B4 - Applied Statistical Methods
Fall 2017
HWA Voting in Sweden

Johan Lyhagen/ Lars Forsberg

December 4, 2017

1 Introduction

The purpose of this home work assignment is to get practical experience in modeling data using logit and probit models, both for binary and ordered data. This includes estimation using the software package R, to intpret and analyze the results.

R is installed in computer room H319 and it is also possible to access through remote desktop to *ts03.dis.uu.se* and log on as if you actually were in H319.

2 Written report

A written report should be handed in no later than the date indicated on the schedule. The written report should be written in such a way that there is no problem to understand what have been done and why.

3 Oral presentation

You shall present your work for the other students in "**seminar-format**". Due to the number of students taking this course, we need to split up in two seminar groups, for simplicity, we will denote these groups:

1. Seminar group 1, and
2. Seminar group 2.

The dates and times for the two seminar groups are the schedule.

4 Group formation

Form working groups of **two maximum three** students. One of the students send an **email** to

Lars.Forsberg@statistik.uu.se.

Please state in that email if you have any preference concerning which *seminar group* you want to belong to. To the extent possible, these preferences will be taken into account when allocating working groups to seminar groups.

You will then get a email-reply saying what

1. **working group number** you have,
2. what **seminar group** you belong to, and
3. what **party** you should analyze.

5 Data

The data used are from the European Social Survey which is a biennial multi-country survey. In this home work assignment we are going to use data from Round 3 and for Sweden, collected in 2006. In Table (1) the questions are classified into topics. A total of 487 variables on 1927 respondents are collected but we are going to use a rather small subset. The subset of questions are shown in Table (2) . For a description of the formulation of each question see the document *ESS3Source_ Mainquestionaire*.

Question B12 is regarding which party the respondent voted for in last election. The parties are as follows: the

1. Centre party (1)
2. The Liberal party (2),
3. The Christian Democrats (3),
4. The Green party (4),
5. The Conservative party (5),
6. The Social Democrats (6),
7. The Left party (7).

The codes 8-10 is for other small parties.

(As mentioned above, once you have formed the groups, you will be assigned a party to analyze, in this presentation, we proceed *as if* we were assigned the **conservatives.)**

Category	Topic
A	Media; Social trust
B	Politics
C	Subjective well-being; Religion
D	Timing of life; The life course
E	Personal and social well-being; Satisfaction (life, work)
F	Socio-demographic profile
G	Human values
H	Test questions
I	Interviewer self-completion questions

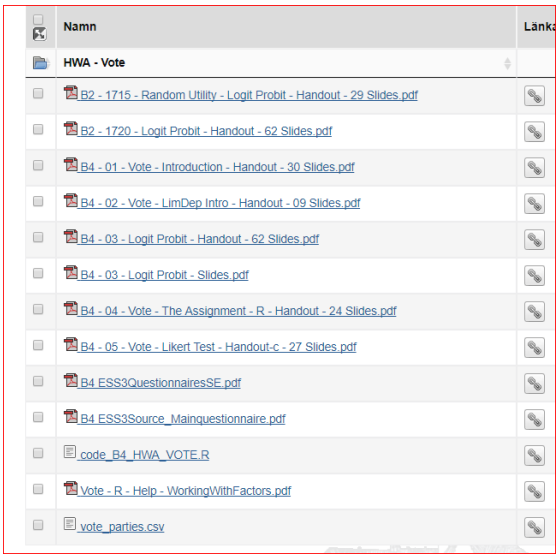
Table 1: Classification of the questions of the European Social Survey, Round 3.

Topic	Questions
A	A1, A5
B	B1, B11, B12
C	C1, C5, C15, C28, C33, C35
D	D4, D6, D9
E	E4, E5, E7
F	F1, F2, F3, F5, F27, F30, F32

Table 2: Variables availble in the file QDATA.wf1

6 Material on studentportalen

Material on studentportalen



7 Assignment

Below, the assignment is stated for the Conservative party, but as mentioned above, you will be assigned what party to analyse.

1. Display and interpret descriptive statistics for how the voters voted in the last general election in Sweden (variable B12).

2. Specify and estimate **logit** model for voters voting on your assigned party. Dependent variable is "voting for the XXX party" (binary variable: 0,1). The model **should** include a Likert scale explanatory variable. Justify/ motivate the model/ choice of variables. You should motivate why a specific variable is in the model and what sign you expect on the coefficient. Estimate and interpret the model.

3. Interpret the parameters of the model in a way you find appropriate. Also, for a typical individual of the sample (pick one), calculate the (model implied) probability that this individual would vote for the conservative party. Change the value of the explanatory variables and analyze how the probability change. (What values of the other explanatory variables would it make sense to use? Discuss and motivate)

4. Given the logit model in task 2: Test, using a **likelihood ratio test**, if the (constant) scaling implied by the Likert scale, i.e. 1,2,3 etc is valid. That is, is the restriction that the distances between the 1 and 2 are the same as the distance between 2 and 3 etc valid? (You need to estimate two models, one restricted and one unrestricted). What is the implication of the test result in terms of the model and how to use the Likert variable in this situation? Can this result be generalized, if so, how, if not, what would be the recommended practice in this situation?

5. Make a two by two table of observed responses versus predicted. Use a 0.5 cut of value and then a cut of value corresponding to the unconditional probability to vote for you assigned (XXX) party. Interpret and discuss. Choose a goodness-of-fit measure to evaluate the model.

6. Using the probit instead of the logit, estimate the model in 2 and compare the implied probabilities, how do they differ? Compare the two by two table of observed responses versus predicted for the two models. Which of the Logit and Probit would you recommend to use for this data? Discuss and motivate.