Selected Quantitative Methods

Lecture 2a

What you already should have known…

Harry B.G. Ganzeboom

Phd Course SQM

November 4 2013

Outline

- I review basic knowledge and skills that should have been acquired in earlier training.

INTRODUCTORY

Descriptive statistics

- Measurement, levels of measurement: nominal, ordinal, interval, ratio[, dichotomy].
- Central tendency and dispersion, by level of measurement: mode, median, means; heterogeneity, range, variance/standard deviation.
- Ranks and n-tile scores (P-standardization).
- Z-standardization.
- The (standard) normal distribution.

Measures of association

- Conditional means / percentages.
- Simple regression:
  - Linear model (intercept / slope).
  - Sum of squares, explained variance.
- Pearson and spearman correlations.
- Correlation matrix.

Inferential statistics

- (Simple) random sampling.
- Sampling distributions and SE (sampling error, standard error).
- T-distribution and Z-distribution.
- Estimation: confidence intervals.
- T-test for differences between two means.
- F-test for differences between multiple means.

Type I / II errors.

- H0 and H1: Null-hypothesis and alternative hypothesis.
- Statistical significance and type I errors.
- Statistical power and type II errors.
- Fundamentals of increasing power:
  - Larger samples
  - Better research designs

Measurement

- Validity and reliability, systematic and random measurement error.
- Elementary multiple indicator measurement: determining uni-dimensionality using a correlation matrix.
- Internal consistency estimates of reliability (Cronbach's alpha).
- Building a multiple indicator index: standardization, dealing with incomplete information, constructing an interpretable metric.
- Validation: criterium validity, construct validity.

Experimental Design

- Randomized groups.
- Before/after designs and other forms of matched designs.
- Double blind.
- Analysis of variance.

SPSS

- Data matrix.
- Principles of use of syntax. Managing and archiving files.
- Recode, compute, count.
- IF and DO IF.
- Aggregate.
- Sorting, splitting, filtering, saving and matching data files.
- Interface between SPSS output and Excel.

Observational designs

- The fundamental difference with experimental designs: confounding / spurious effects.
- Elaboration: eliminating confounding using partial tables.

ADVANCED

Multiple regression

- The linear model with multiple X.
- Standardized and unstandardized models.
- Interpreting stepwise procedures.
- How multiple regression keeps confounding variables constant.
- Consequences of (random) measurement error on causal effects X→Y.

Causal models

- Direct, indirect and total effects.
- Path analysis: correlation = direct effect + indirect effects + confounding effects.
- (The multiplication rule for indirect effects).
- Interpreting multiple regressions models in a causal framework.

Dimensional analysis

- The classical measurement model with multiple latent variables (=factor analysis).
- ‘Extraction’ and ‘rotation’.
- Principal component analysis.
- Use of factor (=component) scores.

Interaction / moderation

- Multiplicative interaction terms.
- Understanding that 2 variables and 1 interaction term together still make 2 variables and an infinite number of effects.
- Interpreting an interactive linear model.

MASTER

Simultaneous equations

- Simultaneous equation models (SEM) move away from single equation model:
  - Causal models to estimate intervening and confounding effects.
  - Estimate measurement models (= factor analysis) together with the causal model. Latent variables.
- Stand alone programs: LISREL, MPLUS.

Latent variable analysis

- In measurement theory we make a distinction between the real thing (‘true score’ or ‘trait’) and the data matrix (=how you measured) and take seriously that these are not the same.
- Multiple factor analysis: how many and which dimensions (= latent variables) account for correlations between the observed variables?
- Latent variable analysis is an integrated method for reliability and validity analysis.
- Confirmatory factor analysis replaces exploratory factor analysis.
- MTMM (multi-trait, multiple measures) allows to separate and repair random and systematic measurement errors.

Latent class analysis

- Latent class analysis is similar to latent factor analyses (and its generalization into SEM), but now:
  - the latent and observed variables are discrete,
  - the associations (measurement relationships) are probabilities.
- LCA is often used to determine how many groups exist (e.g. consumer types).
- Stand-alone programs: LatentGold, LEM.

Nominal dependent variable

- When the Y-variable is dichotomous: binomial logit / logistic and probit regression.
- When the Y-variable is polytomous: multinomial logistic regression.
- When the Y-variable is polytomous and ordered: ordered logistic regression.

Conditional logistic regression

- If you have a multi-nominal dependent variable, with many categories, we would like to bring (multiple) properties of the categories into the model.
- Examples: party choice (left-right, denominational), occupational choice (economic status, cultural status), brand choice (price, color).
- CLOGIT allows to condition the choice of certain X-units on the (multiple) properties of Y.

Complex samples

- What if there was no simple random sample? For instance: hierarchical sampling designs:
  - Students in classes in schools in cities;
  - Voters in cities in regions in countries.
- Multi-level analysis:
    - estimate how much variance there is at various hierarchical levels (variance components);
    - Assess how much of this variance is explained by a linear model with (cross-level) interactions.
- Cluster corrections.

Longitudinal observations

- Longitudinal designs: same unit is observed at multiple time point (with the same measurement).
- Data matrix: multiple observations within a single unit. Similar to hierarchical random sampling.
- Event history (survival or hazard) analysis: how long does it take before an event happens? Cox regression.
- Panel model: analysis quantitative fluctuation between and within the same units. XT-models. Fixed effects and random effects estimators.

STATA or R?

- You cannot beat SPSS in simple data manipulation, but it is not a great program in many other respects.
- Alternatives:
  - Stata: good, expensive, very large user community, extremely good documentation.
  - R: shareware (FREE!), large user community, steep learning curve.