ECOM20001: Econometrics 1

Assignment 3

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 21 October 2022**
- No late assignments will be accepted.
- This assignment is worth 5% of your final mark in ECOM20001.
- There are 35 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 submit your group before 14 October at 8 am; otherwise, you submit as an individual.**
- 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 Assignment 3 folder on your computer, go to the LMS site for ECOM 20001, and download the following data file into the Assignment 3 folder:

- as3_wine.csv

This dataset contains the following 26 variables:

- price: price of a bottle of the wine
- cases: number of cases of the wine produced
- score: WSM wine tasting score from 0 to 100, with 100 being the best taste
- age: how old the wine is in years
- Wine region dummy variables for 9 different wine-producing regions in California and Washington State in the United States, equalling one if the wine is from the region and zero otherwise:
  napa, bayarea, sonoma, scoast, carn, sierra, mendo, wash, othloc
- Wine type dummy variables for 5 different wine types, equalling one if the wine is a particular type and zero otherwise:
  pinotnoir, cabernet, merlot, syrah, nonvarietal.
- Year dummy variables for 10 different years (1990 to 1999), equalling one if the wine is produced in a particular year and zero otherwise:
  d1990, d1991, d1992, d1993, d1994, d1995, d1996, d1997, d1998, d1999.

Data summary

This dataset contains information on 6,979 wines from the United States, including their market prices, cases produced, quality score, age, the region where they are produced, wine type, and year of production.

About the Assignment

In this assignment, we will examine non-linear relationships between age and wine pricing, as well as how these non-linearities differ across wine varieties.

Questions

1. **(2 marks)** Using the ggplot() command in R, produce a scatter plot with price on the vertical axis and score on the horizontal axis. Also, present the fitted values of an estimated cubic regression line in your plot to help visualise the relationship between the variables. Does the relationship appear to be nonlinear? Visually, how many inflection points are prominent in the data?

2. **(5 marks)** Use sequential hypothesis testing to formally determine the appropriate polynomial order for modelling the relationship between price and score. Using stargazer(), report results from estimating cubic, quadratic and linear regression models with price as the dependent variable and (potentially) nonlinear functions of score as the regressors (i.e., there are no other controls in the model).[1] For each step of the sequential hypothesis testing algorithm, state the null and alternative hypothesis for testing a particular polynomial order, report the outcome of the test assuming a 5% significance level, and state whether you should move to the next order or stop. Assume heteroskedasticity-consistent standard errors in conducting your tests, as well as for the remainder of the assignment.[2]

3. **(3 marks)** Based on your sequential hypothesis test results from Question 2, construct another table in stargazer(), with 3 columns reporting different regression results with price as the dependent variable:
   - Column (1): reproduce the model estimation results for the chosen polynomial model for score from Question 2, including the constant.
   - Column (2): same model as in Column (1), but also include all the wine type dummies as controls. Report the coefficient estimates and standard errors for all the wine-type dummies, with the nonvarietal==1 wines being the base group.
   - Column (3): same model as in Column (2), but also include all wine region and year dummies as controls, with the othloc==1 and d1990==1 regions and years further forming the base group. Report the same coefficients and standard errors as in Column (3). Do not report the coefficient estimates or standard errors for the wine region and year dummies, but note in the table that they are controlled for in Column (3).

[1] Note that you may have to construct additional variables from the original dataset to undertake sequential hypothesis testing starting from a cubic model.

[2] Report the output from stargazer() produced by R with heteroskedasticity-consistent standard errors but do not worry about having to update the overall regression F-statistic in the output to account for heteroskedasticity unless otherwise asked to do so in a question.

4. **(3 marks)** Based on the results in the Column (3) model in Question 3, state the null and alternative hypothesis of an appropriate test of whether a nonlinear relationship exists between price and score. Report the outcome of the test assuming a 5% significance level along with the relevant test statistic and its sampling distribution assuming a large random sample.

5. **(3 marks)** Based on the results in the Column (3) model in Question 3, state the null and alternative hypothesis of an appropriate test of whether pinotnoir, cabernet, merlot, syrah wine types have the same associated impact on a wine's price, holding other factors fixed. Report the outcome of the test assuming a 5% significance level along with the relevant test statistic and its sampling distribution assuming a large sample.

6. **(7 marks)** Report the estimated partial effects on price and their standard errors and 95% confidence intervals from increasing score from 80 to 85 and increasing score from 85 to 90, holding all other factors fixed.[3] Interpret the partial effect estimates from the respective changes in score on price. Are your findings foreshadowed by results from any previous questions?

7. **(2 marks)** Estimate the elasticity of price with respect to score. With a multiple linear regression model (e.g., ignore any nonlinearity in the elasticity), and that includes all the same controls as those reported in column (3) of Question (3). Report the regression results in a stargazer() table, reporting all coefficient estimates involving regressors with either price or score or any of the wine-type dummy variables, along with the constant.[4] As in Question 3, do not report coefficient estimates for any other control regressors for wine region or year, but note in the table that they are controlled for. From the point estimates, what is the elasticity?

8. **(3 marks)** Using your Question 8 estimation results, state the null and alternative hypothesis of an appropriate test of whether there is unitary elasticity (e.g., the elasticity is 1). Report the outcome of the test assuming a 5% significance level along with the relevant test statistic and p-value for the test.

[3] That is, you should report a separate partial effect estimate, standard error, and confidence interval for each of these respective changes in score.

[4] Note that in developing your R code, you will be asked in Question 10 below to report the regression results in the same table of results as those from Question 8.

9. **(2 marks)** Modify your regression from Question 7 to allow the elasticity of price with respect to score to differ, depending on whether the wine type is pinotnoir, cabernet, merlot, syrah. Report your Question 7 and 9 regression coefficient estimates in a stargazer() table, reporting all coefficient estimates involving regressors with either price or score or any of the wine-type dummy variables, along with the constant. Based on the point estimates, what are the elasticities of price with respect to score for pinotnoir, cabernet, merlot, and syrah?

10. **(3 marks)** Based on the results in the Column (2) model in Question 10, state the null and alternative hypothesis of an appropriate test of whether the elasticity of price with respect to score is the same for the 4 wine types, pinotnoir, cabernet, merlot, syrah. Report the outcome of the test assuming a 5% significance level along with the relevant test statistic and its sampling distribution.

11. **(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.