METHODS OF QUANTITATIVE DATA ANALYIS
MSR Course, 2011-2012

Harry B.G. Ganzeboom

Lecture 4a: Index Construction

April 27 2012

Index construction

- A common task in data analysis is to combine the information in multiple indicators – that are assumed to measure the same underlying construct – into a single index variable.
- By far the most popular method is the summative index variable, consisting of an unweighted average of the (available) indicators.
- Alternatives: component scores (weighted indicators); cumulative scale models (guttman-, mokken-, rasch-scales).

Steps to take

- Step 1: Inspect and describe the indicators. List the item formulation, min, max, mean, dispersion, valid N.
- Step 2: Dimensional analysis by inspecting or factor-analyzing the correlation matrix. Determine which indicators can represent a single underlying dimension.
- Step 3: Reliability analysis to remove bad indicators and maximize / estimate total reliability.
- Step 4: Compute the index by taking the average of (standardized) indicators.
- Step 5: Fix the metric of the index, e.g. by standardization.

Step 1: Describe

- Always make a table of the indicators with the relevant descriptive information. List the item formulation, min, max, mean, dispersion, valid N.
- Determine direction of the indicators.
- Are the metrics comparable? If not: standardize.
- There are three ways of standardization: (A) Z-scores, (B) P-scores [percentiles], (C) D-scores [dichotomization]. All have their (dis)advantages.

Step 2: Dimensional analysis

- Calculate and inspect the correlation matrix:
  - Dimensionality – are there subgroups of items?
  - Signs: Where and why do we see negative correlations? Can we make them all positive?
- Factor analysis can be used to decide on dimensionality. Preferred method: extraction=PAF, rotation=oblimin.
- If your aim is “simple structure”, then you can also aim for uni-dimensionality.

Step 3: Reliability

- Notice that reliability can only be performed on listwise data.
- Also note that reliability analysis assumes:
  - Uni-dimensionality
  - Consistent direction of the indicators
  - Same metric of (standardized) indicators.
- Note that reliability analysis is NOT a way to determine dimensionality.
- Provides you with:
  - The set of indicators appropriate the most reliable index variable
  - An estimate of the reliability of the index

Step 4: Calculating the index

- Take an unweighted average of indicators.
- Make sure that the indicators have a comparable metric and consistent direction.
- If you use (in SPSS) the **`mean`** function in **`COMP`**, you get rid of a lot of missings.
- Alternative: component score. This maximizes the variance of the constructed variable. There are practical problems with this with respect to missing values.

Step 5: Rescaling the index

- It is of some importance to fix the metric of the index variable to a known quantity.
- Note: when averaging Z-scores or P-scores, the result is not a Z-score, or a P-score (although they look like it…).
- Re-standardization to P or Z score is an obvious method.
- Alternative: rescale between 0 and 1 (natural chice when you would be starting with dichotomized [D] indicators.

More on reliability

- Reliability = statbility of the measure. A measurement is perfectly reliable when it gives an identical result the next time you measure (and there are no changes).
- Test-retest reliability is conceptually the heart of the matter.
- However, this only works for objects that do not change: e.g. gender, year of birth, father's occupation, education (?). If things change, you cannot use this operationally.

Internal consistency

- Internal consistency (cronbach's alpha) is an estimate of test-retest reliability on no change.
- It estimates the reliability of the index (sum) variable.
- Assumptions:
  - Items are truly parallel (no difference in meaning).
  - Items are one-dimensional.
  - Items correlate homogeneously

Simplex reliability

- Reliability can also be estimated with a simplex (causal chain) model. This is also applicable to objects that have no internal consistency whatsoever.
- Here we need a panel design:
  - At least three waves
  - No direct effect of first wave on third wave.
- This can be applied in many situations.