Constructing scales from survey questions

Tomasz Żółtak

2.07.2018

Why to make scales?

Many questions regarding the same topic (1)

European Social Survey - Round 8 (2016):

I am now going to ask you about the effect of social benefits and services on different areas of life in [country]. By social benefits and services we are thinking about things like health care, pensions and social security.

(. . .), please tell me to what extent you agree or disagree that social benefits and services in [country]. . .

[*Agree strongly* — *Agree* — *Neither agree nor disagree* — *Disagree* — *Disagree strongly*]

- . . . place too great a strain on the economy?
- . . . prevent widespread poverty?
- . . . lead to a more equal society?
- . . . cost businesses too much in taxes and charges?

And to what extent do you agree or disagree that social benefits and services in [country]. . .

- . . . make people lazy?
- . . . make people less willing to care for one another?

Many questions regarding the same topic (2)

PISA 2015 - Teacher Questionary:

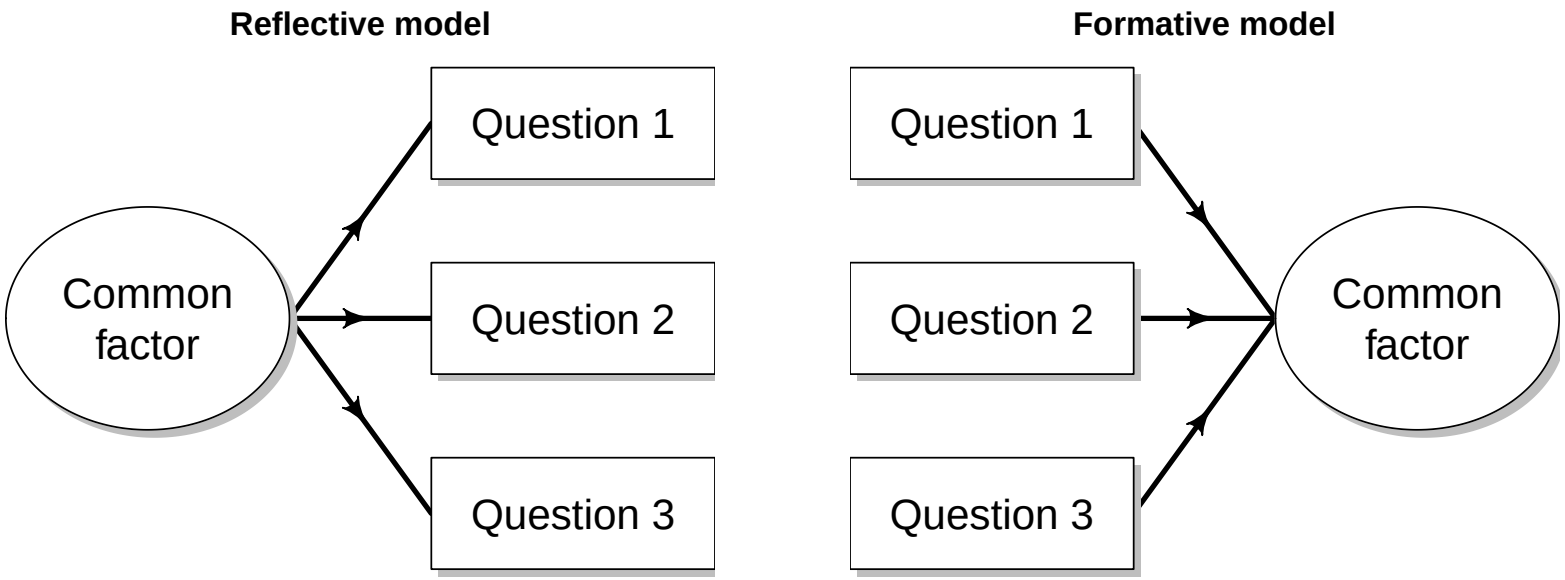
TC026 We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?

[*Strongly disagree* — *Disagree* — *Agree* — *Strongly agree*]

- The advantages of being a teacher clearly outweigh the disadvantages.
- If I could decide again, I would still choose to work as a teacher.
- I regret that I decided to become a teacher.
- I enjoy working at this school.
- I wonder whether it would have been better to choose another profession.
- I would recommend my school as a good place to work.
- I am satisfied with my performance in this school.
- All in all, I am satisfied with my job.

Constructing a scale

- Clearly this questions regards the same topic.
- We should use them all together to draw conclusions.
- We must extract a factor that is a common cause of answers to this questions.
- It will allow us to draw more general conclusions and to do it with more precision (with less random error).
- Nearly the same diagram – fundamental difference in interpretation:



Today we will be thinking in a *reflective* way!

Scaling as hypothesis testing

- Assumption of a common cause in fact is a kind of hypothesis.
- We assume that relationships between observed variables are caused **by, and only by**, the fact, that they share a common cause.
- In practice some questions may not fit to the others.
- There can be also some other factors distracting relationships between observeed variables.
- Scaling gives us a chance to asses, how well given set of questions is suitable to be interpreted as indicators of a more general factor.
- Sometimes we can also extend our model to correct some problems.

There can be more than one common factor

European Social Survey - Round 8 (2016):

Now I will briefly describe some people. Please listen to each description and tell me how much each person is or is not like you. Use this card for your answer. [below is a male version of wording]

[*Very much like me — Like me — Some-what like me — A little like me — Not like me — Not like me at all*]

- Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.
- It is important to him to be rich. He wants to have a lot of money and expensive things.
- He thinks it is important that every person in the world should be treated equally. He believes everyone should have equal opportunities in life.
- It's important to him to show his abilities. He wants people to admire what he does.
- (. . . 15 other items. . .)
- Tradition is important to him. He tries to follow the customs handed down by his religion or his family.
- He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.

Plan of a workshop

1. A little of statistical theory
2. One-dimensional models
    1. Simple model
    2. Assesing model fit
    3. Extracting factor scores
    4. Distracting factors - subscales and question formats
    5. Multiple groups - freeing distributions
    6. Multiple groups - DIFs/invariance
3. Multidimensional models
    1. Estimating Exploratory Factor (and IRT) Analysis model
    2. Choosing the number of dimensions
    3. Rotations
    4. Extracting factor scores in multidmensional models

A little of statistical theory

Two traditions: EFA/CFA and IRT

- Origins of constructing scales comes from Classical Test Theory (CTT):
  - Grounded in measuring physical properties (continuous variables).
  - The most prominent problem: estimating reliability.
- CTT was further developed in order to measure mental factors (cognitive abilities, personality traits, etc.) in the form of Explanatory/Confirmatory Factor Analysis (**EFA/CFA**):
  - often multidimensional;
  - concentrated on reconstructing covariation/correlation matrix of observed variables (linear relationships);
  - for a long time ignoring the fact, that analized variables are often categorical (ordered, but not continous).
- Other approach, not grounded in CTT, was Item-Response Theory (**IRT**):
  - concentrated on modeling - possibly complex - relationship between general trait and answer to each question;
  - statistical techniques drawn from dose-response toxicological research;
  - categorical character of observed variables is a central assumption;
  - in practice for a long time unidimensional (computational complexity).
  - Relationship to *classical* CTT described in detail in late '60 (Lord & Novik).

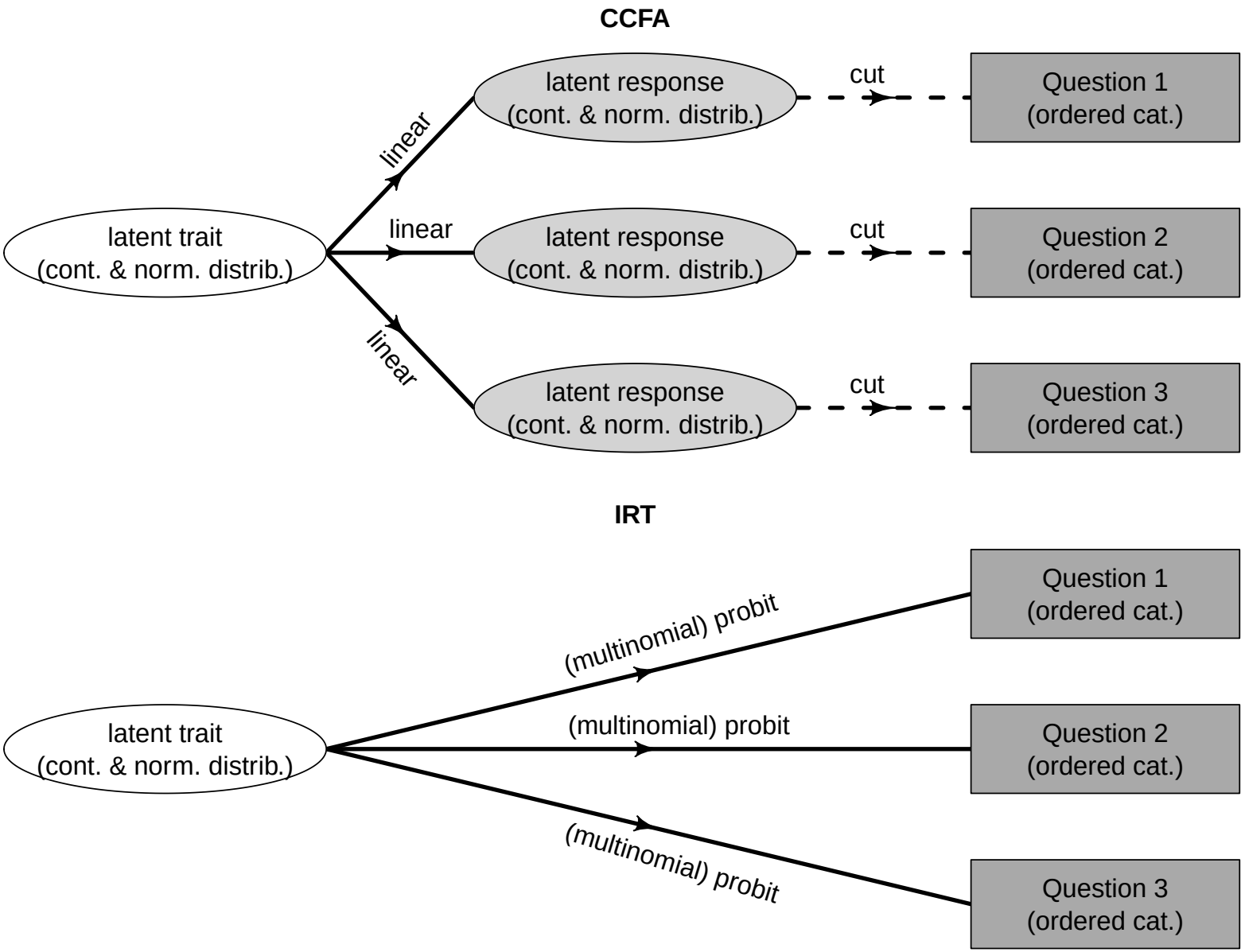
Further developement

- EFA/CFA:
  - taking into account that observed variables are (ordered) categorical by using polychoric correlations marix - so called Categorical Explanatory/Confirmatory Factor Analysis (**CEFA/CCFA**).
- IRT:
  - developement of multidimensional IRT;
  - proofs of equivalence of 2PL/SGRM IRT models with probit link function to CCFA models (Takane, de Leeuw 1987; Bartholomew 1987).
- **Nowadays we may think of CEFA/CCFA and IRT mostly as a different approaches to estimation of model parameters, both having its own advantages and limitations.**
  - CEFA/CCFA:

1. First estimate matrix of polychoric correlations.
2. Proceed (almost exactly the same) as with typical EFA/CFA.

  - IRT: variants of maximum likelihood estimation (or MCMC) with repspect to the *raw* dataset.
- Nevertheless in a research practice CEFA/CCFA and IRT remain rather separated approaches.

Two equivalent solutions to the same problem



CCFA: pros and cons

Pros:

- Less computational demanding.
- Many indices describing overall model fit (ie. if you want to publisch your results in psychology journal).
- Widely recognized.

Cons:

- Can't be applied to typical data with by-design missing values (rather uncommon case in typical surveys, but often appear in cognitive ability tests).
  - For example: all participants answer questions in booklet A. Next, on the basis of their responses, some of them are asked to answer questions in booklet B, and other to answer questions in booklet C. There are no participants that answers questions both in booklets B and C.
- Very little choice of model to describe relationship between latent trait and observed variables.
  - Less possibilities to diagnose problems with individual questions.

IRT: pros and cons

Pros:

- Can be applied to designs with by-design missing values.
- Wide choice of models describing relationship between latent trait and observed variables.
- Can be ilustrated with pretty graphs.

Cons:

- More computationaly demanding - especially for multidimensional models.
  - Computational complexity rises expotentialy as a function of dimensions.
- Assesing overall model fit is rather hard.
  - Methods involving simulations performs the best, but they can be applied only to rather simply models.

Basic concepts

- Observed variables are effects of a common factor that can't be observed directly (or perhaps few such factors - in multidemnsional models).
- The strength (and direction) of relationship between common factor and observed variable is described by **factor loading**.
  - We can think of factor loading as of a correlation.
- The stronger are this relationships, the more precisely we can estimate values of a common factor.
- We prefer a situtation when all the questions have rather similar values of factor loadings than a situation where for some questions relationship with a common factor are much stronger (or weaker) than for the others.
  - In the first case we can say all questions are *equally good* indicators of a general concept that we are measuring.
  - Questions with very low factor loadings should be excluded from a model - we can say, that they don't measure the same factor (concept) as the other questions.

Our tools

Lavaan

- *The lavaan package is developed to provide useRs, researchers and teachers a free open-source, but commercial-quality package for latent variable modeling. You can use lavaan to estimate a large variety of multivariate statistical models, including path analysis, confirmatory factor analysis, structural equation modeling and growth curve models.* (official package page)
- We will use *lavaan* to perform CCFA but it can be easily used to estimate wide variety of SEM models (estimated on the basis of covariance/correlation matrix).
- *laavan* itself do not enable user to perform typical EFA/CEFA. It can be done with use of functions from the *semTools* package.

mirt

- In my opinion first-choice IRT package for R.
- Flexible but can be easy used with default settings.
- Providing summaries calculated in an EFA/CFA manner (factor loadings, rotations).
- Allowing for multiple group analysis.
- Allowing for latent regression (some kind of simply SEM model).