Statistical Methods for Business Research

Department of Management, Birkbeck, Autumn 2019

Coursework 2019

**Words Limit (2,500)**

**Deadline: 13th December 2019**

THE UK LONGITUDINAL SMALL BUSINESS SURVEY (2016)

(UK Data Service and data collection copyright owner)

*The Longitudinal Small Business Survey (LSBS)* is a large-scale telephone survey of small business owners and managers, commissioned by the Department for Business, Innovation and Skills (BIS). This survey is the latest in a series of annual and biennial Small Business Surveys (SBS) dating back to 2003.  
  
From 2015, the survey methodology changed to include a longitudinal tracking element. A large sample size was recruited in Year One (2015) to establish a panel of businesses that will be re-surveyed in subsequent years. This will allow a detailed analysis of how combinations of factors affect business performance. However, our data file only include a part of the cross-sectional data from 2016 survey. The detailed of the survey questions, as well as the coding information will be provided together with the dataset on the Moodle.

For each question you must provide the correct Stata coding. This will be included at the end of the coursework in appendix (APPENDIX DO FILE).

At the end of your work You should provide:

1. **A Word or PDF file, which includes your analysis of the outcomes, together with and all graphs and tables.**
2. **At the end of the word doc (or PDF) there should be an appendix with the Stata coding of the do-file as in the “example of coursework structure”**

*NOTE: the word document (PDF) must not exceed 2,500 words, excluding tables and graphs of course*

*NOTE: Any graph should be labelled and titled in such a way that it can be understood. Graphs and tables should be numbered (Figure 1, 2, …. Table 1, 2….). Do* ***not*** *include a graph or table if it is not mentioned in your written analysis.*

*You may discuss the methods required to answer the questions with classmates. You are required, however, to prepare your own do-files and write your own commentaries: the usual cautions about plagiarism apply.*

Questions Descriptive and Inferential Statistics

1. Provide the following label to the dataset:

“UK Longitudinal Small Business Survey 2016”

1. Rename P1\_2016 with “TURNOVER\_2016”, A6\_2016 with “AGE\_2016”, A2\_2016 with “EMPLOYEESIZE\_2016”
2. Generate a new variable consisting of the natural log of employee size (A2\_2016) and name this new variables “lnEMPLOYEESIZE\_2016”; Generate a new variable with the name of PRODUCTIVITY\_2016, equals to turnover (P1\_2016) divided by employee size (A2\_2016); Generate a new variable consisting of the natural log of PRODUCTIVITY\_2016 and name this new variables “lnPRODUCTIVITY\_2016”.
3. In a unique table, provide a simple summary statistics of the data by considering PRODUCTIVITY\_2016, employee size (A2\_2016), age of the business (A6\_2016), and family business in 2016. The table should contain the number of observations, means, standard deviation, minimum and maximum values of each variable. Give a brief description of the results.
4. Set a table with summary statistics of the variable PRODUCTIVITY\_2016, employee size (A2\_2016), age of the business (A6\_2016) indicating the 25th percentile, the median and the 75th percentile and interpret the results
5. We want to test differences among means of PRODUCTIVITY\_2016 between family business and other business.
6. Construct a binomial variable called EXPORT\_2016 which assumes value 1 if the business export goods or services during 2016 and 0 otherwise, based on C1\_C2\_2016. Then consider the variable FAMILYBUSINESS\_2016 which assumes value 1 if the business declares to be family owned and 0 otherwise. Test whether EXPORT\_2016 and FAMILYBUSINESS\_2016 are independent, show the cross tabulate table with the p-value and interpret the results.
7. By using a correlation table, show whether PRODUCTIVITY\_2016, employee size (A2\_2016), age of the business (A6\_2016), and family business in 2016 are correlated. The table should show the correlation coefficients and the respective p-value. Interpret the results
8. Given the following function: 𝑙𝑛𝑦𝑖=𝛼𝑖+𝛽𝑖𝑙𝑛𝑥𝑖+𝜀𝑖, is the relationship between y and x linear? What the coefficient 𝛽 may represent?

Questions Multiple Regressions, Dummy Variables and Logistic Regressions

1) Estimate the following regression models

lnPRODUCTIVITY\_2016 = a + b1 AGE\_2016 + b2 lnEMPLOYEESIZE\_2016 + b3FAMILYBUSINESS\_2016 + b4 EXPORT\_2016 (model 1)

lnPRODUCTIVITY\_2016 = a + b1 AGE\_2016 + b2 lnEMPLOYEESIZE\_2016 + b3FAMILYBUSINESS\_2016 + b4 EXPORT\_2016 +b5 X1+ b6 X2 (model 2)

-X1, X2 refer to 2 variables selected from the dataset, that you think are relevant with the dependant variable (lnPRODUCTIVITY\_2016)

a) For each model show the regression table and interpret the estimations coefficients, the R-squared and the F-statistics

b) For each model conduct the test for omitted variables, heteroskedasticity, normality of the errors and multicollinearity and interpret the results of each test

c) If your tests indicate that it is appropriate to re-estimate one or more of the models with robust (Huber-White) standard errors, do so, show the regression table and interpret the estimations coefficients.

d) Indicate which model in your opinion fits better the data and according to which measure of fit.

2) Estimate the following logistic regression models

HIGHPROCUTIVITY\_2016 = a + b1 AGE\_2016 + b2 lnEMPLOYEESIZE\_2016 + b3FAMILYBUSINESS\_2016 + b4 EXPORT\_2016 (model 1)

HIGHPROCUTIVITY\_2016 = a + b1 AGE\_2016 + b2 lnEMPLOYEESIZE\_2016 + b3FAMILYBUSINESS\_2016 + b4 EXPORT\_2016 +b5 X1+ b6 X2 (model 2)

Where:

- HIGHPROCUTIVITY\_2016 is a binomial variable equal to 1 if PRODUCTIVITY\_2016 is above the median value and 0 otherwise

-X1, X2 refer to 2 variables selected from the dataset, that you think are relevant with the dependant variable (HIGHPROCUTIVITY\_2016)

a) For each model show the regression table and interpret the estimations coefficients in terms of marginal effects

b) For each model conduct the test specifications LR test and Wald test and interpret the results

c) Re-estimate the two models by considering only the business operating in the England (NATION\_2016) and for each model show the regression table and interpret the estimations coefficients in terms of marginal effects

17 October 2019