MT5762: Individual Project Instructions

Steve Drasco, School of Mathematics and Statistics, University of St Andrews

Introduction

Read these instructions completely before you begin.

Read the associated article available on Moodle (**Herald_article.pdf** *New Zealand Herald* 4/07/2000).

A forensic scientist has approached you to analyse his data at the end of an experiment. He has been growing Cannabis plants in different soil types and analysing the chemical composition of the leaves. He is specifically interested in whether there are differences between those plants grown in standard store-bought potting mix, and those that are grown in general outdoor soil. It is thought that if a plant's chemical composition is very specific to its soil type, then this may assist in the prosecution of persons accused of drug production and distribution.

The data for this project is contained in the file **potplants_MT5762.csv**. The variables contained in the dataset are the reference codes (not of interest to us), the type of soil used (there are 4 types), and the levels of various chemical elements measured in the leaves. The variables are:

- **Sample Name** Reference number of the samples.
- **Group** The four soil types: Potting mix (pm) and three locations about New Zealand. These are bhb, mb, and nth, being: Blockhouse Bay (Auckland suburb), Mission Bay (Auckland suburb) and Northland (a northern region).
- **Mg-Th** The 38 chemical elements measured in the leaves. Note the units of measurement are not given, so you need not declare them in your work.

What to do

- Read the data into your computer and explore it.
- Work out answers to each of the four questions in the next section of these instructions.
- Briefly **answer the questions below in the form of a 1-page PDF file**. Use section headings or other indicators of where the answer to each question begins and ends.
- On additional pages of that same PDF file, include any plots, figures, or tables that you produced and referenced in your answers. Do not crowd these pages (in some cases, one object per page is appropriate). Make sure that each object is complete, and that any text is large enough to be easily read.
- On further additional pages of the same PDF file, Include all of the code you wrote and used.
- Add a single cover page to the beginning of your PDF file indicating your name and student ID number.
- Proofread your PDF file, especially the page with your answers to the questions.

- Name your PDF file **Project1.pdf**, and upload it on the moodle site for our module (do not upload anything to MMS). Only upload one single PDF file.

Questions

Each of the following questions is worth the number of marks indicated in square brackets at the end of the question.

1. You will analyse only 5 of the chemical elements in the data. The choice of which 5 elements you analyse is yours. Which 5 elements did you choose, and why did you choose them? **[1 mark]**
2. Do the data indicate differences in the elemental composition of Cannabis leaves grown in different soil types? **[5 marks]**
3. Are some of the elements related to one another in terms of their levels in the sampled leaves? Present no more than 5 pairs of elements, choosing pairs with interesting relationships. **[5 marks]**
4. Do you think the results of this experiment will ultimately allow the determination of what soil the plants were grown in, just from the elemental composition of the leaves? Be sure to consider what **population(s)** we are generalising to. Give a detailed response i.e. if you do not think the aims of the study can be achieved, explain why. **[5 marks]**

Statistical Requirements in Assessment of Answers

The following properties will impact how many marks your answers earn.

- You are expected to use formal statistical tests and estimates.
- Use a type-1 error of 5% to determine statistical significance. Confidence intervals for estimates should be given and be 95%.
- Check all assumptions formally, and explain what any violations mean for the validity of conclusions.
- Always be clear what hypotheses are being tested.

Remaining marks

There are a total of 20 marks possible for this project, 16 of which are allocated in the Questions section above. The remaining 4 marks will be allocated as follows

- Figures and tables were easy to find and to read (all text was large enough, references from answers to questions were easy to follow, etc). **[2 marks]**
- Code was easy to find and to read (it was tidy, easily followed, and easy to find at the end of the PDF file). **[2 marks]**

Deadline

Submit your PDF file via moodle before 23:59pm, Friday, 2nd of October. Penalties for late submissions will follow [school policy](an initial penalty of 15% of the maximum available mark, then a further 5% per 8-hour period, or part thereof). Because of the significant initial penalty of 15%, you are strongly advised to not submit your file at the last minute. Assume there will be glitches in the upload process, and plan to upload hours in advance of the deadline.