ECOM20001: Econometrics 1

Assignment 1

Student Information

To receive an assignment grade, you must fill out the information in this table and include this table on the front cover page for your assignment. **Only students whose names and student ID numbers are included on the cover page will receive marks for the assignment.** Groups of up to 3 students are allowed.

Name	Student ID Number
Sally Probability	422552
Xiaosong Statistics	653223
Ipsa Regression	294480

Due Date and Weight

- **Submit via LMS by 8 am on 2 September 2022**
- No late assignments will be accepted.
- This assignment is worth 5% of your final mark in ECOM20001.
- There are 40 marks in total.

What You Must Submit via the LMS

- **Assignment answers** no more than 8 A4 pages with 12-point font.
  5 marks will be deducted if your answers exceed 8 A4 pages.
- The **R code** that generates your results. Specifically, copy and paste your R code in an Appendix at the end of your assignment document (e.g., in the .docx file) so it can be viewed and tested by markers. The R code Appendix does not count toward your 8-page answer limit. You may alter and shrink the R code font to less than a 12-point font so that it is easier to read.
  2 marks will be deducted if you do not include your R code.

Additional Instructions

- You may submit this assignment in groups of up to 3. Students in a group are allowed to be from different tutorials.
- You must complete the assignment in no more than 8 A4 pages with 12-point Arial, Times New Roman, Helvetica, Cambria or Calibri font. The assignment cover page does not count toward the 8 A4 page limit.
- To save time, you may copy RStudio output directly into your answers in reporting empirical results. You are also free to create your better-formatted tables based on your RStudio output, which is, of course, good practice in learning how to present empirical results.
- Figures may also be copied and pasted directly into your assignment answers. They may be scaled down in size to meet the 8-page limit, but please ensure that your figures are readable. If they are not, marks will be deducted.
- Marks will be deducted if interpretations of results are incorrect, imprecise, unclear, or not well-scaled. Similarly, marks will be deducted if figures or tables are incorrect, unclear, not properly labelled, not well-scaled, or missing legends.
- When in doubt, work with 3 digits past the decimal throughout.
- This R code in the Appendix at the end of your assignment (as discussed on the previous page) must be commented on and easy for the subject tutors to follow. If the code is not well commented and easy to follow, marks will be deducted. Commenting and code clarity must be at the level of tutorial code, or marks will be deducted.
- Students with a genuine reason for not being able to submit the assignment on time can apply for special consideration to have the assignment mark transferred to the exam at the following link:
    - https://students.unimelb.edu.au/admin/special/

Getting Started

Please create an Assignment1 folder on your computer, go to the LMS site for ECOM 20001, and download the following data file into the Assignment1 folder:

- as1_beer.csv

This dataset contains the following 5 variables:

- state: US state
- year: sample year
- beercons: per capita beer sales in terms of gallons of ethanol equivalent
- beertax: beer tax in terms of inflation-adjusted (real) dollars per gallon
- cigtax: cigarette tax in terms of inflation-adjusted (real) dollars per pack

Data summary

This dataset contains annual information on beer sales, beer taxes, and cigarette taxes across 47 US states from 1981 to 2007.

About the Assignment

In this assignment, we will investigate the empirical relationship between beer sales and "sin taxes" for beer and cigarettes. Understanding how sales (and hence individuals' consumption decisions) are affected by government-set tax rates is a fundamental area of research called *public economics*, that is, the economics of government decision-making.

The data are drawn from the article: Bernheim, Douglas B., Meer, Jonathan, and Neva K. Novarro (2016): "Do Consumers Exploit Commitment Opportunities? Evidence from Natural Experiments Involving Liquor Consumption," *American Economic Journal: Economic Policy*, 8(4), 41-69.

Questions

1. **(4 marks)** Report summary statistics (number of observations, mean, standard deviation, min, max) for beercons, beertax, and cigtax. Interpret the means in words to characterise a "typical" observation in the sample. Also, comment on the min and max values, highlighting the extremes each variable takes on.

2. **(3 marks)** Compute the 95% confidence intervals for the respective means of beercons, beertax, and cigtax.

3. **(3 marks)** Create a new variable, hightax, which equals one if beertax is equal to or greater than the median value of beertax and 0 otherwise. Plot 2 separate densities within the same graph for beercons when hightax = 1 and for beercons when hightax = 0. Interpret the differences in the conditional densities and provide a potential explanation for their differences in means.

4. **(5 marks)** Conduct the following test for difference in means:

   - H0: mean(beercons if hightax =1) = mean(beercons if hightax =0)
   - H1: mean(beercons if hightax =1) != mean(beercons if hightax =0)

   where the symbol "!=" means "not equals." Report the difference means, 95% confidence interval for the difference in means, the p-value for the test, and whether the test implies a statistically significant result at the 5% significance level. Provide a brief interpretation of your findings by computing the per cent change in the conditional mean of beercons when going from hightax = 0 to hightax = 1.

5. **(3 marks)** Construct a scatter plot with beertax on the horizontal axis and beercons on the vertical axis. Use an appropriate single linear regression and abline() R to visualise the relationship in the scatter plot using predicted values from the single linear regression. Also, report the correlation coefficient, corr(beercons, beertax). Does the pattern in the plot align with your findings from questions 3 and 4?

6. **(8 marks)** Run the following (separate) single linear regressions:

   - Regression 1: dependent variable: beercons, independent variable: beertax
   - Regression 2: dependent variable: beercons, independent variable: cigtax

Report coefficient estimates and standard errors assuming homoskedasticity for each regression in a single table.[1] In addition, for each regression, please interpret the magnitude of the predicted change in beercons corresponding to a one-standard-deviation increase in the independent variable. Test whether the predicted change is statistically significantly different from 0 at the 5% level and report the 95% confidence interval for the predicted change.

7. **(4 marks)** Suppose you ran a third (multiple) linear regression:

   - Regression 3: dependent variable: beercons, independent vars.: beertax, cigtax

   The coefficient estimate on beertax is -0.422 in this regression. Provide two additional relevant scatter plots that help explain the change in the direction and magnitude of the coefficient with the coefficient estimate on beertax in Regression 3 compared to Regression 1 from Question 6. In each scatter plot, use an appropriate single linear regression and abline() to visualise the relationship using predicted values from the single linear regression.

8. **(8 marks)** Re-run Regression 1 from Question 6 on two mutually-exclusive subsamples:

   - Sub-sample 1: year <= 1994
   - Sub-sample 2: year > 1994

   Report the regression results using summary() or stargazer(), again assuming homoskedasticity. Discuss whether there is a change in the magnitude of the coefficient on beertax between the respective sets of estimates, whether the respective coefficients on beertax are statistically significantly different from 0 (assume a 5% significance level), and discuss what any change in the coefficient's magnitude and statistical significance means in plain language.

9. **(2 marks)** R-code: we will review and mark your R code as follows:

   - 2/2 if the R code is correct and organised and commented like the solution code for the assignment.
   - 1/2 if the R code is correct but hard to follow or not well commented.
   - 0/2 if the R code is incorrect and/or a complete mess or not submitted.

[1] See discussion of how the summary() command in tute5.R to obtain regression results that assume homoskedasticity. You can use this command, like we did in Tutorial 5, to run your regressions and obtain regression output for interpretation and testing. If you want, you can alternatively use stargazer() as we use in tute6.R around line 220.