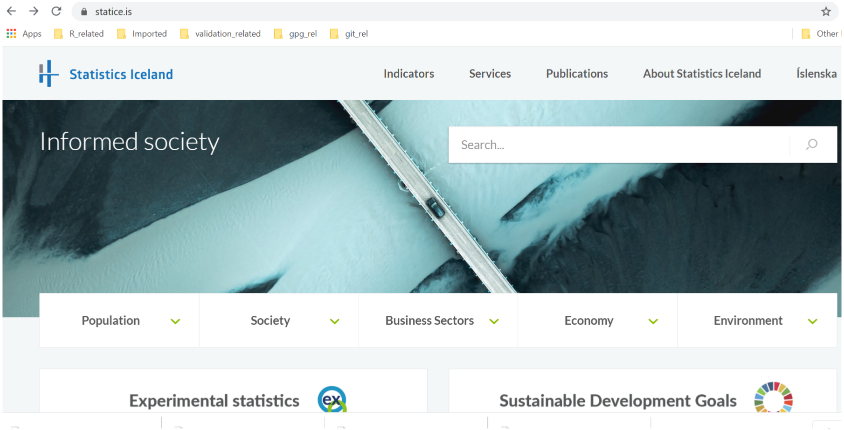
Correcting for population overestimates by using statistical classification methods

Violeta Calian and Margherita Zuppardo
Statistics Iceland

NTTS-2021

About us



statice.is

Context and focus

- Solving an official statistics problem, **using** *R*
- Local context: basic principles at Statistics Iceland, while increasing **efficiency**
  - reproducibility, flexibility
  - transparency, peer review, collaboration
  - based on scientific/statistical knowledge
- General context: evolution of goals in official statistics:
  - description
  - modeling, estimation, **prediction**
  - reporting **uncertainty**
  - up-to-date!

Formulation of the problem

The social statistics problem:

- Overestimating population size of a given country/area at a given time

- Main cause: de-registration issues
- Impact: bias in estimates of demographic/social measures

- Input data: from multiple registers / databases (i.e. large set of attributes)
- To predict: status as present/absent
- Status of solutions across NSIs: *fuzzy* and *SOL*

Examples of NSIs' solutions

- indices on SOL (Estonia)
- scores on SOL (Sweden)
- few classification models (Iceland, 2011 Census: cumulative link for ordinal response)

While in the whole wide world:

- deep learning
- ensemble learning
- ML applications for: face/speech recognition, financial and medical fields, . . .

Specific questions:

- *training* data and issues with the *status* variable: delay and noise
- a completely different formulation of the problem: macro vs micro
- interpretability and solutions

R / open source code:

Key role in modernizing official statistics

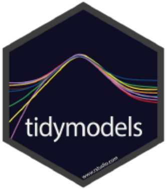
illustrated by the present work

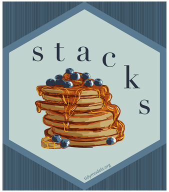
developed for the purpose of Census 2021

but also for routine population estimates

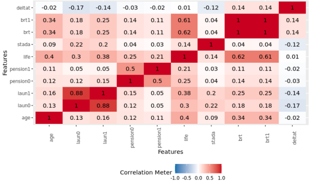
The main R - packages

tidyverse
www.rstudio.com

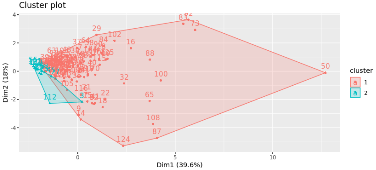


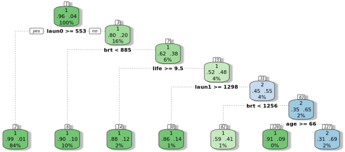


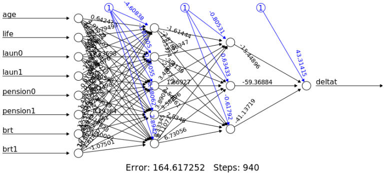
Results



	age	laun0	laun1	pension0	pension1	life	stada	brt	brt1	deltat
deltat	-0.02	-0.17	-0.14	-0.03	-0.02	0.01	-0.12	0.14	0.14	1
brt1	0.34	0.18	0.25	0.14	0.11	0.61	0.04	1	1	0.14
brt	0.34	0.18	0.25	0.14	0.11	0.62	0.04	1	1	0.14
stada	0.09	0.22	0.2	0.04	0.03	0.14	1	0.04	0.04	-0.12
life	0.4	0.3	0.38	0.25	0.21	1	0.14	0.62	0.61	0.01
pension1	0.11	0.05	0.05	0.5	1	0.21	0.03	0.11	0.11	-0.02
pension0	0.12	0.12	0.15	1	0.5	0.25	0.04	0.14	0.14	-0.03
laun1	0.16	0.88	1	0.15	0.05	0.38	0.2	0.25	0.25	-0.14
laun0	0.13	1	0.88	0.12	0.05	0.3	0.22	0.18	0.18	-0.17
age	1	0.13	0.16	0.12	0.11	0.4	0.09	0.34	0.34	-0.02







Error: 164.617252   Steps: 940

Next?

https://github.com/violetacln/SLOPA

violeta.calian@hagstofa.is and margherita.zuppardo@hagstofa.is

Thank you!