POLM086 Quantitative Data Analysis (30 credit module)

Assessed Coursework Part 2: This is the second part of the assessed work for this module, Part 1 is a separate document. Each consists of 50 percent of the marks awarded for the assignment. Both parts must be completed by those taking the module for credit. All assessed work to be handed in using BART. We will discuss this assignment in detail and provide assistance in the classes before this work has to be handed in.

1. Conduct a review of data relevant to your research questions or broader research topic for your dissertation (for example as part of your MA/MSc/PhD). You can use internet searches for data and/or a set of data archives (you may need to register):

http://ukdataservice.ac.uk/

http://www.icpsr.umich.edu/icpsrweb/ICPSR/

https://dataverse.harvard.edu/

http://w.issp.org/menu-top/home/

https://ec.europa.eu/eurostat/data/browse-statistics-by-theme

http://data.un.org/

https://data.oecd.org/

If you cannot find data directly on you research question use data that is related to the broader topic area of your research and discuss the data's relevance. The review must include:

A specification of the research topic, theoretical background/issues and your research questions.

A list of some relevant data (give URL links to the data if possible), a statement of how the data could be used to address your research questions, a discussion of the opportunities for research and/or limitations of current data sources. Comment on whether and how variables in the dataset measure the main concepts of interest to you in your research (for example are they valid and reliable measures. Consider the dataset(s) potential for considering relationships of interest as dependent and independent variables for statistical analysis in a multivariate regression model. Comment on any limitations of the data for using regression to address your research questions.

2. For one interval/ratio variable in a dataset that measures a concept of interest to you in addressing your research question(s), plot a histogram of the variable's distribution. Report the mean, median, mode values and a measure of skewness. Comment on the overall shape of the distribution you see visually in the histogram (for example does it resemble the normal distribution, can you 'see' any skew?). Comment on whether there is missing data for this variable and whether there are outliers. Discuss whether and how the results of this analysis help you address your research question and the usefulness of this variable in addressing the question.

3. Use multivariate linear regression to analyse a relationship between a Y dependent variable and more than one X, independent, variables. The variables you choose for inclusion as Y and X variables in this analysis should include the variable you described in section 2 above. Report a two tailed t

test of the coefficients of the X variables with the null that the coefficient is equal to zero. Comment on the results and the implications, if any, for your research questions. Discuss how the inclusion of multiple variables in your regression (compared to a bivariate model with only one of the X variables you included) affects your results. Discuss the strengths and limitations of the regression model you have applied for assessing causal relationships between the X variables and the Y variable. In your discussion include the definition, and consider the potential relevance to your analysis, of omitted variable bias.

Length guide: 3,000 words equivalent maximum in total all sections -this total includes the lists of datasets and variables of interest to you and (it is likely you will not need to use the full word length but it is given as a maximum).