INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Contributes of Statistics Portugal in Teaching and Research in Official Statistics

Francisco Lima (INE), Pedro Campos (INE), Carlos Marcelo (INE)

Statistics Portugal

Outline

- Training in official statistics
- Research
- EMOS
- A Matching Between National Statstical Offices' Demand and Universities' Offer
- Data Science skills
- Discussion

Introduction

- Over the years, Statistics Portugal has sought to keep pace with **scientific developments** in the area of Statistics and in other related fields.
- In addition, it has been particularly important to continue to train its staff internally.

Training in Official Statistics

- This training has occurred in various ways
  - international courses organized within the framework of Eurostat
  - tailor-made courses administered by external entities
  - in-house training courses taught by other colleagues with knowledge and experience in the areas.

Training in Official Statistics

- Main areas of training
  - Statistical software (Stata, R, Python)
  - Econometrics
  - Macroeconomics
  - Data Science/Machine Learning
  - National Accounts

Training and Research in Official Statistics

- EUROPEAN SYSTEM TRAINING PROGRAMME (ESTP - 2021)
  - Flash estimates and nowcastig
  - Digital Dissemination
  - Big Data Tools for Official Statistics
  - Data Linkage
  - Privately held data based on partnerships
  - Statistical disclosure control
  - Nacional Accounts
  - Advanced methods for Sample Surveys
  - Impacts of COVID-19

eurostat

Main Jounals

- Journal of Official Statistics
  - Published by Statictics Sweden

https://sciendo.com/journal/JOS

- Statistical Journal of the IAOS

Official Statistics

https://officialstatistics.com/

EMOS

– Recently, Statistics Portugal became associated to EMOS, (**European Master in Official Statistics**).

- **EMOS** is a network of Master programs that was created to strengthen the collaboration between academia and the producers of official statistics and to help develop professionals able to work with European official data at different levels in the rapidly changing production system of the 21st century.

EMOS Learning Outcomes

- Learning outcomes are usually defined by **Eurostat (2021),** and are aligned with the goals of the European Statistical System (ESS):

1. System of Official Statistics

2. Production model and methods (quality issues, sources, architectures)

3. Specific Themes (economic, Finance, Population, Environment, Energy...)

4. Statistical Methods (sampling, index numbers, ...)

5. Dissemination

A Matching Between National Statstical Offices' Demand And Universities' Offer

Statistical theory, Econometrics, Classical and Bayesian Inference, Machine Learning and other topics are already covered in the syllabus of many EMOS courses.

However, there are other areas in which **research and training** is needed:

- Administrative Data
- Big Data
- Statistical Confidentiality

A Matching Between National Statstical Offices' Demand And Universities' Offer

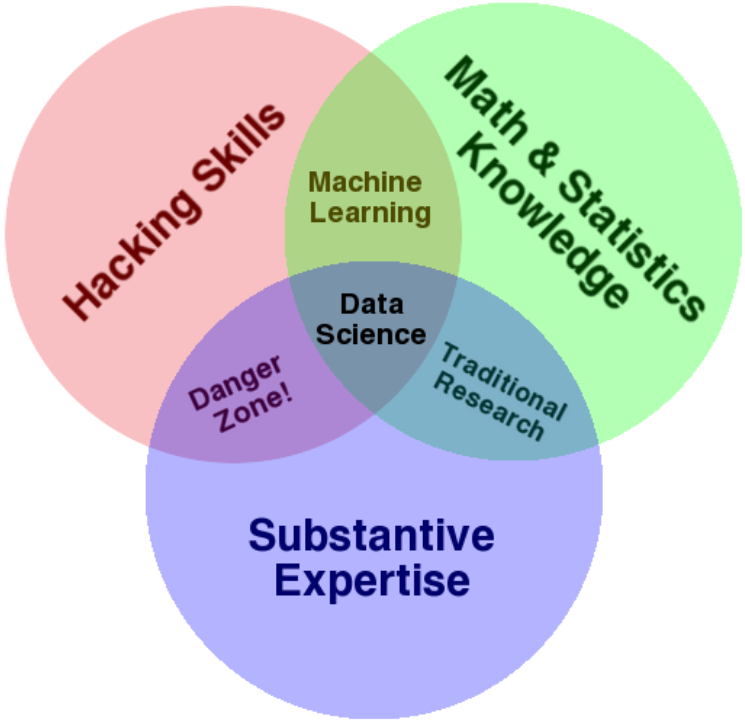
- There are a number of scenarios in which Big Data could be used in statistical organizations (HLG project on Big Data in Official Statistics, 2014):
  - **Scenario 1: use as auxiliary information to improve all or part of an existing survey**
  - **Scenario 2: supplementing/replacing all or part of an existing survey with Big Data**
  - **Scenario 3: producing a predefined statistical output either with or without supplementation of survey data**
  - **Scenario 4: producing a statistical output guided by findings from the data (HLG project on Big Data in Official Statistics, 2014).**

A Matching Between National Statstical Offices' Demand And Universities' Offer

Required EMOS learning outcomes, and the currently most important official statistical needs (Pratesi and Campos, 2021).

EMOS learning outcomes	Administrative Data	Big data/Smart Statistics	Privacy and Confidentiality
1. System of Official Statistics	√	√	
2. Production model and methods (quality issues, sources, architectures)	√	√	
3. Specific Themes (economic, Finance, Population, Environment, Energy...)	√	√	
4. Statistical Methods (sampling, index numbers, ...)	√	√	√
5. Dissemination	√	√	

Data Science skills



» Drew Conway Venn's diagram:
http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Discussion / Conclusions

- The business informatisation, the internet popularisation and the emergence of the big data phenomenon lead to non-conventional data sources which may contribute to official statistics.
- However, to make this contribution effective the next generation of statisticians must be able to master methods to analyse data characterized by Volume, Velocity, Variety.

Discussion / Conclusions

- The new statisticians must be familiar with methods that allow to check for Veracity and to extract Value from the data.
- In addition, it is important to know how to deal with new sources of data, namely administrative data, and the corresponding issues of metadata and quality.

Discussion / Conclusions

- In a world with more and more data sources, it is increasingly important to ensure the protection of individual data. To this end, it is urgent to teach methods of Statistical Disclosure Control.
- These requirements point to the need of extending the skills of competencies of the statisticians into the realm of Data Science.
  - National Centre of Ccompetences in Data Science
  - EMOS track