

Essayopgave 3

Logit and Probit Regression

Due: 28. March at 12:00 to Thomas via email (tleeper@ps.au.dk)

Your task:

1. Identify a binary outcome available in the General Social Survey 2002 dataset. *Remember that you can sometimes recode categorical or ordinal variables into a binary outcome by collapsing categories.*
2. Develop a research question about one or more possible causes of that outcome.
3. State and briefly justify a clear theoretical expectation (i.e., hypothesis) about one or more causal variables.
4. Test your hypothesis using a binary outcome logistic regression model that accounts for possibly confounding factors.
5. Estimate the size of effects using predicted probabilities and/or marginal effects. Represent those results both as text (or a table) and as a graph.
6. Compare the effect size estimates from the logit model to those generated from a probit specification of an identical model (i.e., reestimate the exact same model using probit rather than logit regression and compare the substantive interpretations from each model to one another).
7. Summarize your results in a brief conclusion.

Data:

For this assignment, please use the General Social Survey 2002 dataset. This dataset represents a nationally representative survey of United States residents. The GSS has been conducted every 1–2 years since 1972 and is considered very high quality survey data covering a range of social and political issues in addition to including a large number of demographic variables. The dataset in Stata (.dta) format can be found at http://publicdata.norc.org/GSS/DOCUMENTS/OTHR/2002_stata.zip.

The output of the Stata `codebook` command has been uploaded to Blackboard to help you explore the dataset more easily.

Note: The dataset includes three types of missing values: `.d` meaning the respondent said “don’t know,” `.n` meaning the respondent supplied no answer, and `.i` meaning the respondent was not asked the question.