



STAT270 AND STAT680  
APPLIED STATISTICS

Assignment Semester 2.

2019

You are expected to write your assignment using R Markdown (see Lecture 6) or MS Word and submit a PDF. You need to submit your assignment via the provided submission link on iLearn.

You may discuss the assignment in the early stages with your fellow students. However, the assignment submitted should be your own individual work.

The R Markdown 'Cheatsheet' from the RStudio team is given [here](#).

In your answers to the questions below, produce the appropriate R output and/or explanation of the steps and results. Don't include any more R output than necessary and include only concise explanations.

### Question 1 [21 marks]

Information was recorded on 376 sampled whiskys and is available in the file `whisky-chemistry.csv` on iLearn.

| Variable | Description                                                                                     |
|----------|-------------------------------------------------------------------------------------------------|
| Taste    | Taste score achieved for the Whisky                                                             |
| Alcohol  | Level of Alcohol in % for the Whisky                                                            |
| Esters   | Level of Esters: Typically add fruity flavour                                                   |
| Lactones | Level of Lactones: Found in the barrel the Whisky is aged in.<br>Contributes to woody flavours. |
| PhenComp | Level of Phenolic Compounds: Typically giving a smokey flavour.                                 |

- Consider first a full regression model with all the predictors used to explain the **Taste** response.
  - [4 marks] Write down the full statistical multiple regression model for quality explained by the other predictors. Carefully define all necessary parameters in your answer.
  - [4 marks] Fit and validate the full regression model.
  - [3 marks] Compute a 95% confidence interval for the regression coefficient (slope) for the **Esters** variable. Explain what the confidence interval represents in the context of the data.
- Consider now a reduced model that can explain the taste response with a reduced set of predictors.
  - [2 marks] Using the appropriate backward model selection method discussed in the course, determine the best regression model for the data.
  - [4 marks] Write down the final model and interpret it in the context of the data.
- For both the full model considered in a) and the reduced model in b), answer the questions below.
  - [2 marks] State the  $R^2$  and explain what it means in the context of the data.
  - [2 marks] Explain why the adjusted  $R^2$  should be reported over the  $R^2$  for assessing the goodness of fit.

## Question 2 [10 marks]

An internet service provider would like to know the levels of performance of its four competitors, which all offer a competitive product near Macquarie Park. A random sample of people that in that area and subscribe to one of those four providers' plans were surveyed to see the average download speed achieved under a standard speed test. The achieved speed and service provider is recorded. The data is available in the file `internet-speed.csv` on iLearn.

| Variable    | Description                                                   |
|-------------|---------------------------------------------------------------|
| TestedSpeed | The measured download speed                                   |
| Provider    | The four internet service providers with levels A, B, C and D |

Define the general contrasts,

$$C_1 = \frac{\mu_A + \mu_B + \mu_C}{3} - \mu_D \quad C_2 = \frac{\mu_A + \mu_B}{2} - \frac{\mu_C + \mu_D}{2}.$$

Conduct two significance tests on the contrasts  $C_1$  and  $C_2$  for an overall significance level of  $\alpha = 0.05$ . In each of your answers:

- [1 mark] State the appropriate hypotheses.
- [1 mark] Calculate the observed value of the (raw) contrast.
- [2 marks] Calculate the standard error of the contrast.
- [1 mark] State the null distribution of the test statistic.
- [2 marks] Compute or give the best approximate bound on the P-Value.
- [3 marks] Draw your overall conclusions (both statistical and contextual).