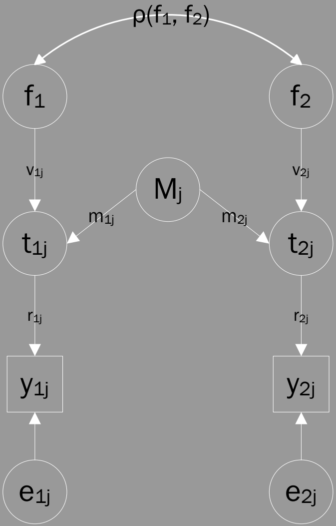
Separating random, systematic, and categorisation errors in surveys

Daniel Oberski
Willem Saris
Jacques Hagenaars

Faculty of Social and Behavioural Sciences
Tilburg University

Survey Research Centre
ESADE, Barcelona
Universitat Ramon Llull

The basic response model



f_1, f_2 = variables of interest

v_{ij} = validity coefficient for variable i

M_j = method factor for both variables

m_{ij} = method effect on variable i

t_{ij} = true score for y_{ij}

r_{ij} = reliability coefficient

y_{ij} = the observed variable

e_{ij} = the random error in variable y_{ij}

The basic response model

- The quality coefficient q is the product of the reliability and validity coefficients:
- $q = vr$
- The square q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.

The basic response model

- The quality coefficient q is the product of the reliability and validity coefficients:
- $q = vr$
- The square q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.

The basic response model

- The quality coefficient q is the product of the reliability and validity coefficients:
- $q = vr$
- The square q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.

The basic response model

- The quality coefficient q is the product of the reliability and validity coefficients:
- $q = vr$
- The square q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.

First trait measured with three methods

CARD 73 Using this card, please tell me how true each of the following statements is about your current job.

		Not at all true	A little true	Quite true	Very true	(Don't know)
G64	There is a lot of variety in my work.	1	2	3	4	8

iS19 The next 3 questions are about your current job. Please choose one of the following to describe how varied your work is.

Please tick one box.

Not at all varied ☐ 1

A little varied ☐ 2

Quite varied ☐ 3

Very varied ☐ 4

iS32 Please indicate, on a scale of 0 to 10, how varied your work is, where 0 is not at all varied and 10 is very varied.

Please tick the box that is closest to your opinion

Not at all varied										**Very varied**
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Three traits measured with first method

CARD 73 Using this card, please tell me how true each of the following statements is about your current job.

		Not at all true	A little true	Quite true	Very true	(Don't know)
G64	There is a lot of variety in my work.	1	2	3	4	8
...						
G66	My job is secure	1	2	3	4	8
...						
G70	My health or safety is at risk because of my work.	1	2	3	4	8

Three traits measured with second method

iS19 The next 3 questions are about your current job. Please choose one of the following to describe how varied your work is.
Please tick one box.

- Not at all varied ☐ 1
- A little varied ☐ 2
- Quite varied ☐ 3
- Very varied ☐ 4

iS20 Please choose one of the following to describe how secure your job is.
Please tick one box.

- Not at all secure ☐ 1
- A little secure ☐ 2
- Quite secure ☐ 3
- Very secure ☐ 4

iS21 Please choose one of the following to say how much, if at all, your work puts your health and safety at risk.
Please tick one box.

- Not at all at risk ☐ 1
- A little at risk ☐ 2
- Quite a lot at risk ☐ 3
- Very much at risk ☐ 4

Three traits measured with third method

Skip details of the model

iS32 Please indicate, on a scale of 0 to 10, how varied your work is, where 0 is not at all varied and 10 is very varied.
Please tick the box that is closest to your opinion

Not at all varied — ***Very varied***

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

iS33 Now please indicate, on a scale of 0 to 10, how secure your job is, where 0 is not at all secure and 10 is very secure.
Please tick the box that is closest to your opinion

Not at all secure — ***Very secure***

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

iS34 Please indicate, on a scale of 0 to 10, how much your health and safety is at risk from your work, where 0 is not at all at risk and 10 is very much at risk.
Please tick the box that is closest to your opinion

Not at all at risk — ***Very much at risk***

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Different models for MTMM experiments

- Classic MTMM model
- Correlated uniqueness (Kenny & Judd)
- Direct product (Browne)
- True score model
- MTM-1 (Eid 2000)

Different models for MTMM experiments

- Classic MTMM model
- Correlated uniqueness (Kenny & Judd)
- Direct product (Browne)
- True score model
- MTM-1 (Eid 2000)

Different models for MTMM experiments

- Classic MTMM model
- Correlated uniqueness (Kenny & Judd)
- Direct product (Browne)
- True score model
- MTM-1 (Eid 2000)

Different models for MTMM experiments

- Classic MTMM model
- Correlated uniqueness (Kenny & Judd)
- Direct product (Browne)
- True score model
- MTM-1 (Eid 2000)

Different models for MTMM experiments

- Classic MTMM model
- Correlated uniqueness (Kenny & Judd)
- Direct product (Browne)
- True score model
- MTM-1 (Eid 2000)

Different models for MTMM experiments

- We use the true score model
- Equivalent to the classic MTMM model
- Sometimes necessary to remove one method factor
- In that case our model is the equivalent to Eid's MTM-1 model.

Different models for MTMM experiments

- We use the true score model
- Equivalent to the classic MTMM model
- Sometimes necessary to remove one method factor
- In that case our model is the equivalent to Eid's MTM-1 model.

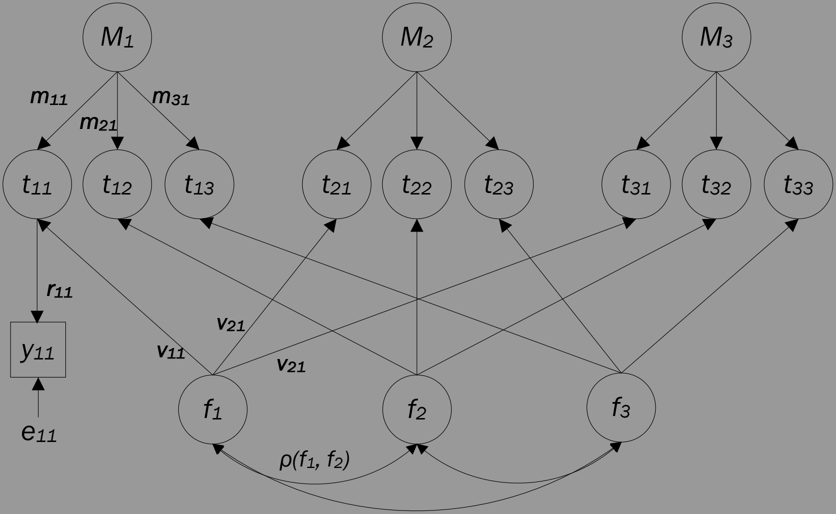
Different models for MTMM experiments

- We use the true score model
- Equivalent to the classic MTMM model
- Sometimes necessary to remove one method factor
- In that case our model is the equivalent to Eid's MTM-1 model.

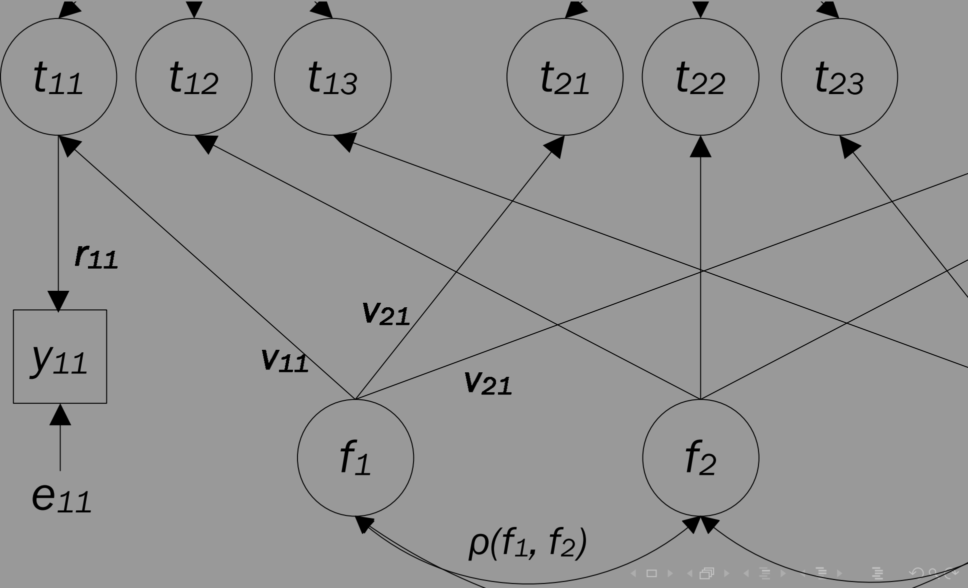
Different models for MTMM experiments

- We use the true score model
- Equivalent to the classic MTMM model
- Sometimes necessary to remove one method factor
- In that case our model is the equivalent to Eid's MTM-1 model.

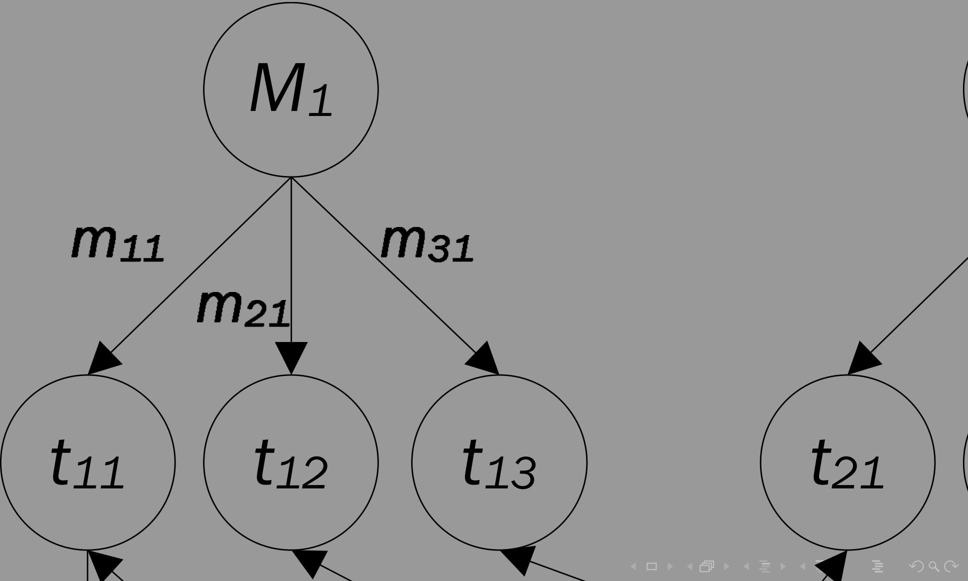
True score model



True score model



True score model



True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

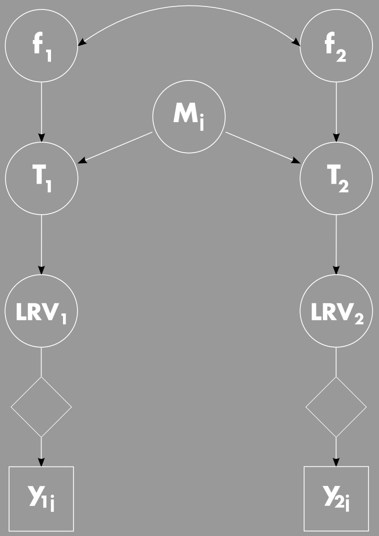
True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

True score model assumptions

- No correlations among methods
- No correlations between traits and methods
- Equal method effects
- Linear and additive effects
- Normal errors, independent of all unobserved variables
- All variables are continuous

The basic response model, revised



The 'categorisation factor'

- The quality was defined as:
- $$q^2 = \frac{Var(f)}{Var(y)}.$$
- However, we have seen that y is itself a categorization of an unobserved continuous variable (c), and therefore the above equation can be 'decomposed' into
- $$q^2 = \frac{Var(f)}{Var(LRV)} \cdot \frac{Var(LRV)}{Var(y)}.$$
- We call this ratio of the quality coefficient $q = v.r$ from the categorical model to the same coefficient from the continuous model the 'categorisation factor'.

The 'categorisation factor'

$$q^2 = \frac{Var(f)}{Var(LRV)} \cdot \frac{Var(LRV)}{Var(y)}. \tag{1}$$

- It can be seen that the quality normally estimated from the continuous model is a product of two terms:
- $$q^2_{cont} = q^2_{categ} \cdot c,$$

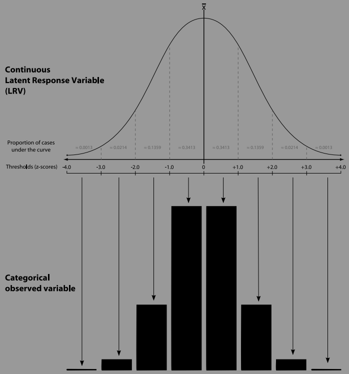
  where c is a categorisation factor.
- If $c < 1$, the quality in the categorical model is higher than in the continuous model
- If $c > 1$, the quality in the categorical model is lower than in the continuous model

Categorisation of continuous variables

Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.

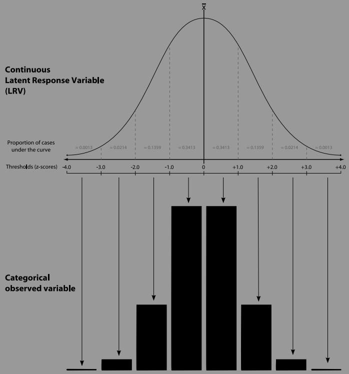
Categorisation of continuous variables

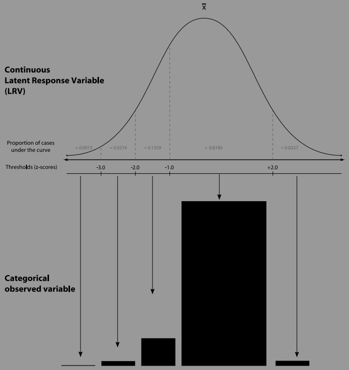
Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.



Categorisation of continuous variables

Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.





Consequences of categorisation for the correlations between observed variables

- The fewer categories, the smaller the Pearson correlation;
- The more skew in observed variables, the smaller the Pearson correlation;
- The corrected ('polychoric') correlations are always higher than the Pearson correlations, but not necessarily equally so for all variables.

Consequences of categorisation for the correlations between observed variables

- The fewer categories, the smaller the Pearson correlation;
- The more skew in observed variables, the smaller the Pearson correlation;
- The corrected ('polychoric') correlations are always higher than the Pearson correlations, but not necessarily equally so for all variables.

Consequences of categorisation for the correlations between observed variables

- The fewer categories, the smaller the Pearson correlation;
- The more skew in observed variables, the smaller the Pearson correlation;
- The corrected ('polychoric') correlations are always higher than the Pearson correlations, but not necessarily equally so for all variables.

Therefore,

- If the skewness of observed variables is higher for variables measured by one particular method, then the corrected correlations between those variables will go up more than the others, and the method effects in the categorical model will be higher;
- As method-induced correlation goes up, the estimates of the quality will go down instead of up!

Therefore,

- If the skewness of observed variables is higher for variables measured by one particular method, then the corrected correlations between those variables will go up more than the others, and the method effects in the categorical model will be higher;
- As method-induced correlation goes up, the estimates of the quality will go down instead of up!

Analysis of the experiments

- We analysed the 4 experiments from the ESS which involved variables with 5 categories or less
- The topics: role of women, GP's, political efficacy, job.



- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Analysis of the experiments

- We analysed the 4 experiments from the ESS which involved variables with 5 categories or less
- The topics: role of women, GP's, political efficacy, job.



- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Analysis of the experiments

- We analysed the 4 experiments from the ESS which involved variables with 5 categories or less
- The topics: role of women, GP's, political efficacy, job.



- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Quality (q^2)and method effects (m) in the efficacy experiment, in Denmark

Results of continuous MTMM model, main questionnaire (first method)

	'Efficacy'		
	Complex	Active	Mind
q^2	0.77	0.83	0.79
m	0.00	0.00	0.00

$df = 19, \chi^2 = 40.0, p = 0.003.$

Efficacy experiment: Denmark

Polychoric correlations

		Method 1			Method 2		
Method 1	Complex	1.00					
	Active	-0.44	1.00				
	Mind	-0.51	0.47	1.00			
Method 2	Complex	0.66	-0.45	-0.51	1.00		
	Active	-0.44	0.74	0.46	-0.51	1.00	
	Mind	-0.52	0.51	0.67	-0.56	0.56	1.00

Pearson correlations

		Method 1			Method 2		
Method 1	Complex	1.00					
	Active	-0.38	1.00				
	Mind	-0.46	0.41	1.00			
Method 2	Complex	0.60	-0.37	-0.44	1.00		
	Active	-0.39	0.67	0.40	-0.43	1.00	
	Mind	-0.46	0.43	0.62	-0.49	0.48	1.00

$n \approx 916$

% Increase in the correlations after correction for categorisation

Efficacy experiment: Denmark

		Method 1			Method 2	
Method 1	Complex					
	Active	**16**%				
	Mind	**11**%	**15**%			
Method 2	Complex	10%	22%	16%		
	Active	13%	10%	15%	**19**%	
	Mind	13%	19%	8%	**14**%	**17**%

Mean percentage increase of the polychoric correlations: 14.5%

Quality (q^2)and method effects (m) according to the continuous and categorical models, with categorisation factors

	'Efficacy'		
	Complex	Active	Mind
Continuous analysis			
q^2	0.77	0.83	0.79
m	0.00	0.00	0.00
Categorical analysis			
q^2	0.63	0.70	0.63
m	0.11	0.08	0.11
Categorisation factor			
	1.23	1.18	1.25

Correction for categorisation: conclusions

- The general 'push' is that all coefficients go up, because the polychoric correlations are always higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is estimated at zero in the continuous model, but cannot be so constrained in the categorical model.

Correction for categorisation: conclusions

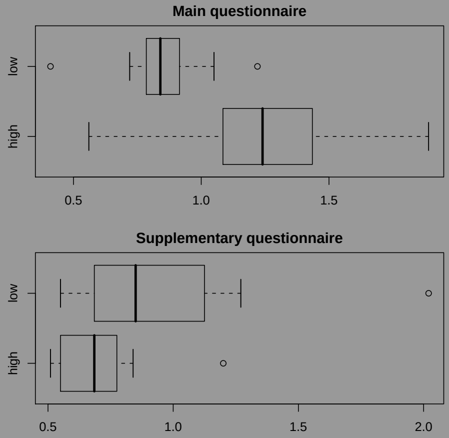
- The general 'push' is that all coefficients go up, because the polychoric correlations are always higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is estimated at zero in the continuous model, but cannot be so constrained in the categorical model.

Correction for categorisation: conclusions

- The general ‘push’ is that all coefficients go up, because the polychoric correlations are always higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is estimated at zero in the continuous model, but cannot be so constrained in the categorical model.

A meta-analysis of the categorisation error studies

The categorisation factor, q_{cat}/q_{cont}:



How general are the findings?

	Estimate	S.E.	95% C.I lower	95% C.I upper
(Intercept)	1.04	0.36	0.31	1.77
Topic				
Doctors	(reference category)			
Efficacy	0.06	0.10	-0.14	0.27
Job	0.04	0.40	-0.71	0.78
Women	0.38	0.26	-0.14	0.90
Scale				
Direct	(reference category)			
Agree-disagree	-0.11	0.35	-0.81	0.59
True-false	0.17	0.32	-0.48	0.81
Negative	-0.50	0.23	-0.96	-0.02
Main questionnaire	-0.30	0.29	-0.88	0.29
Highest quality	-0.19	0.09	-0.37	-0.01
Highest quality × main	0.66	0.15	0.35	0.96

Multiple R-Squared: 0.45; Adjusted R-squared: 0.35

Some implications of the findings

- Potentially, if one method produces more categorisation errors than another, the quality coefficients may be estimated higher in the continuous model.
- If this happens more in some countries than others, differences in quality will result due to the way the LRV's have been categorised.

The latent traits

Estimated correlations between the latent traits under the two different models

	Complex	Active	Mind
Continuous model	1		
	-.625	1	
	-.754	.663	1
Categorical model	1		
	-.632	1	
	-.752	.696	1

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error in this experiment were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this case.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

That's it for now. Moltes gràcies per la seva atenció!

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. http://www.sqp.nl