**SC968 PANEL DATA METHODS FOR SOCIOLOGISTS**

**COURSE ASSESSMENT PAPER 2010-11**

**This assessment counts as 50% of the final mark for SC968, unless the student’s exam mark is higher, in which case the course is assessed on the exam mark alone.**

**Please answer both questions. They carry equal weight for your assessment grade. For this exercise, use the data set stored in the SC968 current area as “teaching.dta”. Please type your answers into this assignment sheet; you may also copy and paste output from STATA.**

**The deadline for this assessment is Thursday 28th April 2011.**

1. **A RESEARCH PROPOSAL ON THE DETERMINANTS OF DIVORCE AND SEPARATION**

The government’s Department for Work and Pensions has invited you to prepare a proposal for a programme of research into the determinants of divorce and separation – focusing particularly, but not necessarily restricted to, couples who have children. The Department is especially interested in the role that poverty, and other factors linked with poverty, play as determinants of divorce and separation. They have expressed an interest in research which uses longitudinal data.

For this exercise you should prepare a research proposal, which includes the following elements.

1. An outline of why the proposed area of research is important, and what previous studies have found. For this assignment, you are not required to present a comprehensive literature review, but you may like to refer to a couple of studies in the area.

2. Details of the research questions which you propose to address, together with associated hypotheses.

3. A section making the case for using panel data in the research, and explaining how panel data can inform this question in ways which cross-sectional data could not.

4. A description of the methods which you propose to use for the analysis (which would include the methods taught in SC968, but which need not be restricted to these methods), and the particular aspects of the research question which will be addressed by each of these methods. The potential strengths and limitations of each of the methods should be pointed out.

5. A brief outline of the data you will use (here, the BHPS). This section should also include brief descriptions of how you will define the variable(s) describing divorce and/or separation, and of the explanatory variables which you will be seeking to use in multivariate analysis (these variables should not be restricted to the variables present in the SC968 teaching data set, but should be present in the BHPS).

It is not usual for research proposals to contain a large volume of analytical results. However, they would often contain a couple of tables showing descriptive statistics, cross-tabulations, transition matrices, etc. These may serve to illustrate the importance of the topic, or to underline the reasons why a particular approach is to be taken.

Proposals of this type are often subject to strict limits on length. For this part of the assignment, you are limited to 2000 words (excluding tables) – although an excellent proposal could be rather shorter than this.

**2 A COMPARISON OF A SURVIVAL ANALYSIS WITH A LOGISTIC REGRESSION ANALYSIS**

You are asked to carry out a study investigating the influences on the onset of a health problem that limits daily activities (variable **hllt**).

Potential explanatory variables that you are asked to consider include

* age
* gender
* marital status
* employment status
* income
* educational qualifications

Select an appropriate sample for this analysis from the teaching.dta to investigate this issue and explain why.

Show how you have recoded the explanatory variables and give your reasons where appropriate.

You may also include in your analysis other explanatory variables, any confounding variables that you think are appropriate and any plausible interaction effects.

Carry out an analysis using regression models for panel data with a dichotomous outcome (NOTE: you are not expected to use random coefficients models). Comment on the decisions you made when proceeding from one step to the next. Include output from any post estimation commands that you run to aid interpretation. Interpret the output from your final model.

Suppose you submitted this work for publication and a reviewer suggests that you should have used survival analysis to investigate influences on the time to the first onset of a limiting health problem.

Run an appropriate Cox regression model to do this based on the findings from your regression analyses. Remember to show all your steps and provide a commentary on the results. Include any supporting figures and supplementary tests that you carry out.

Finally, summarise the findings from the two approaches and comment on the differences in interpretation of the models and the conclusions that can be drawn from them.

Be aware that there is no one “correct answer” to this question. Please provide a printout of your Stata do file containing all your syntax with your answer.