

ECOM20001: Econometrics 1

Assignment 1

Student Information

To receive an assignment grade, you must fill out the information in this table and include this page as your assignment cover page.

Name	Student ID Number	Tutor	Tutorial Day & Time	Tutorial Location
Sally Probability	422552	Richard Hayes	Tue 10:15am	The Spot 4452
Markus Statistics	653223	Sahiba Narang	Wed 4:15pm	The Spot 3054

Due Date and Weight

- **Submit via the LMS by 8am on Monday, 8 April 2019.**
- 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 the LMS

- **Assignment answers**, no more than 10 A4 pages with 12 point font. 5 points 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. 5 points out of 25 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 then go to the LMS site for ECOM 20001 and download the following data file into the Assignment1 folder:

- [as1_wine.csv](#)

This dataset contains the following 7 variables:

- **price**: price of a bottle of wine
- **cases**: number of cases of the wine produced
- **score**: WSM wine tasting score from 0 to 100, with 100 being the best taste
- **napa**: binary variable equalling one if the wine is produced from the Napa Valley wine region, and equals 0 otherwise
- **sonoma**: binary variable equalling one if the wine is produced from the Sonoma wine region, and equals 0 otherwise
- **d1995**: binary variable equalling one if the wine is produced in the year 1995, and equals 0 otherwise
- **d1999**: binary variable equalling one if the wine is produced in the year 1999, and equals 0 otherwise

In total, the dataset contains this information for 2060 different wines from wine regions in California and Washington on the west coast of the United States.

About the Assignment

In this assignment, we explore the relationships between price, the amount of wine produced, and wine quality as measured by its WSM wine tasting score. We will also investigate whether wine prices and quality differ in the Napa Valley and Sonoma wine regions compared to all other wine regions on the west coast of the United States. Finally, we will see whether wine prices and quality differ across the 1995 and 1999 years in the Napa Valley and Sonoma wine regions.

Questions

1. **(3 marks)** Report summary statistics for **price**, **cases**, **score**, **napa**, **sonoma**, **d1995**, **d1999**. Interpret each of the means in plain language, thereby characterising a typical wine in the dataset. Your answer should be no more than four sentences long.
2. **(3 marks)** Compute 99% confidence intervals for **price**, **cases**, **score**
3. **(3 marks)** Construct the following density plots:
 - Plot 3a: display 2 separate densities for **price** for Napa wines (**napa=1**) and for Sonoma wines (**sonoma=1**) within the same graph.
 - In other words, the 2 different density plots for **price** for Napa and Sonoma wines should be overlaid with each other such that the densities for **price** in these respective wine regions are easily comparable visually.
 - Plot 3b: display 2 separate densities for **cases** for Napa wines (**napa=1**) and for Sonoma wines (**sonoma=1**) within the same graph.
 - Plot 3c: display 2 separate densities for **score** for Napa wines (**napa=1**) and for Sonoma wines (**sonoma=1**) within the same graph.

Provide a brief interpretation of your findings in Plots 3a, 3b, and 3c. Does the wine price, cases sold, and wine quality appear to be systemically different in one of the regions? Your answer should be no more than four sentences long.

4. **(4 marks)** Construct the following density plots:
 - Plot 4a: display 2 separate densities for **price** for Napa wines (**napa=1**) for years 1995 (**d1995=1**) and 1999 (**d1999=1**) within the same graph.
 - Plot 4b: display 2 separate densities for **score** for Napa wines (**napa=1**) for years 1995 (**d1995=1**) and 1999 (**d1999=1**) within the same graph.
 - Plot 4c: display 2 separate densities for **price** for Sonoma wines (**sonoma=1**) for years 1995 (**d1995=1**) and 1999 (**d1999=1**) within the same graph.
 - Plot 4d: display 2 separate densities for **score** for Sonoma wines (**sonoma=1**) for years 1995 (**d1995=1**) and 1999 (**d1999=1**) within the same graph.

Provide a brief interpretation of your findings in Plots 4a, 4b, 4c, 4d. Do prices and quality appear to be changing across the 1995 and 1999 vintages in the different wine regions? Your answer should be no more than five sentences long.

5. **(4 marks)** Construct the following scatter plots:

- Plot 5a: **score** on the vertical axis, **price** on the horizontal axis
- Plot 5b: **score** on the vertical axis, **cases** on the horizontal axis

Report the covariance and correlation for each plot. Provide a brief interpretation of your findings. Your answer should be no more than three sentences long.

6. **(4 marks)** Test the following differences in means:

- Test 6a: $H_0: \text{mean}(\text{price for Napa}) = \text{mean}(\text{price all other wines})$
vs $H_1: \text{mean}(\text{price for Napa}) \neq \text{mean}(\text{price all other wines})$
- Test 6b: $H_0: \text{mean}(\text{score for Napa}) = \text{mean}(\text{score all other wines})$
vs $H_1: \text{mean}(\text{score for Napa}) \neq \text{mean}(\text{score all other wines})$
- Test 6c: $H_0: \text{mean}(\text{price for Sonoma}) = \text{mean}(\text{price all other wines})$
vs $H_1: \text{mean}(\text{price for Sonoma}) \neq \text{mean}(\text{price all other wines})$
- Test 6d: $H_0: \text{mean}(\text{score for Sonoma}) = \text{mean}(\text{score all other wines})$
vs $H_1: \text{mean}(\text{score for Sonoma}) \neq \text{mean}(\text{score all other wines})$

where the symbol “ \neq ” means “not equals.” For each 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 four sentences long.

7. **(4 marks)** Test the following differences in means:

- Test 7a: $H_0: \text{mean}(\text{price for Napa in 1999}) = \text{mean}(\text{price for Napa in 1995})$
vs $H_1: \text{mean}(\text{price for Napa in 1999}) \neq \text{mean}(\text{price for Napa in 1995})$
- Test 7b: $H_0: \text{mean}(\text{score for Napa in 1999}) = \text{mean}(\text{score for Napa in 1995})$
vs $H_1: \text{mean}(\text{score for Napa in 1999}) \neq \text{mean}(\text{score for Napa in 1995})$
- Test 7c: $H_0: \text{mean}(\text{price for Sonoma in 1999}) = \text{mean}(\text{price for Sonoma in 1995})$
vs $H_1: \text{mean}(\text{price for Sonoma in 1999}) \neq \text{mean}(\text{price for Sonoma in 1995})$
- Test 7d: $H_0: \text{mean}(\text{score for Sonoma in 1999}) = \text{mean}(\text{score for Sonoma in 1995})$
vs $H_1: \text{mean}(\text{score for Sonoma in 1999}) \neq \text{mean}(\text{score for Sonoma in 1995})$

For each 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 four sentences long.