STAT344: Group Project

This is a group project, with 3 to 5 students in each group. Each group should have a *leader*, chosen by yourselves. You are free to choose your group members, not necessarily the same as your lab groups, and you can have group members from different lab sessions. You may consider choosing group members with different strengths, e.g., A is good in writing, B is good in analysis, C is good in R, etc. This project has **two parts**: Part I (80%), Part II (20%).

PART I

You will need to collect data by yourselves using *two different* sampling methods/designs, such as a simple random sample (SRS) and a stratified sample. The two samples should have the same sample sizes from the same population, and the sample size should not be too small, (say) at least 50. For example, we may be interested in the *average number of pages* of all the mathematics and statistics books in a UBC Library, i.e., the parameter is the population mean. We may obtain a SRS of 50 books, check the number of pages of each book, and obtain the average number of pages of the books in your sample. We may also obtain a stratified sample, say 20 randomly selected books in mathematics and 30 randomly selected books in Statistics, and combined them into a single sample. If you choose stratified sampling, please state clearly how you choose the strata and *why*, how you decide sample size for each stratum, etc. You should also consider a *second different parameter*, say, *proportion of books over 200 pages*. Of course, you may also consider sampling people or other units of interest.

It is fine to use available data on internet. For example, if you find an interesting "big data" on internet, you may treat the "big data" as "population" and then do the sampling to obtain a smaller sample (e.g., a SRS and a stratified sample). The key is to clearly describe your *sampling procedures* and briefly justify your approaches. In other words, you cannot use available samples without doing the sampling yourselves.

Once you obtain the samples, you should analyze the data and write a comprehensive report. Your report should contain the following materials:

- A cover page: list the *names* and *student numbers* your group members, the *group leader*, and indicate clearly the role and *contributions* each group member makes in the project;

- Clearly state the objectives, background, and why the problem is important or of interest.

- Clearly describe how and where the samples are obtained (e.g., how you obtain a SRS, how you choose strata, how you determine the sample sizes, etc), what is the targeted population (e.g., all science books in UBC Koerner Library), and the parameters of interest. Please consider two different types of population, such as one being continuous (e.g., number of pages of a book) and one being binary (e.g., whether the book is over 200 pages or not), based on the same samples.

- In data analysis, please show estimates, standard errors, and confidence intervals. Please *interpret the results*, and state any assumptions you have made. Discuss advantages and disadvantages of different methods being used.

- Final conclusions and discussion: overall conclusions, any limitations of your conclusions and sampling methods, whether your conclusions can be generalized to larger or other populations.

Requirements for your report:

- Your report should have at most 10 pages, plus Appendix (data, R code, etc).

- Your report should be typed. No hand-writing. Upload a pdf or word file to Canvas.

- The quality of the writing is an important evaluation criterion.

- Put data and R code in an Appendix.

- Each group should work *independently*.

PART II

Read the paper "The Emperor's New Tests" (Perlman and Wu, 1999, *Statistical Science*, 14, pages 355-369) posted on Canvas. For this paper, you just need to understand the main points, not the technical/mathematical details (*hint: you can mostly focus on the first and the last sections of the paper and safely skip the other sections*). Write **one paragraph** to a non-statistican *in words* explaining the key message of the paper.

SUBMISSION

Each group submit **one** report (upload to Canvas) by **Sunday, November 12, 2023** (midnight). Please use your group leader's name as the file name in the format of

LastName-FirstName.pdf

,
e.g., Wu-Lang.pdf. The report can be submitted by any member of your group.