

SOS3003
**Applied data analysis for
social science**
Lecture note 09b-2009

Erling Berge
Department of sociology and political
science
NTNU

Spring 2010

© Erling Berge 2010

1

Producing a term paper

- At the beginning of the class we were talking about the requirements of the paper and a bit about how to write it
- I hope you have studied it carefully:
<http://www.svt.ntnu.no/iss/Erling.Berge/2010%20SOS3003%20SemOppgKravEN201001.pdf>
- Today we shall consider the practicality of building and estimating model

Spring 2010

© Erling Berge 2010

2

Step 1: Dependent variable

- Investigate the distribution of cases on the dependent variable
 - Think about what mechanisms may generate high or low variable values for particular cases
 - Make a list of such mechanisms
 - Can you find information on these mechanisms in the data?
 - Make suitable variables of those you find

Spring 2010

© Erling Berge 2010

3

Step 2 Types of research problems

- The dependent variable will usually be either an
 - Indicator of obtained status of some kind (education level, marriage status)
 - Indicator of activity of some kind (work, industry, leisure activity, voting behaviour)
 - Indicator of strength of attitude or belief of some kind (political preferences, trust, type of entertainment)
- The problem of modelling the variation is different for the different types of variables
 - They will have different causal structures

Spring 2010

© Erling Berge 2010

4

Step 3 Types of causal mechanisms I

- Structural causation
 - Social structure does have causal impacts that are not well understood. In a framework of methodological individualism one may say that it limits and orders the options that actors can choose from. Hence, variables such as age, sex, and place of living can be used as proxies for poorly understood causal factors.
 - Budget constraints (time and income constraints) have the same character. They limit and orders the options that actors can choose from. However, they enter the model more through the way the dependent variable is constructed, and the kind of link function (linear or logistic) used to mediate between observations and dependent variable.

Spring 2010

© Erling Berge 2010

5

Step 3 Types of causal mechanisms II

- Individual causation
 - Preferences (norms, values, attitudes) may be difficult to observe in detail but are assumed to be present
 - Resources (income/ capital, education/ human capital, access to networks/ social capital) are usually measured extensively even if unevenly. Here there are budget constraints
 - Perception of opportunities will often depend on position in social structure

Spring 2010

© Erling Berge 2010

6

Step 4 Explanatory variables

- All kinds of explanatory variables are allowed
- Make a list of conceivable variables
- Look for direct or indirect indicators for the variables. Approximations are allowed
- Construct new variables where variation and codes match as well as possible the intended indicator
- Then the first model can be estimated

Spring 2010

© Erling Berge 2010

7

Elaborating the model

1. Distribution of residual normal?
2. If no:
 - i. Curvilinearity? If yes, fix the problem and go back to 1. Else
 - ii. Missing variables? (correlates with both y- and x-variables)
 - iii. Heteroscedasticity? If yes, fix the problem and go back to 1. (Fixing this may entail transformation to symmetry.)
 - iv. If tests are trustworthy remove obviously irrelevant variables and go back to 1
3. If yes: you have a first estimate of your model
4. Consider how it may be improved!

Spring 2010

© Erling Berge 2010

8