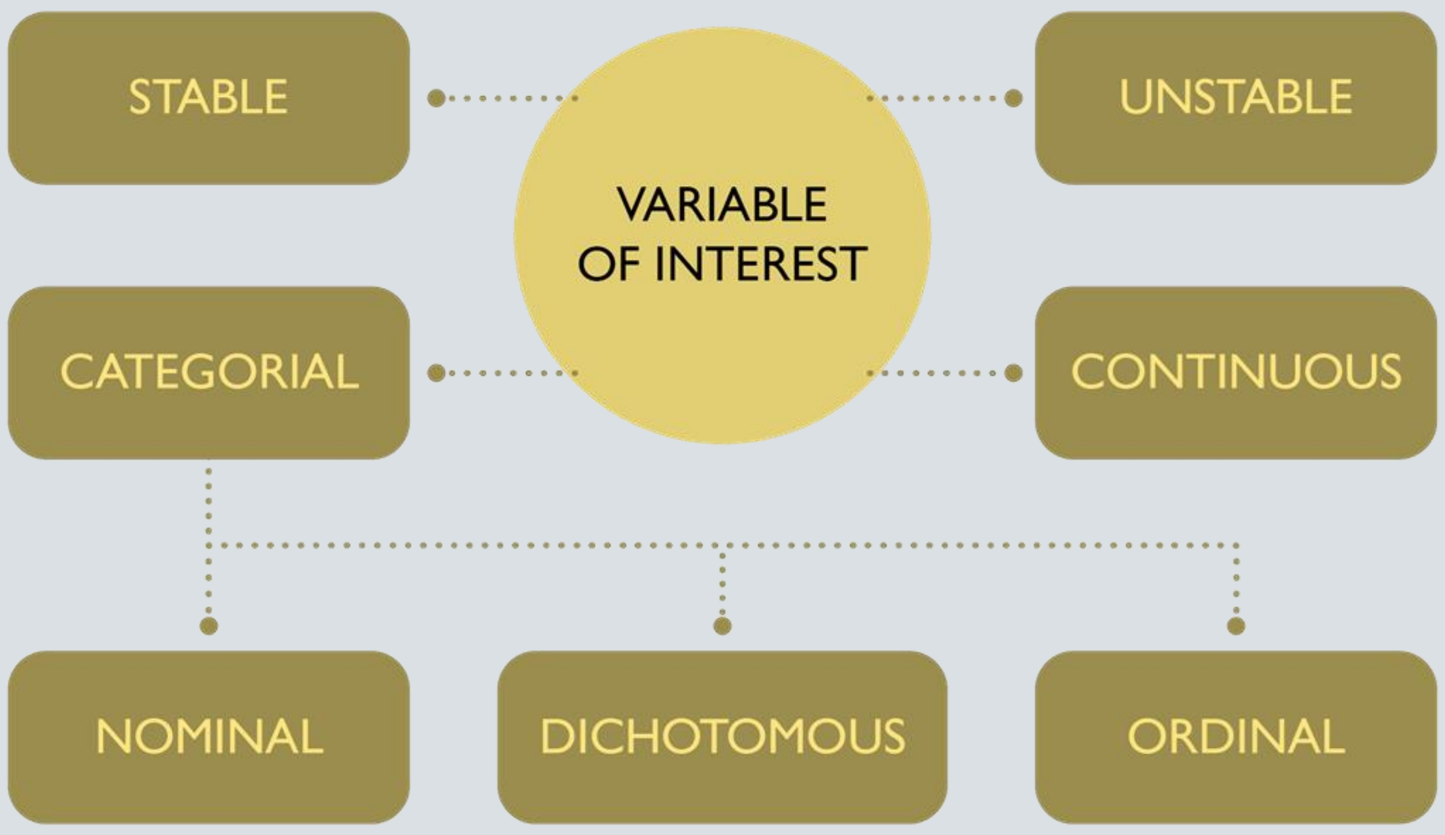
European Master in Official Statistics

Estimation of measurement errors in social survey

EMOS program @ KU Leuven

The EMOS aims at providing students with an advanced training in the specific themes of statistics in general and official statistics in particular, supported by the complementary quantitative and statistical tools offered by the hosting university. The main objective of the EMOS is to enhance the abilities of students to understand and to be able to analyse European official data at different levels: quality, production process, dissemination, and analysis in a national, European and international context.



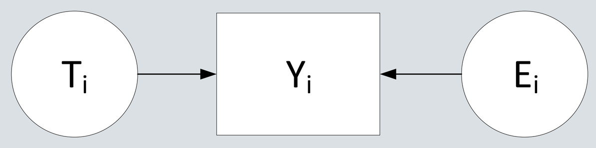
Measurement error

Measurement error is the difference between the value of a characteristic provided by the respondent and the true but unobserved value of that characteristic.

Reliability

Reliability R is defined as the proportion of the observed variance from the survey response that is accounted for by the variance of the true score.

$$R = Var(T)/Var(E)$$



Stable continuous variables

The Intraclass correlation coefficient:

$$R = Cor(Y_1, Y_2)$$

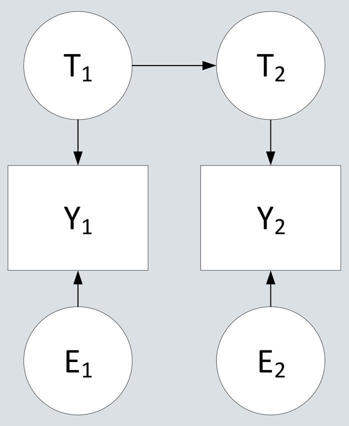
Main assumption: **T_1=T_2 or T_1= *c* + T_2 with *c* being a constant**

Advantage: The theory is simple to understand and communicate

- This method is easy to apply

Limitation: The assumption of stable true score is difficult to achieve

- The data collection cost can be high, especially for large social surveys



Stable categorical variables

Cohen's Kappa coefficient (Nominal and Dichotomous): $\kappa = \frac{P_t - P_e}{1 - P_e}$

P_t: proportion of the total number of households with unchanged responses

P_e the proportion of the households that have the same response by accident

For ordinal variables, we can add either a quadratic or linear weight to the calculation of the proportions

Advantage: The Cohen's Kappa coefficient is simple to apply

- This method is widely used and therefore easy to communicate

Limitation: The reliability estimate is biased downward if the class distribution is unbalanced

- Same as ICC (The stability assumption is difficult to achieve and the data collection cost is high)
- The choice of weighting scheme is arbitrary and can be difficult to justify

Unstable continuous variables

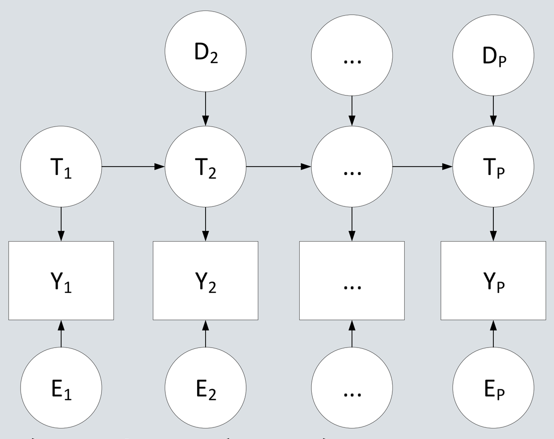
Quasi-Markov simplex model: $Y_w = T_w + E_w$ and $T_w = \beta_{w,w-1} T_{w-1} + D_w$

Advantage: The simplex model accounts for the instability of true scores

- There are different sets of assumptions for different data

Limitation: The data of at least 3 replications (waves) is required

- The data must follow the simplex structure $Cor(Y_3, Y_1) < Cor(Y_3, Y_2)$ and $Cor(Y_3, Y_1) < \mathrm{Cor}(Y_2, Y_1)$
- The assumptions can be restrictive and difficult to achieve



Unstable ordinal and dichotomous variables

- Quasi-Markov simplex model for categorical variables:
- Dichotomous variables: Tetrachoric correlation
- Nominal variables: Possible after transforming a nominal variable into multiple dichotomous variables
- Ordinal variables: Polychoric correlation

Advantage: Same advantages ad the continuous case

Disadvantage: Same limitations as the continuous case

- This method assumes an underlying continuous variable of interest being measured by these categorical variables, which may not be true

Unstable nominal variables

The Latent Markov transition model (LTM) can be used to estimate reliability by analyzing the latent class membership and its transition over time

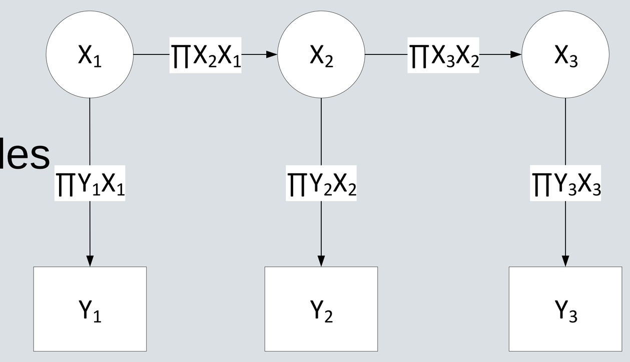
Yule's Q (Clogg & Manning, 1996): $R = Q_{Y_m X} = \frac{\theta_{Y_m X} - 1}{\theta_{Y_m X} + 1}$

Advantage: Item-response probability of each level of the nominal variables

- Directly estimate the reliability of unstable nominal variables

Limitation: This method is difficult to implement

- Can only be used to estimate reliability of large samples
- The estimate can be biased downward if the class distribution is unbalance



Type of variable		Stable variable	Unstable variable
Continuous		Intraclass correlation coefficient (1,1)	Quasi-Markov simplex model with Pearson correlation/covariance matrix
Categorical	Nominal	Cohen's Kappa coefficient OR Latent class model	Quasi-Markov simplex model with Tetrachoric correlation on dummy variables OR Latent transition model
	Dichotomous	Cohen's Kappa coefficient OR Latent class model	Quasi-Markov simplex model with Tetrachoric correlation
	Ordinal	Weighted Kappa coefficient OR Weighted Latent class model	Quasi-Markov simplex model with Polychoric correlation

Leuven Statistics Research Centre (LStat)
KU Leuven
Celestijnenlaan 200 B, 3000 Leuven
http://lstat.kuleuven.be/masterBologna/emos-track

Programme Director:
Geert.Molenberghs@kuleuven.be
Programme Coordinator:
An.Carbonez@kuleuven.be