

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 name and student ID number 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
Markus Statistics	653223

Due Date and Weight

- **Submit via Canvas by 8am on 13 April 2020**
- No late assignments will be accepted.
- This assignment is worth 5% of your final mark in ECOM20001.
- There are 25 marks in total.

What You Must Submit via Canvas

- **Assignment answers**, no more than 10 A4 pages with 12 point font. 5 marks out of 25 will be deducted if you answers exceed 10 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 that it can be viewed and tested by markers. The R code Appendix does not count toward your 10 page answer limit. You may alter and shrink the R code font to less than 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 one or two. Students in a group of two are allowed to be in different tutorials.
- You must complete the assignment in no more than 10 A4 pages with 12 point Arial, Times New Roman, Helvetica, Cambria or Calibri font. The assignment cover page does not count as one of the 10 A4 pages.
- To save time, you may cut and paste RStudio output directly into your answers in reporting empirical results. You are also free to create your own better-formatted tables based on your RStudio output, which is of course good practice in learning how to present empirical results.
- Figures may also may be copied and pasted directly into your assignment answers. They may be scaled down in size to meet the 10 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 labeled, not well-scaled, or missing legends.
- This R code in the Appendix at the end of your assignment (as discussed on the previous page) must be clearly commented and easy for the subject tutors to follow. If the code is not well commented and easy to follow, 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, and go to the Canvas site for ECOM 20001 and download the following data file into the Assignment1 folder:

- [as1_obama.csv](#)

This dataset contains the following 6 variables:

- **id**: anonymous identification number for an individual donor
- **city**: city where the donor lives
- **state**: US state where the donor lives
- **amount**: amount donated to the Democratic Party in the United States during the 2008 and 2012 elections.
- **share_under25**: percent share of people under 25 years old in the 3-digit ZIP code where the donor lives.
 - 3-digit ZIP codes break up the United States until 929 distinct geographic areas similar in size, for example, to counties
 - so **share_under25** captures how youthful the demographic area/county is in which a given voter lives

In total, this dataset contains information for 86,964 individual voters from across the United States during the 2008 and 2012 presidential elections, which were run by Barack Obama.

About the Assignment

In this assignment, we explore the relationship between the amount individuals' donate to the Democratic Party in the United States and how it relates to the youthfulness of the area in which a given voter lives.

Questions

1. **(2 marks)** Using the `share_under25` variable, create a binary variable called `young` that equals one if `share_under25` is greater than 0.5, and is equal to 0 otherwise. Report summary statistics (mean, std. dev, min, max) for `amount`, `share_under25`, `young`. Interpret each of the sample means in plain language, and comment strictly based on the mean for `young` whether or not 0.5 is the sample median for `young`. Your answer should be no more than 5 sentences long in total.
2. **(3 marks)** Compute 95% confidence intervals for the respective means of `amount`, `share_under25`, `young`.
3. **(4 marks)** Display 2 separate densities for `amount` where `young`=1 and for `amount` where `young`=0 within the same graph. Describe whether each the distributions are (a) skewed; and (b) bimodal. Also comment which part of the respective densities exhibit the largest difference when `young`=1 vs. when `young`=0. Your answer should be no more than 4 sentences long.

4. **(3 marks)** Test the following difference in means:

- $H_0: \text{mean}(\text{amount} \text{ if } \text{young}=1) = \text{mean}(\text{amount} \text{ if } \text{young}=0)$
- $H_1: \text{mean}(\text{amount} \text{ if } \text{young}=1) \neq \text{mean}(\text{amount} \text{ if } \text{young}=0)$

where the symbol “!=“ means “not equals.” For the test, report the difference means, 95% confidence interval for the difference in means, p-value for the test, and whether the test implies a statistically significant result at the 5% level of significance. Provide a brief interpretation of your findings. Your answer should be no more than 3 sentences long.

5. **(3 marks)** Construct another new binary variable called `amount_zero` which equals 1 if `amount`=0 and equals 0 if `amount`> 0. Using this variable test the following difference in means:

- $H_0: \text{mean}(\text{amount_zero} \text{ if } \text{young}=1) = \text{mean}(\text{amount_zero} \text{ if } \text{young}=0)$
- $H_1: \text{mean}(\text{amount_zero} \text{ if } \text{young}=1) \neq \text{mean}(\text{amount_zero} \text{ if } \text{young}=0)$

where the symbol “!=“ means “not equals.” For the test, report the difference means, 95% confidence interval for the difference in means, p-value for the test, and whether the test implies a statistically significant result at the 5% level of significance. Provide a brief interpretation of your findings. Your answer should be no more than two sentences long.

6. **(3 marks)** Construct a scatter plot of `amount` vs. `share_under25` where `amount` is on the vertical axis and `share_under25` is on the horizontal axis. Visually, does there appear to be a positive or negative relationship between `amount` and `share_under25`? Also compute and report the correlation coefficient, `corr(amount,share_under25)`. Your answer should be no more than two sentences.
7. **(5 marks)** Finally, run a single linear regression where the dependent variable is `amount` and the independent variable is `share_under25`. Discuss your results by:
- Interpreting the magnitude of the regression intercept estimate.
 - Interpreting the magnitude of the predicted change in `amount` corresponding to a one-unit increase in `share_under25`.
 - Interpreting the magnitude of the predicted change in `amount` corresponding to a one standard deviation increase in `share_under25`.

Finally, if you were on Obama's data analytics team, which of the latter two interpretations do you think would be more useful for understanding how changes in age levels impacts donation amounts? In total, your answer to question 7 should not be more than one page of text (given font and size restrictions from page 2 above).

8. **(2 marks)** R-code: we will review and mark your R code according to the following scheme:
- 2/2 if R code is correct and organised and commented like the solution code for the assignment.
 - 1/2 if R code is correct, but hard to follow or not well commented.
 - 0/2 if R code is incorrect and/or a complete mess, or not submitted.