The quality of survey questions

Cinzia Meraviglia
University of Eastern Piedmont, Italy

Department of Sociology, KTU
14th March 2013

Overview

1. Measurement error, from Saris and Gallhofer (2007) and the work of RECSM (http://www.upf.edu/survey/)
   - Validity, reliability, quality of questions and method effect
   - Why to correct for systematic error, with some examples
2. How to correct for systematic error using SQP 2.0 (www.sqp.nl)
   - An example from ISSP 2009 UK

Questions and answers

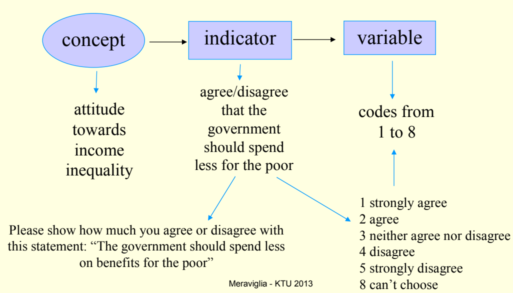
- A questionnaire is formed by a number of questions, which are pre-defined as for both the question wording and the response options
- Good questions (and answers) bring quality data, bad questions (and answers) bring poor data
- Can we always trust the data we get?
  - Are they free from error?
  - More specifically, are they free from systematic measurement error?

Types of measurement error

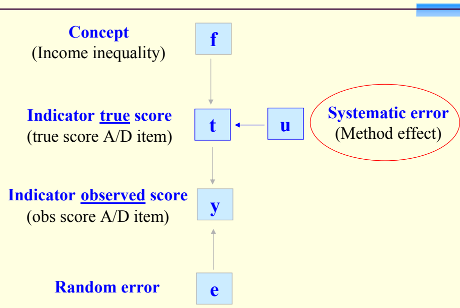
- Imagine we want to measure attitudes towards income inequality using an A/D (agree/disagree) item like this:
  - *Please show how much you agree or disagree with this statement: "The government should spend less on benefits for the poor"*
  - Answers = 1 strongly agree/ 2 agree/ 3 neither agree nor disagree/ 4 disagree/ 5 strongly disagree/ 8 can't choose
- Random error: eg. a R choses 2 instead of 1, or 3 instead of 4
- Systematic error: eg. the A/D response scale biases the "true" score of Rs on the measured trait (= the opinion on the government duties towards the poor) by eliciting a higher level of agreement
- Note that:
  - All variables are always affected by measurement error
  - We can (relatively safely) disregard random error, but we have to deal with systematic error, because it affects the magnitude of the estimated relationship between variables (either by lowering or increasing it)

Systematic error

- It is produced in this process:

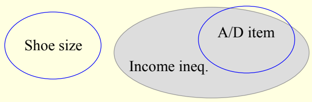


A measurement model
(Saris and Gallhofer 2007)



Validity

- v: refers to the degree in which our indicator (t) is related to the concept of interest (f)
- Example:
  - Concept = attitude towards income inequality
  - Valid indicator = agree/disagree that the government should spend less for the poor (93% valid)
  - Non-valid indicator = R's shoes size (0% valid)

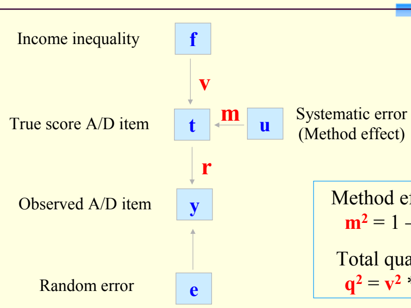


Reliability

- r: refers to the degree in which the observed indicator (y) is close to the true indicator (t)
- Reliability is influenced by many features of the question/answer:
  - Number of response categories (eg. 5, 7, 11)
  - Labelling of categories (full/partial/none)
  - Ordering of categories (from higher to lower, or vice versa)
  - Position of the question in the questionnaire
  - …

A measurement model with v and r

(adapted from Saris and Gallhofer 2007)

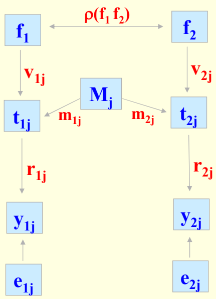


Method effect:
$m^2 = 1 - v^2$

Total quality:
$q^2 = v^2 * r^2$

A model for 2 concepts and 1 method

(Saris and Gallhofer 2007, 187)



$\rho(f_1\ f_2)$= correlation between the variables of interest

v_{ij} = validity ocefficient

M_j = method factor for both variables
m_{ij} = method effect on variable i
t_{ij} = true score

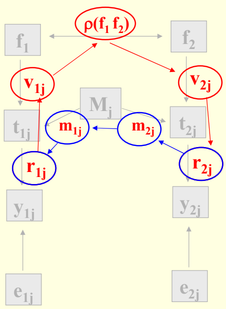
r_{ij} = reliability coefficient

y_{ij} = observed variable

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

Correlation between obs. variables

(Saris and Gallhofer 2007, 188)



Correlation between obs indicators equals the actual correlation between concepts ρ(f1,f2) weighted by vs and rs, increased/decreased by the method effects weighted by rs:

$$\rho(y_1\ y_2) = r_{1j}\ v_{1j}\ \rho(f_1,f_2) v_{2j}\ r_{2j} +$$
$$+\ r_{1j}\ m_{1j}\ m_{2j}\ r_{2j}$$

Why to correct for systematic error

- Correlations between variables may be due to:
  - Substantive relationships among them, or
  - A non-substantive, method-driven effect
  - Usually, to a combination of the above
- What if we give substantive meaning to a correlation that has none?
- An example from Saris and Gallhofer (2007, ch. 14) on political efficacy (let's say, concept 0), which is formed by:
  - Concept 1: subjective competence (3 items, "Subj")
  - Concept 2: perceived system responsiveness (2 items, "Res")

Political efficacy

1. Subjective competence:
   - *Subj1: "Sometimes politics and government seem so complicated that I can't really understand what is going on"*
   - Subj2: "I think I can take an active role in a group that is focused on political issues"
   - *Subj3: "I understand and judge important political questions very well"*
2. Perceived system responsiveness:
   - *Res1: "Politicians do not care much about what people like me think"*
   - Res2: "Politicians are only interested in people's votes but not in their opinions"

- Answers = 5 points A/D scale

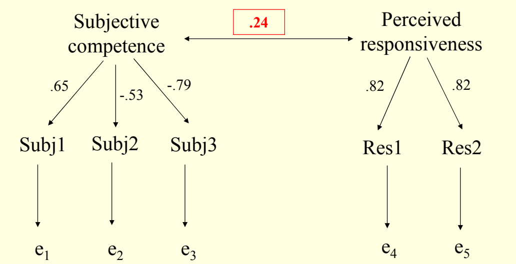
Correlations

	Subj1	Subj2	Subj3	Res1	Res2
Subj1	1.00				
Subj2	-.33	1.00			
Subj3	-.52	.43	1.00		
Res1	.16	-.13	-.13	1.00	
Res2	.12	-.18	-.13	.68	1.00

About the correlation matrix

- The "subj" variables seem to be correlated more with one another than with the "res" variables
- We suppose this is because each set of variables measures a different concept:
  - "Subj" variables → Subjective competence
  - "Res" variables → Perceived responsiveness

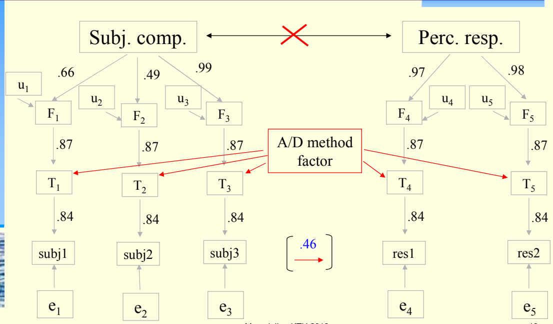
A perfectly meaningful model



Conclusion 1

- After estimating a factor model, we see that our hypothesis is met:
  - The “subj” variables correlate to a concept, the “res” variables correlate to the other
  - The correlation between the two concepts is .24, meaning that knowing something on the first set of variables tells us something about the other two, and vice versa
  - The fact that the two concepts are correlated confirms the hypothesis that they are conceptual dimensions of political efficacy

A more accurate model



Conclusion 2 (the right one)

- The correlation between the observed variables is due to a non-substantive effect, ie. the method effect of the A/D scale
- The two concepts are NOT correlated: the method effect accounts for all the correlation between them (see first model)
- Were we to rely on the meaningful but not accurate model, we would draw the wrong conclusion about these two concepts being conceptual dimensions of the political efficacy

How to correct for systematic error

1. Structural equation modelling with multiple indicators
   - Requires knowledge of the theory, the use of Lisrel, and collection of suitable data
2. Multi-trait multi-method experiments (Campbell and Fiske 1959; Andrews 1984; Saris and Andrews 1991)
   - Same
3. SQP 2.0 (Oberski 2011)
   - Online, free of charge, simple, effective, useful!
   - Requires coding of questions according to 60 criteria (see Saris and Gallhofer 2007, chapter 13)
   - Allows to correct observed correlations for systematic measurement error due to method effects, with a bit of elementary algebra

SQP 2.0

- www.sqp.nl , see Oberski, Gruner and Saris (2011)
- A software based on the theory by Saris and Gallhofer (2007)
- Its data basis comes from the MTMM experiments run in the ESS 1-3 + questions studied in the past (3000 questions in total)
- Many languages: English (IR, UK, USA), German (A, CH, D), Dutch (B, NL), French (F, L), Danish, Finnish, Portuguese, Norwegian, Turkey, Polish, Greek, Estonian, Italian, Czech, Spanish, Slovak, Slovenian, Ukrainian

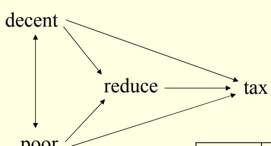
When to use it

- Two main cases:
  1. **Data analysis**: to correct the observed correlations for systematic error
     a) Either in the case of ESS questions,
     b) Or in the case of questions of your own survey
  2. **Research design:** to design a new question(naire) and to know which wording/answer options bring the highest quality, and which improvements can be done, again:
     a) Either in the case of ESS questions,
     b) Or in the case of questions in your own survey

Example (ISSP 2009 Social Inequality, GB)

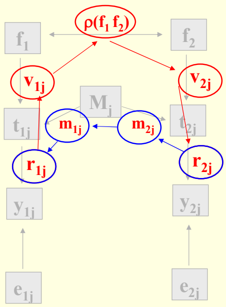
- See questionnaire
- Research question:
  - How the opinion on the role of the government concerning income inequality affects the opinion on taxation for those with high incomes?
- Concepts and indicators:
  - Taxation: Q7 “tax”
  - Role of government: Q6a “reduce”, Q6b “decent”, Q6c “poor”

Model and correlations without systematic error



	reduce	decent	poor	tax
reduce	1			
decent	0.38	1		
poor	-0.17	-0.401	1	
tax	0.34	0.171	-0.085	1

Model with systematic error (any 2 items)



Recall that the correlation between obs. variables is:

$\rho(y_1,y_2) = r_{1j}\, v_{1j}\, \rho(f_1,f_2) v_{2j}\, r_{2j}\, +$
$+\, r_{1j}\, m_{1j}\, m_{2j}\, r_{2j}$

We get:

- $\rho(y_1,y_2)$ from the data
- r, v and m coefficients from SQP

Then we can calculate $\rho(f_1,f_2)$, which is what we are interested into

Let's go to .SQP.nl and see the coded questions

- www.sqp.nl
  - See Oberski, Gruner and Saris (2011) for reference on SQP 2.0; Saris and Gallhofer (2007) use an earlier version of SQP
- The 4 items have been coded according to the 60 characteristics
  - See Saris and Gallhofer (2007), ch. 12 par. 2
- We can get the predicted coefficients and put them in a table (see slide 27)

Example: the "reduce" item

The SQP output for the 4 items

	reduce	decent	poor	tax
r^2	0.689	0.690	0.689	0.609
v^2	0.933	0.936	0.934	0.913
q^2	0.643	0.646	0.644	0.556
$CMV_{i.}$	0.046	0.044	0.045	0.053

Let's correct for systematic error

- Now that we have the validity and reliability coefficients, we can use them for correcting the observed correlations for systematic measurement error
- We will then compare the observed and the corrected correlations in order to:
  - Assess how the method effect influences the coefficients
  - Understand how bad we can go if we draw our conclusions not correcting our observed correlations

Recall that…

- CMV = common method variance = the % of the variance of the obs variables explained by the method effect = $\sqrt{r_{1j}^2\, m_{1j}^2\, m_{2j}^2\, r_{2j}^2}$
- Quality = $q^2 = r^2 * v^2$
- From this equation:

$$\rho(y_1,y_2) = r_{1j}\, v_{1j}\, \rho(f_1,f_2)\, v_{2j}\, r_{2j} + r_{1j}\, m_{1j}\, m_{2j}\, r_{2j}$$

- With a bit of algebra we get:

$$\rho(f_1,f_2) = \frac{\rho(y_1,y_2) - CMV_{12}}{q_1^2 q_2^2}$$

The reduce-decent correlation

- Using the last equation, we calculate the actual correlation between the concepts:

$$\rho(f_1,f_2)=\frac{\rho(y_1,y_2)-CMV_{12}}{\sqrt{q_1^2q_2^2}}=\frac{.38-.045}{\sqrt{.643*.646}}=.520$$

- Then we see that the observed correlation is deflated by measurement error:

$\rho(y_1,y_2) = .38$

$\rho(f_1,f_2) = .52$

	reduce	decent
$\rho(y_1,y_2)$	.38	
r^2	0.689	0.690
v^2	0.933	0.936
q^2	0.643	0.646
CMV_i	0.046	0.044

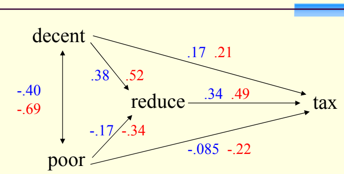
Observed vs corrected correlations

Observed				
	reduce	decent	poor	tax
Red.	1			
Dec.	0.38	1		
Poor	-0.17	-0.401	1	
Tax	0.34	0.171	-0.085	1

Corrected				
	reduce	decent	poor	tax
Red.	1			
Dec.	.520	1		
Poor	-.336	-.693	1	
Tax	.492	.209	-.219	1

- Measurement error deflates all correlations
- The “poor” item is severely affected by systematic error
- Conclusions based on the obs correlations would be biased!

Our model



- Blue coefficients = observed correlations
- Red coefficients = corrected correlations

Conclusions

- Measurement error is always there
- It affects our measures of the relationships between variables, even the simplest ones (eg. correlation coefficients)
- There is a simple and effective way to correct for measurement error, ie. using coefficients predicted by SQP
- Let's use it!

References

- Andrews F.M. (1984), Construct validity and error components of survey measures: A structural equation approach, *Public Opinion Quarterly*, 48, 409-442
- Campbell D.T. and Fiske D.W. (1959), Convergent and discriminant validationby multitrait-multimethod matrices, *Psychological Bulletin*, 56, 81-105
- Goldthorpe J.H. (2007), *On sociology*, vol. 1, Stanford, Stanford University Press
- Oberski D., Gruner T. and Saris W.E. (2011), The program SQP 2.0 for prediction of quality of questions and its applications, in W.E.Saris et al. (2011), *Final report about the project JRA3 as part of the ESS infrastructure*, RECSM Working Paper n. 24, Universidada Pompeu Fabra, Barcelona
- Saris W.E. and Andrews F.M. (1991), Evaluation of measurement instruments using a structural modeling approach, in P.P.Bleimer, R.M.Groves, L.E.Lyberg, N. Mathiowetz and S.Sudman (eds), *Measurement errors in survey*, New York, Wiley, 575-599
- Saris W.E. and Gallhofer I. (2007), *Design, evaluation and analysis of questionnaires for survey research*, Hoboken Wiley