

Imputation of occupation in the Occupational Register

Jens Malmros
Statistics Sweden



Introduction



The Occupational register

- Formed in 2001
- Swedish population 16 years and older
- Official statistics on gainfully employed 16-64 years old



SSYK classification

- Based on ISCO-08
- SSYK2012
- 425 occupations
- 4 digit-code
 - Hierarchical structure
- Imbalanced population distribution



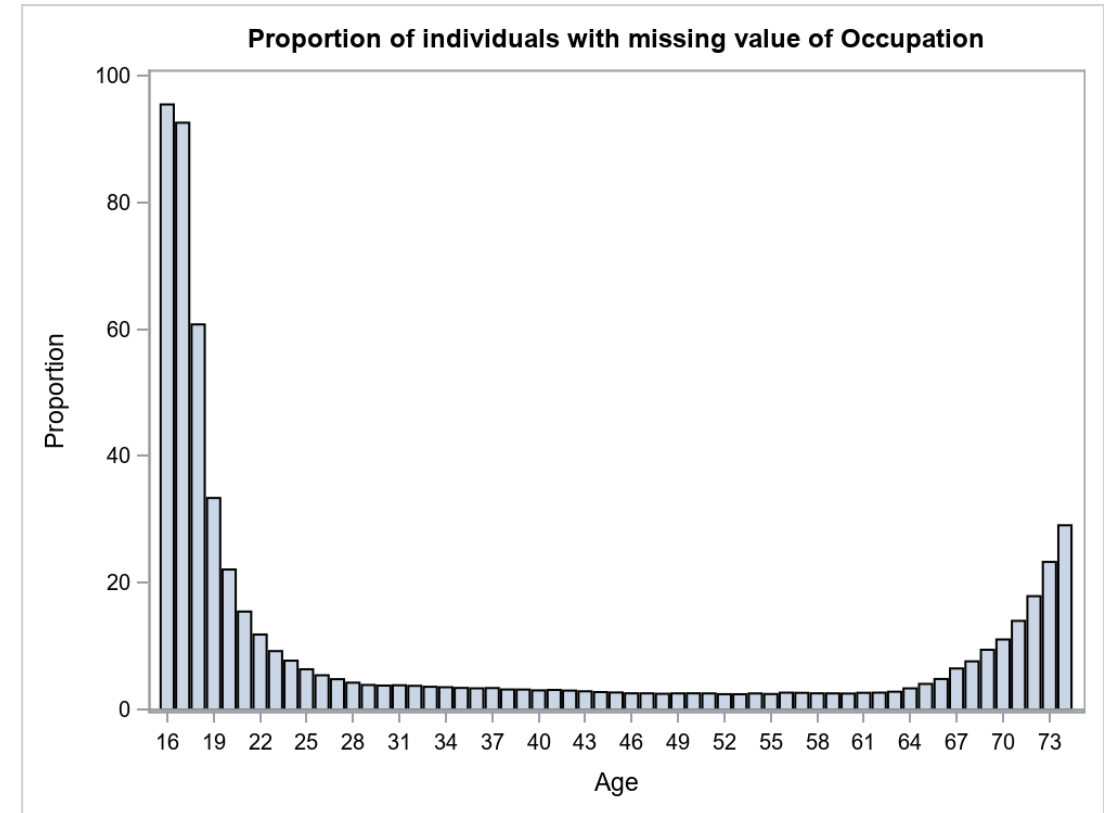
Register data

- 20+ survey sources
 - 70 % from Wage and salary structures
 - Separate survey for Occupational register
 - Intermittently collected
- Quality rating for observations
- Missing values



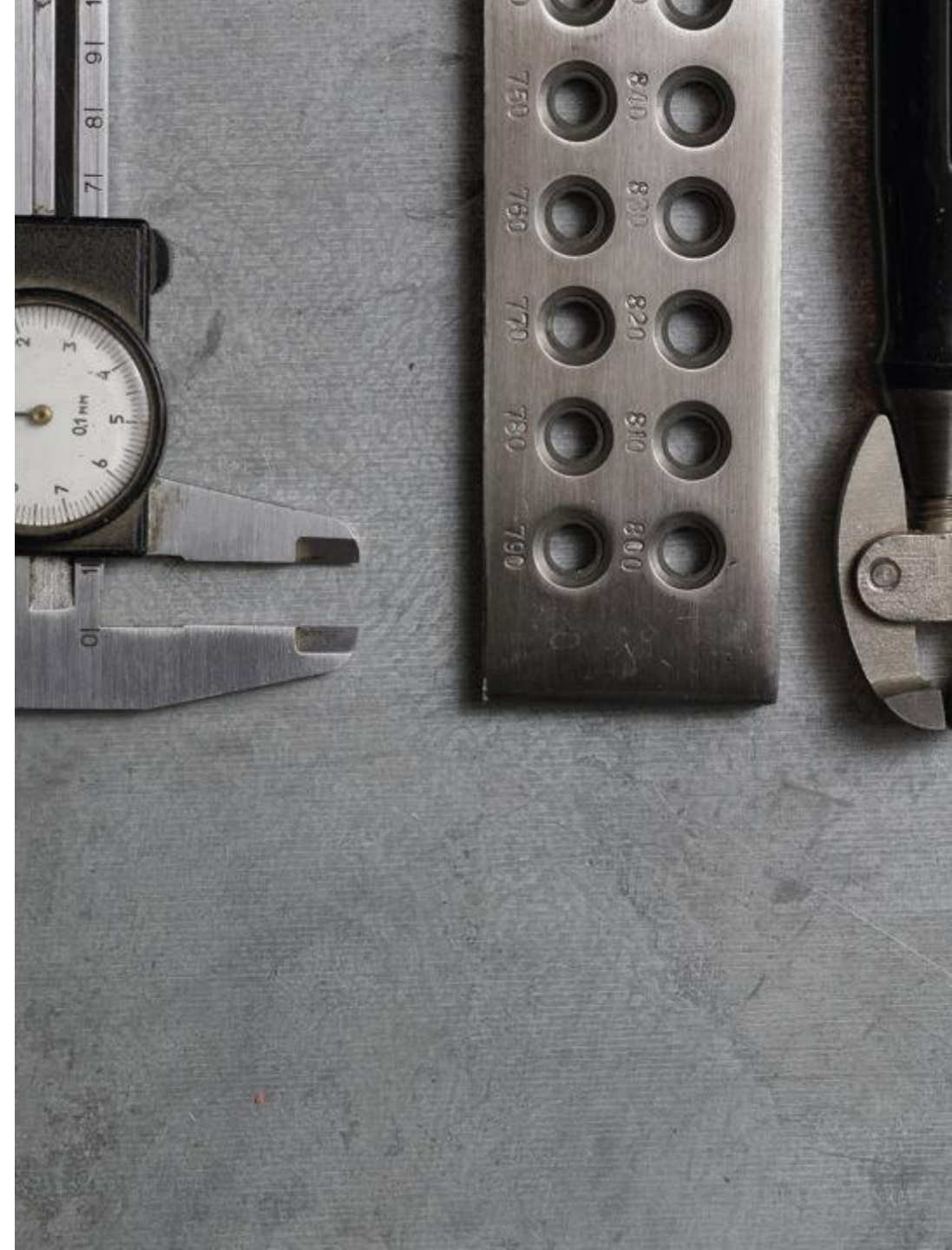
Motivation for imputation

- Missing values for 10 % of population
 - 6 % missing
 - 4 % imputed using old model
 - > SSYK96
- Population expansion to 15-74 years
- Imputation may improve the quality of the statistics



Quality

- Missing values 5-10 %
- Misclassifications 10-15 %



Methods



Methods

- Data
 - 2019 Occupational register
 - Gainfully employed 16-74 years old
 - 5 million observations
- Model
- Evaluation
 - Predictive performance
 - Distributional accuracy



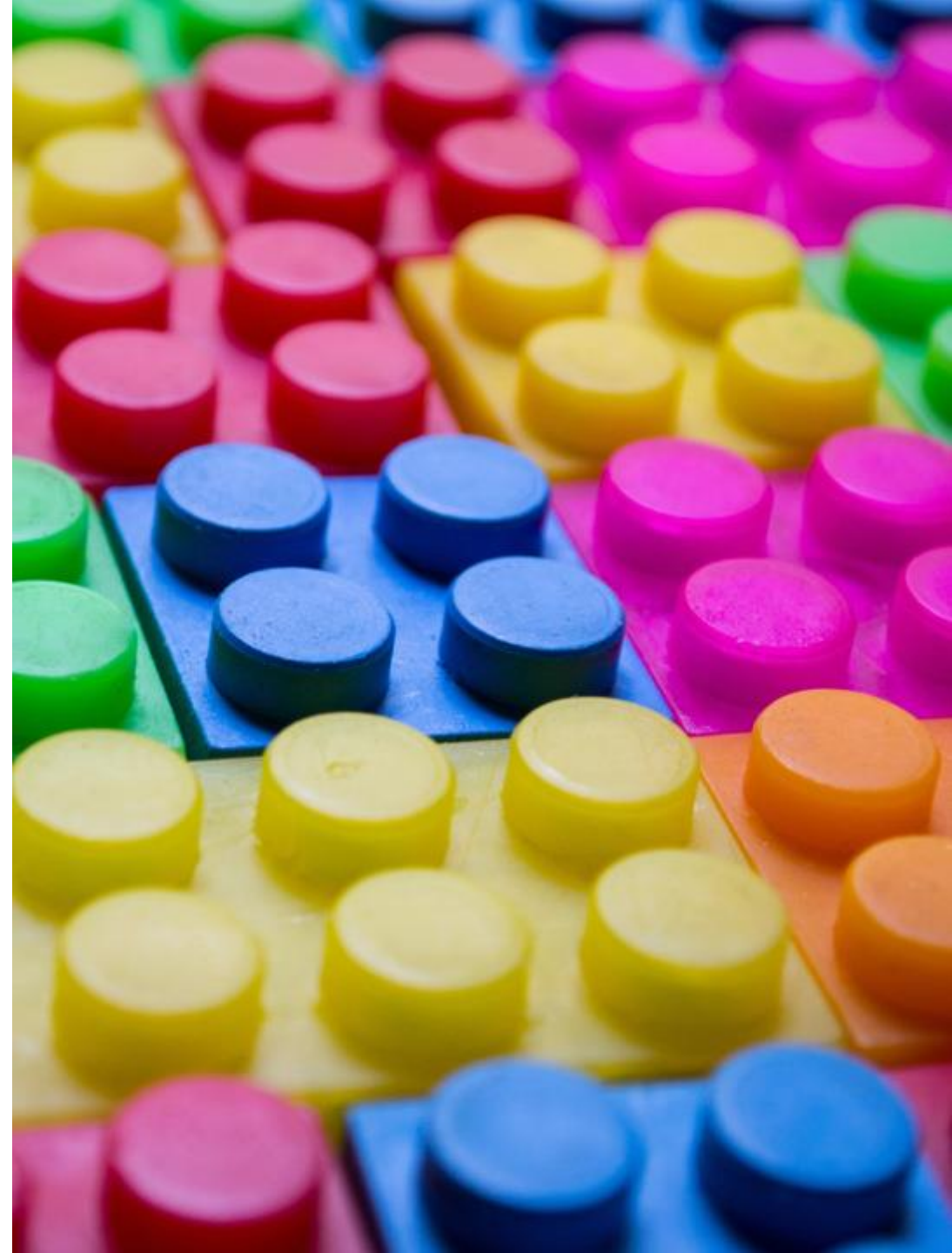
Training and evaluation

- High quality observations
 - 80 % training data
 - 20 % validation data
- Observations with missing values
 - "test data"



Misclassifications by SSYK

- Proportion of misclassifications
 - By SSYK
 - Predictive performance
 - Proportion of missing values
- Simulation of missing values
 - Validation data
 - Missing probability
 - Imputation

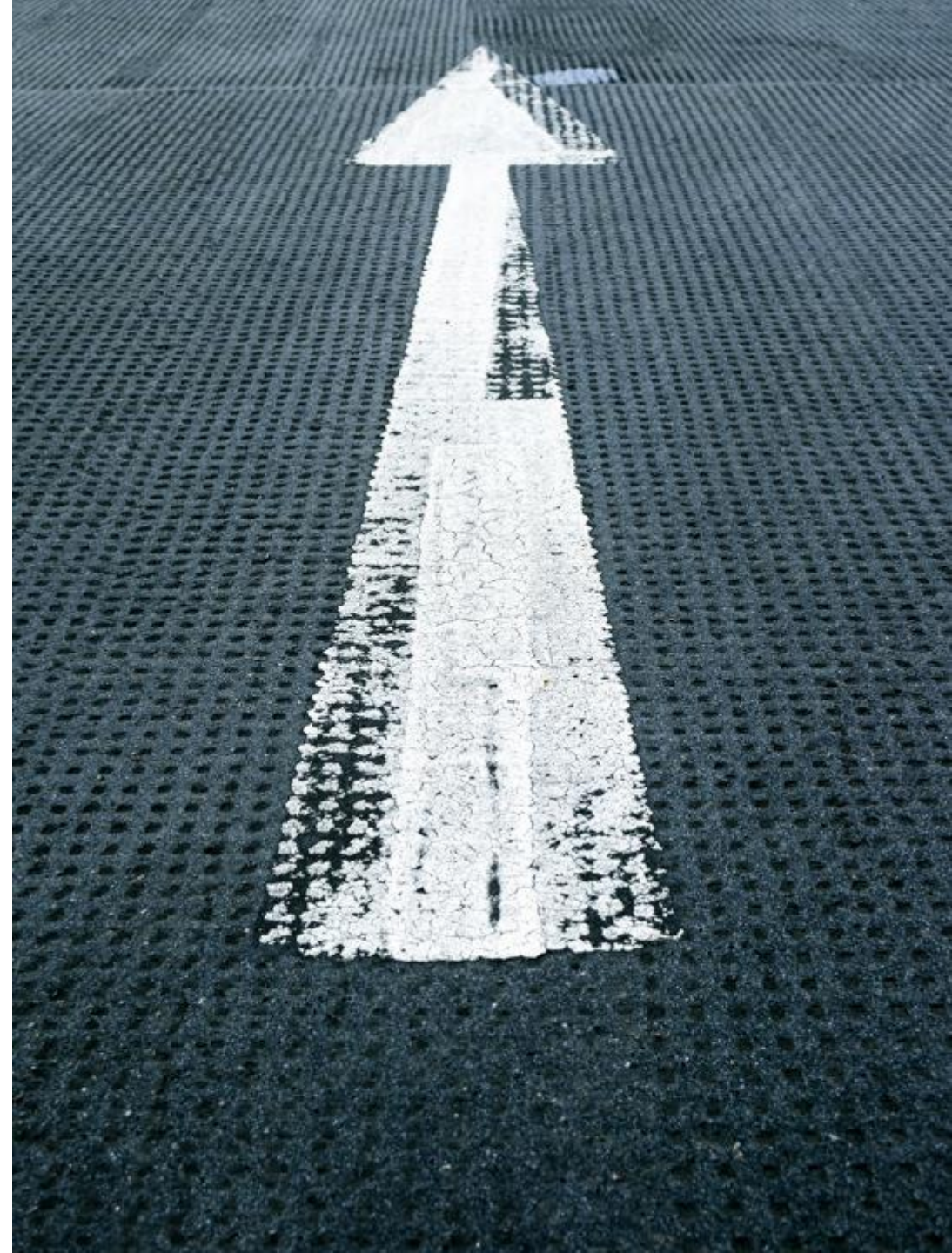
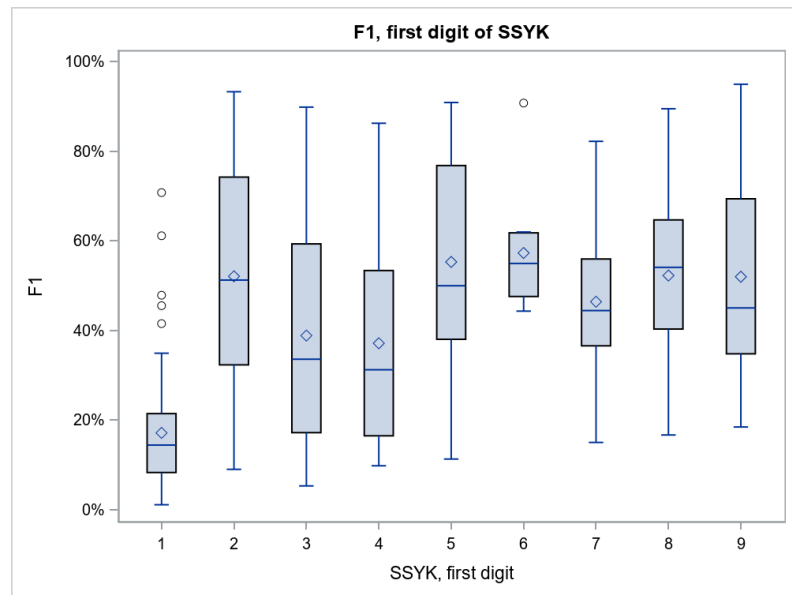


Results



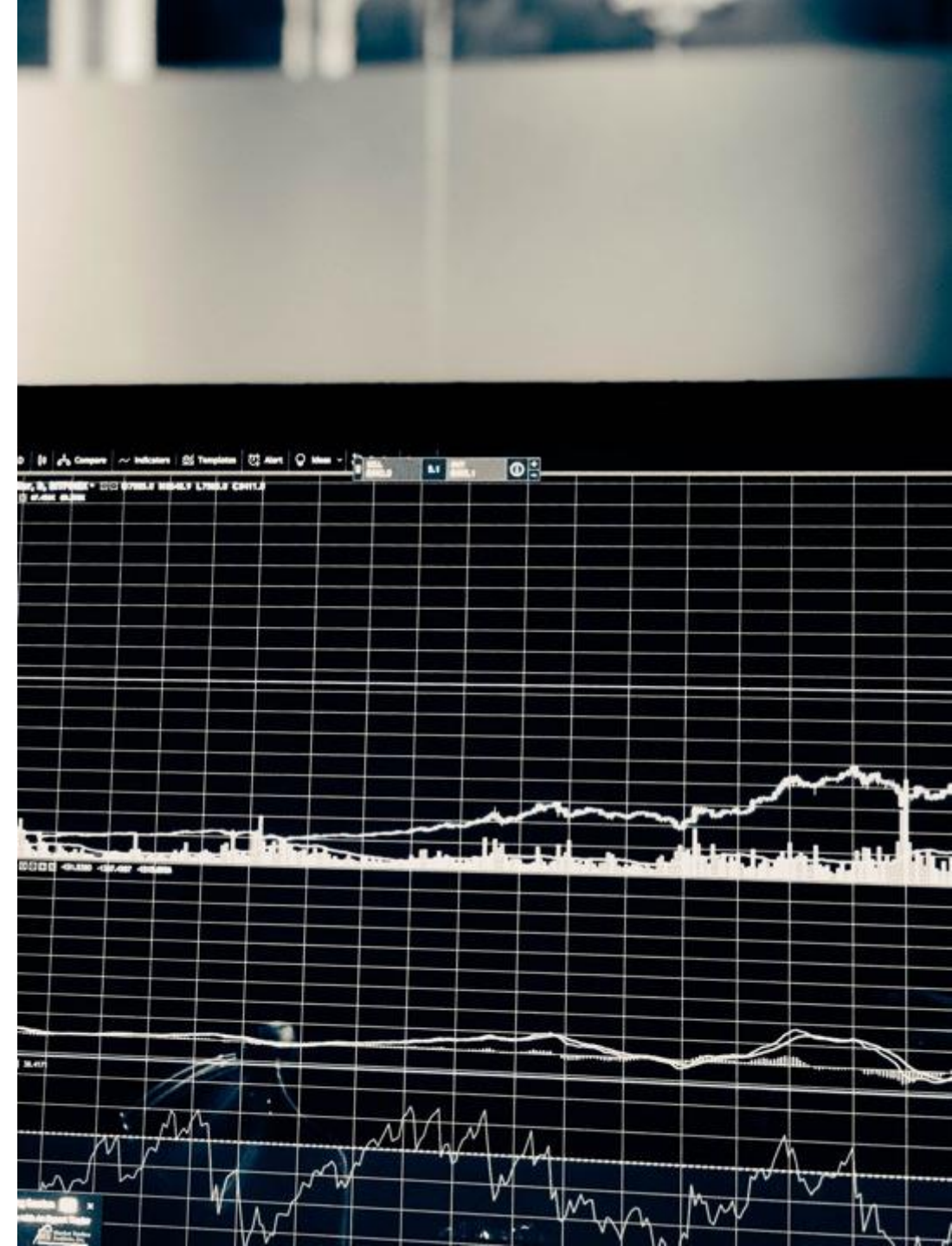
Individual predictive performance

- Accuracy
 - 44 % (54 %)



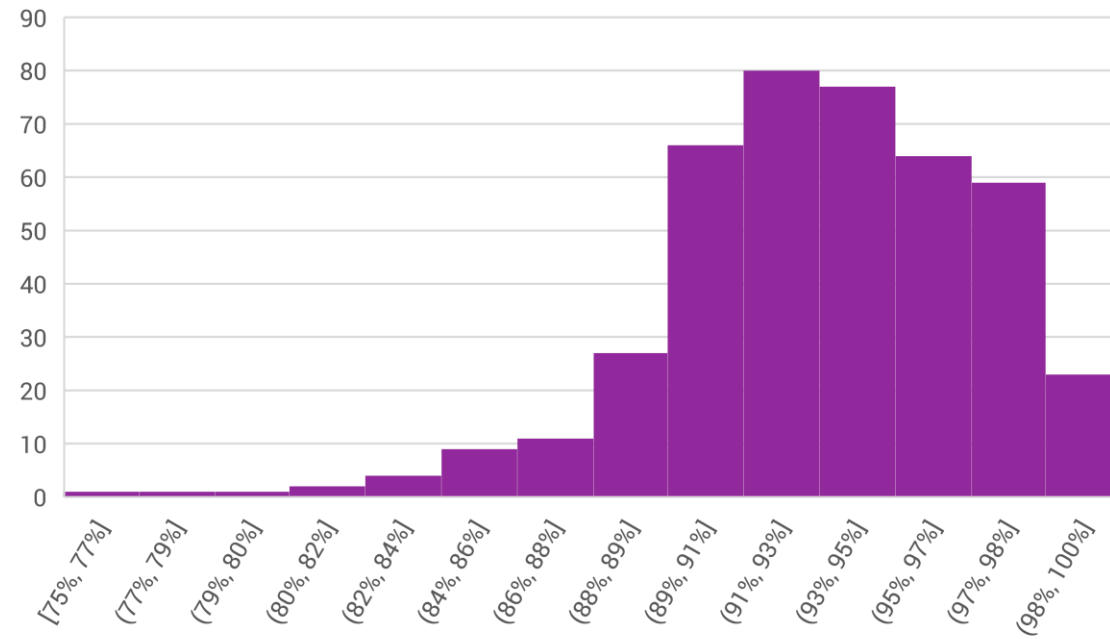
Distributional accuracy

- Imputation method
 - Majority vote
 - Sampled from predictions
- Permutation test of similarity
 - H_0 not rejected if sampled



Class predictive performance

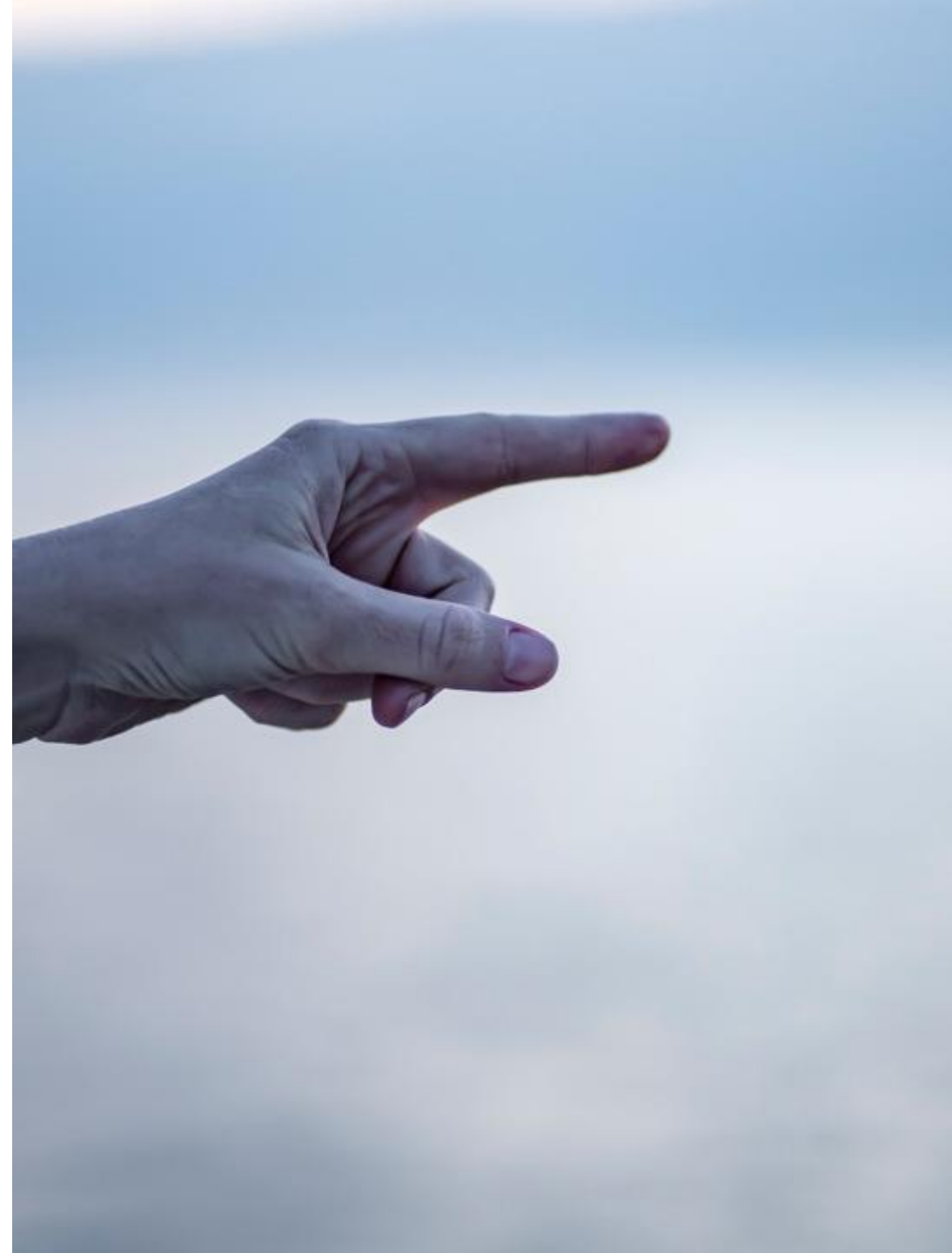
"Recall" for all SSYK



Discussion

Discussion

- Model development
 - Imbalanced data
 - Threshold on quality
- Clarify needs and quality requirements
 - Process for ML



Implementation

- Autumn 2023 (hopefully)
- Re-train model
 - Implementation
 - Yearly
- Monitoring
- Upcoming SSYK revision



Imputation of occupation in the Occupational Register

Jens Malmros

Statistics Sweden

