Skills for the new generation of statisticians

#CES2017Skills

Riikka Mäkinen*, Training Manager, Statistics Finland
rikka.makinen@stat.fi  @riikkamakin

Dario Buono*, Team Leader of Methodology, Eurostat
dario.buono@ec.europa.eu  @darbuo

*The views expressed are the author’s alone and do not necessarily correspond to those of the corresponding organisations of affiliation
A Data-Driven Life

The New Data Lake

- S-DATA
  - surveys
- D-DATA
  - administrative data
- G-DATA
  - geospatial data
- B-DATA
  - Big Data
New data sources:
key capabilities and skills profile impacted

New Statistics Development
- New data sources exploration
- Legislative work participation
- Methods and tools development for new statistics
- Understand user needs
- Statistical product innovation

Statistical Design
- Design statistical outputs
- Design production system
- Process and workflow design
- Process methods design

Statistical Data Collection
- Metadata collection
- Primary data collection
- Secondary data collection
- Provision agreement management

Information Resource Management
- Data management
- Metadata management

Statistical Dissemination
- Products and services promotions
- Statistical content management
- Release management

Statistical Processing
- Statistical data preparation
- Statistical Registers Management
- Statistical output analysis
- Calculation and output finalisation

New variables and units derivations

Statistical Analysis
- Statistical output analysis
- Information extraction and management

Quality assessment
- Quality assessment
- Quality control mechanisms
- Quality improvement mgmt.

Process and Methods management
Information extraction and management
Technology management
Human and other capabilities

skills profiles:

MODERN DATA SCIENTIST

Data Scientist: The second half of the 21st century requires a new breed of multidisciplinary skill sets ranging from an understanding of mathematics, statistics, technology, language, communication, and knowledge of data science at scale. Finding people who understand both data science and product design is equally hard. So here is a little cheat sheet on what the modern data scientist really is.

MATH & STATISTICS
- Modern linear algebra
- Probability theory
- Statistical modeling
- Experiment Design
- Bayesian statistics
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS
- Persistence
- Effective written and oral communication
- Influence without authority
- Focus under pressure
- Problem solving
- Business acumen
- Attention to detail

PROGRAMMING & DATABASE
- Computer languages: R
- Statistical analysis in Python
- Predictive analytics (SAS, Matlab)
- Databases (SQL and NoSQL)
- Data integrity and data quality
- Data visualization and interactive data tools
- Time series analysis
- Experience with Hadoop (Hive)

COMMUNICATION & VISUALIZATION
- Analytical insights, algorithmic thinking
- Storytelling
- Effective use of visualizations tools
- Communication with LOB teams
- Knowledge of data science textbooks and journals
- Knowledge of open data resources

Human and other capabilities
Means for acquiring skilled statisticians and supporting data science-driven work

**Recruiting & long-term personnel planning**
- Data scientist’s competence profile? Skills are seldom found in one employee
- NSIs need to have a realistic understanding of tasks and competence profiles of the future personnel

**Multidisciplinary data science teams**
- [The competency framework for Big Data team and Big Data team leader](#) (CSO Ireland & High-level Group for MOS)
- Challenges NSIs to find ways of building and maintaining horizontal teams
Means for acquiring skilled statisticians and supporting data science-driven work

**Personnel training**
- Courses and training programmes in specific competence areas
- *Statistical education in times of big data* (Destatis)

**Collaboration with universities**
- Courses and training programmes
- Trainees and thesis workers

**Networks and outsourcing services**
- Partnerships with research and data science networks

Means for acquiring skilled statisticians and supporting data science-driven work

*Statistical offices need to adopt an innovative culture where experimental activities are commonplace; this is important since future statistical work requires creative thinking and abilities and a will to do things differently.*

**Supporting innovative culture**
- Acknowledging current personnel’s interest and potential
- Having experimental activities
- "Letting the past go", statisticians need to identify themselves with new competence requirements and new working methods
Means for acquiring skilled statisticians and supporting data science-driven work

*Management and leadership has a crucial role in*
- realising personnel planning
- enabling team work in organisation structure
- supporting innovative culture
- allowing time and space for experimental activities
- using participatory ways in defining the common future in statistical offices: the ways statistics are produced, what the needed competences are and how can these be acquired

How can an official statistician become a data scientist?

“Being a data scientist is not only about data crunching. It’s about understanding the business challenge, creating some valuable actionable insights to the data, and communicating their findings to the business.”

Jean-Paul Isson, Monster Worldwide, Inc.

“The real question is not how a statistician can become a data scientist but how statistical organisation can include “data sciences” in their toolbox and how the needed competences, skills and features can be fused into the professional identity of official statisticians.”

Skills for the new generation of statisticians by Statistics Finland and Eurostat
Thanks!

Riikka Mäkinen, Training Manager, Statistics Finland
riikka.makinen@stat.fi
@riikkamakin

Dario Buono, Team Leader of Methodology, Eurostat
dario.buono@ec.europa.eu
@darbuo