thinking and promotes experimental and innovative activities within an NSI. We have seen, that having thesis trainees has contributed to research and development at Statistics Finland as well as helped in implementing new statistical techniques.

**6. References**

Andersson, A. (2016), The shadow economy in Finland: The Effects of the Reverse Value-Added Tax Reported Revenue and Deductions, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/163860> (Accessed: 16 April 2018).

Asikainen, H. (2010), The effects of public news on consumers’ economic confidence (in Finnish) Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/18052> (Accessed: 16 April 2018).

Beniard, H. (2010), A Composite Leading Indicator of the Finnish Economy, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/17591> (Accessed: 16 April 2018).

Fornaro, P. (2011), Dynamic Factor Models and Forecasting Finnish Macroeconomic Variables, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/26610> (Accessed: 16 April 2018).

Gellagos, A (2017) Occupational code correction via outlier analysis and hierarchical classification, Faculty of Science, University of Helsinki.

Ikäheimo, J. (2001): Editing and imputation of industrial tax register data. Comparison of selected methods (in Finnish). Faculty of Science, Univestity of Jyväskylän.

Kaukoranta, I. (2009), Do measurement error in GDP announcements cause output fluctations, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/16496> (Accessed: 16 April 2018).

Konnu, J. (2006), Statistical disclosure control methods for micro data in social statistics (in Finnish). Faculty of Science, Univestity of Jyväskylän.

Korhonen, K. (2008), Estimating employment figures on immigrants from Statistics Finland’s LFS data (in Finnish), Luonnontieteellinen tiedekunta, Oulun yliopisto.

Luukko, J. (2016), Over-qualification of highly educated immigrants (in Finnish) Jyväskylän yliopiston kauppakorkeakoulu, Jyväskylän yliopisto. Available at: <https://jyx.jyu.fi/dspace/handle/123456789/53251> (Accessed: 16 April 2018).

Mustonen, A. (2017) Nowcasting Turnover of trade in Finland, Faculty of Social Sciences, University of Helsinki. <https://helda.helsinki.fi/handle/10138/225228> (Accessed: 16 April 2018).

Nikula, J. (2010), The effects of data collection and weightning methods on data quality and estimates. Analysing nonresponse in the EU crime victim survey, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/17592> (Accessed: 16 April 2018).

Nissinen, A. (2011), Statistical disclosure control methods for tabular data (in Finnish) Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/29258> (Accessed: 16 April 2018)

Pasanen, A. (2005), Mass-imputation and analysis of mass-imputed healt survey data. Comparison of model-based and design-based methods. Faculty of Science, Univestity of Jyväskylän.

Pehkonen, S. (2011) The effect of enterprises size on salary. Study on 2003-2009 wage structure data (in Finnish). Jyväskylän yliopiston kauppakorkeakoulu, Jyväskylän yliopisto.

Peltola, J. (2017) Digitalization of music and Long Tail effect in Finnish music industry, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/232013> (Accessed: 16 April 2018).

Piela, P. (1999), Mass-imputation in short-term retail statistics (in Finnish), Faculty of Science, Univestity of Jyväskylän.

Toivola, J. (2014), Comparison of non-response in websurveys and telephone interviews. Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/153081> (Accessed: 16 April 2018).

Turunen, R. (2005), Analysis of SARIMA-models and docomposition on seasonal adjusment method TRAMO/SEATS (in Finnish). Faculty of Science, Univestity of Jyväskylän.

Tzu-Jou, H. (2017), Web scraping applied for official statistics. Politecnica de Catalunya.

Vartiainen, M. (2011), Studying representativity on R-indicator in the Finnish crime victim survey (in Finnish) Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/27986> (Accessed: 16 April 2018).

Vertanen, V. (2004), Imputation of regional retail flow data. Comparison of selected methods. (in Finnish) Faculty of Science, Univestity of Jyväskylän.

Wallenius, M. (2013), Multi-stage survival tables in studying the divorce probality (in Finnish). Monitilaiset elinajantaulut avioeron todennäköisyyden tarkastelussa, Faculty of Social Sciences, University of Helsinki. Available at: <https://helda.helsinki.fi/handle/10138/40172> (Accessed: 16 April 2018).