Statistical Thinking (ETC5242)

Assignment 1 – extra question for ETC5242

Semester 2, 2024

Instructions

This assessment task is to be done individually. It is not a group task.

This task is only for students enrolled in ETC5242. It is an extra task that forms part of Assignment 1, in addition to the group task that is labelled as "Assignment 1" on Moodle.

Since this is an individual task, your submission should be entirely your own work.

Preparing your submission

Your responses should be written in an R Markdown file that compiles to produce the desired results.

Complete the task using a combination of written text and output from R, in the format of a mini-report. All answers need explanation and justification.

All plots must be properly labelled and explained. For example, the definition of each axis should be clear and also what each visual element (points, lines, etc.) represents.

Show and evaluate all code chunks using the `echo = TRUE` and `eval = TRUE` chunk options, and format the output so that it does not run off the page when printed. You can also suppress all other messages and warnings, as per the following command (to be included in the first code chunk of your R Markdown file):

```
knitr::opts_chunk$set(echo = TRUE, eval = TRUE, warning = FALSE,
                      message = FALSE, error = FALSE)
```

Anything that is not part of your answer should **not** be included in the R Markdown file, even if it is not evaluated or does not appear in the rendered PDF file.

Your name and student ID number must be stated in the YAML section of the R Markdown file and on any other files submitted.

Submission

This task is due on Thursday 10 October 2024 at 23:55.

You need to upload **two (2) separate files** via Moodle: the Rmd file and the corresponding PDF file. You may create an HTML or Word document from your Rmd file, but you **must** convert this to PDF for submission.

Your files should be named according to the following templates:

`FirstName_Surname_studentID_A1_extra.Rmd`
`FirstName_Surname_studentID_A1_extra.pdf`

Replace `FirstName` and `Surname` with your own names, using the same spelling as shown in Moodle. Replace `studentID` with your student ID number. For example: `Damjan_Vukcevic_123456_A1_extra.Rmd`

Submission of incorrect files or file names will lead to loss of marks.

Your files must be submitted **separately**, rather than combined together as a ZIP file or similar.

Marking

There is a total of 12 marks available for this task.

Your overall mark for Assignment 1 for ETC5242 consists of your mark for the group task (maximum 100 marks) plus your mark for the extra task described here (maximum 12 marks). This overall mark contributes 20% to your final score for ETC5242.

The 12 marks will be awarded as follows:

- Appropriate methods used to answer the question [4 marks].
- Clear and adequate description of methods and conclusions [4 marks].
- Quality and neatness of presentation [2 marks].
- Reproducibility [2 marks]. Whether the submitted Rmd files compile as submitted without error.

Task

This task is a statistical exploration of the phenomenon called the "wisdom of the crowd", which refers to the idea of taking the collective opinion of a group of people.

As a concrete example, in one of the lectures this semester we asked each student to guess the number of chocolates that were in a jar displayed at the front of the lecture theatre. The individual guess varied substantially. We could take the mean of the guesses as a way to combine everyone's opinions together to form a collective opinion.

Broadly, we want you to explore and describe what conditions would typically lead to the mean of the guesses (the collective opinion) to be more accurate than the individual guesses. To make this more specific, consider the following situation:

- The individual guesses can be modelled as iid random variables that follow a $\mathrm{N}(\mu, 1)$ distribution.
- The true value is 50.
- There are 15 individuals.

Let A be the probability that the mean guess is closer to the true value (of 50) than **all** of the individual guesses.

Your task is to answer the following question:

- For what values of μ (to 2 decimal places) is it true that $A > 0.1$?

You can use any method to solve this, including simulation.

Your response should include at most 3 paragraphs of text, and any R code and output (including plots and tables) as you judge necessary to answer the question.