**SO5012 Analysing Data in the Real World**

**Seminar X – Multinomial Regression**

**Introduction and Data**

﻿The main data file used today is UKVOTE2010. It is based on survey responses about voting preferences in the 2010 UK elections. The dependent variable of interest VoteIntention is voting preference among the three major parties in the UK: Labour, Conservative, and Liberal Democrat. Voting preferences for minor and regional parties are grouped together in a fourth category. There is a lot of missing data in this data set, as many respondents to the survey stated no preference at all. We will focus on analysing just those respondents who did state a preference, trying to understand how preferences varied as a function of Gender, Age, and Qualifications. We are going to focus on the qualifications variable, which has four levels in decreasing order of educational attainment: BA, GCE (previously A-Levels), GCSE (previously O-levels), and a residual Technical/None/Other category.

**Seminar tasks**

1. ﻿Cross-tabulate the variable VoteIntention with the variable Qualifications, setting the table to include the conditional probabilities of VoteIntention, given Qualifications. Which parties do better among those with higher qualification levels and which do worse?
2. Fit a multinomial logistic regression model for VoteIntention, with Labour as the baseline outcome category, using only the variable Qualifications as an explanatory (factor) variable. How can we see from the coefficients which parties do better among those with higher education levels and which do worse? Check that you see the same general patterns as you saw when you cross-tabulated the same data.
3. Run a logistic regression for a voting intention of Labour (as the baseline) vs conservatives. What do you notice about the results compared to those found in question 2?
4. Add the variable Age to the model. For which outcome levels is there a significant association between age and vote intention, controlling for qualifications?
5. Perform a likelihood ratio test to see whether Age is a significant predictor across all outcome levels.
6. Now, add the variable Gender to the model. Relative to men of the same age and qualifications, which parties are women more/less likely to vote for?

**Further work**

﻿If you finish all this, and are stuck for more work to do, or you find this a bit boring and want to try this out on real data – try it on British Election Study from last term. Choose a voting variable as a dependent, recode the education variable and off you go. Are the results the same on that dataset as they were are on this old data?