# Psychometrics meets machine learning

Marjolein Fokkema
Associate Professor @ Leiden University
www.marjoleinfokkema.nl
m.fokkema@fsw.leidenuniv.nl

1

Prediction

Decision making

Interpretability and inference

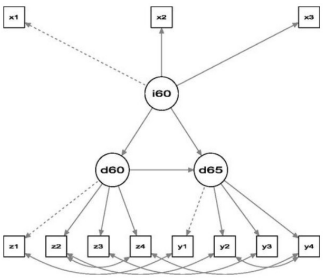
2

## Menu

- Explanation versus prediction
- Human versus actuarial decision making
- Interpretable machine learning

3

## Traditional SEM perspective (explanation)



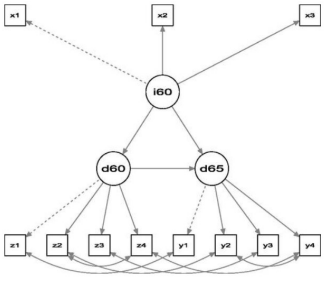

Latent variables
- Industrialization in 1960
- Political democracy in 1960
- Political democracy in 1965

Quality of the model

How well are the training sample's means and covariances reproduced?
- $\chi 2(35) = 38.1$
- CFI = 0.995
- RMSEA = 0.035

4

## Prediction perspective



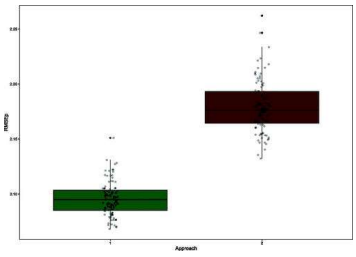

Latent variables
- Industrialization in 1960
- Political democracy in 1960
- Political democracy in 1965

Quality of the model

Given *new* observations on x1-x3 (and z1-z4), how accurate is the *prediction* of y1-y4?
Cross-validated MSE or $R^2$.

5

- SEM-based approach versus multivariate multiple linear regression



$$\text{RMSEp} = \sqrt{\frac{1}{n \times R}\sum_{i=1}^{n}\sum_{r=1}^{R}(y_{ir} - \hat{y}_{ir})^2}$$

- The linear model estimated 7 (predictors) x 4 (outcomes) = 28 regression coefficients

6

Pick a side:

a) Psychometrics should be all about prediction

b) Psychometrics should be all about measurement

7