POL 450C, Homework 3

April 20th, 2017

Assigned: 4/20/2016
Due: 4/27/2016

In this problem set we're going to analyze American's propensity for supporting budget reduction using the data set in `ANES.RData`. In particular, we're interested in analyzing an originally 7-pt scale on support for budget deficit reduction as a 3-pt scale. The three levels are:

$$Y_i = \begin{cases} 1 & \text{If respondent favors reducing budget deficit} \\ 2 & \text{If respondent does not lean either way} \\ 3 & \text{If respondent opposes reducing budget deficit} \end{cases}$$

We have the following independent variables

X_{i1} = Republican (0/1)

X_{i2} = Democrat (0/1)

X_{i3} = Family income, thousands

We will use an ordered probit model to infer the relationship between family income and support for deficit reduction, conditional on partisan identification.

1) Call the coefficients $\boldsymbol{\beta}$ and the cutoffs $\boldsymbol{\Psi} = (\psi_1, \psi_2)$. Write the data generating process for the ordered probit, the likelihood for the coefficients $\boldsymbol{\beta}$ and the cutoffs $\boldsymbol{\Psi}$ and the log-likelihood.

2) This question asks you to think more carefully about the cutoff values.

   a) State in words an interpretation of the cut off values $\boldsymbol{\Psi}$.

   b) In class we stated that the ordered probit generalizes the probit, this question helps you see why. Suppose you have two levels: Y_i is either equal to 1 or 2. Show that the ordered probit with two levels is equivalent to a probit regression with a dichotomous dependent variable.

3) From the `MASS` library use `polr` to obtain maximum likelihood estimates of deficit-reduction preferences on $\boldsymbol{X}_i$ and the cutoff values $\boldsymbol{\Psi}$. Report the maximum likelihood estimates of the coefficeints and cut-off values and the accompanying standard errors for these estimates.

4) Using the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ we will calculate predicted probabilities for all three levels of support for deficit reduction across the range observed values of family income for Republicans, Democrats, and Independents and then generate 95-percent confidence intervals for the predicted probabilities. To obtain the estimates and 95-percent confidence intervals we will use both the bootstrap and simulation from the multivariate normal. We walk through both procedures in this question.

   a) Bootstrap:

      - For 10,000 iterations
         i) Sample N observations, with replacement
         ii) Use `polr` to fit an ordered probit on the sample
         iii) Using the estimates from this iteration, calculate the predicted probabilities for each level of support for deficit reduction, for a sythetic respondent who is a Republican, Democrat, and Independent, varying the level of income from the minimum to the maximum in the sample. To be specific, for Republicans you will calculate the predicted probability $Y_i = 1$, $Y_i = 2$ and $Y_i = 3$ for all observed levels of family income. Store the predicted probabilities for each level and party group (this will be easiest if you create a separate matrix for Democrats, Republicans, and Independents; or create an array).
      - Create three plots (one for Republicans, Democrats, and Independents) and present the average probability for each level of deficit reduction support againt family income. Present the 95-percent confidence intervals for the calculated predictions (hint: you could draw lines connecting the top and bottom of the confidence intervals, or use `arrows` to draw the confidence band at each point, or some other preferred method). Use a `rug` to show where there are observed data for Republicans, Democrats, and Independents.

   b) Multivariate Normal:

      - Draw 10,000 realizations of

$$(\boldsymbol{\beta}, \boldsymbol{\Psi})^t \quad \sim \quad \text{Multivariate Normal}\left((\boldsymbol{\beta}^*, \boldsymbol{\Psi}^*), I_N(\boldsymbol{\beta}^*, \boldsymbol{\Psi}^*)^{-1}\right)$$

        where $\boldsymbol{\beta}$ is a vector of regression coefficients and $\boldsymbol{\Psi} = (\psi_1, \psi_2, \ldots, \psi_J)$ is a vector of cutoff values.
      - For each realization,$(\boldsymbol{\beta}, \boldsymbol{\Psi})^t$ calculate the predicted probabilities for each level of deficit reduction and party ID varying income levels. Create plots that include average predicted probabilities and 95-percent confidence intervals. Include a `rug` to show where there are data to support the inference.

5) Interpret the plots. What do you notice about the relationship between income and support for deficit reduction?

6) Fit the same model using a linear model and create a plot describing the relationship and 95-percent confidence intervals. What differences (if any) do you note? Compare the coefficients of the ordered probit and linear regression in a table.