ME104 Linear Regression Analysis: Problem Set 8
Models for categorial and ordinal data.

1. The main data file used today is `science2.dta`. It is based on data from two social surveys, one carried out in Europe and one in the US, which were used in an article by Gaskell et al. (2005). One of the research questions of Gaskell et al. (2005) was to examine what characteristics of individual citizens affect their opinions about governance of science, i.e. the way in which decisions about science and technology in a society should be made. Two dichotomies were first considered, based on survey respondents opinions on whether such decisions should be based on (i) expert advice or views of average citizens, and (ii) moral and ethical issues or scientific evidence of risk and benefit. Combinations of these give four positions, which the authors labelled as follows:

   -Scientific elitism: expert advice and scientific evidence
   -Scientific populism: public opinion and scientific evidence
   -Moral elitism: expert advice and ethical considerations
   -Moral populism: public opinion and ethical considerations

`valueposition`	shows each respondents position.
`female`	a dummy variable for women.
`age`	age, recorded as mid-points of intervals of a grouped age variable, but treated as continuous here.
`higheduc`	a dummy variable for whether the respondent has been in full-time education at the age of 20 or over.
`highrelig`	a dummy variable for attending at least once a month
`us`	dummy variable for US respondents

   (a) Fit a multinomial logistic regression model for `valueposition` given the five explanatory variables. Multinomial logistic models are obtained with the `mlogit` command. The option `baseoutcome` can be used to set the reference level of the response variable.

   (b) The command `mlogtest, lr` produces the results of likelihood ratio tests when we remove each of the explanatory variables in turn from the model while keeping the others in it. Are there any nonsignificant explanatory variables in this? (Hint: To use the `mlogtest, lr` command you need an add-on package "SPost", see the do.file for this).

   (c) If there are nonsignificant explanatory variables in the model, refit the model with one of them omitted, and repeat this process untill the likelihood ratio statistics are significant for all the remaining variables (at, say, 5% level of significance).

   (d) Generate an interaction between location (Europe or US) and religiosity and add it to the previous model. Are the likelihood ratio tests for age, education and the interaction between region and religiosity significant?

(e) Examine the regression coefficients for `higheduc` and the Wald tests and confidence intervals for them. Compared to the reference level of Scientific elitism, which value position appears to have its probability affected by the level of education? What is the interpretation of the coefficient corresponding to this effect?

(f) Calculate the fitted probabilities considering respondents of ages 20-75, fixing the other explanatory variables at us=0, highed=1 and highrel=0 using the `prgen` command.

(g) Plot the fitted probabilities against age.

(h) Calculate fitted probabilities at all possible combinations of the three categorical explanatory variables, fixing age at 50. Interpret some of these probabilities.