NATIONAL UNIVERSITY OF SINGAPORE

Department of Statistics and Data Science

ST1131 Introduction to Statistics and Statistical Computing

(Semester 1 : AY 2021/2022)

Individual Assignment

Due Date: 15 October 2021 by 5 pm

INSTRUCTIONS TO STUDENTS

1. Students are supposed to submit the answers on time. Any submission after 5pm of the due date are not accepted.
2. 10% of the given mark will be deducted for each 2 hours late in submission.
3. **Students are required to complete this assignment individually.**
4. All submission is done online.
5. Your submission has **two separate files**. One is a .pdf file of report, and the second file of R code. Make sure that there is no error when the graders open and run your R code file.
6. Be sure to lay out systematically the various parts and steps in your working.
7. Your submission files should be named as `A0123456B.pdf` and `A0123456B.R` where A0123456B is your student matric number.

A dataset documents all United Nations General Assembly votes since its establishment in 1946 till 2015. The data is broken into three different files: the first (`resolutions.csv`) lists each UN resolution, subject, and vote records; the second (`states.csv`) records individual member state votes per resolution; the third file (`votes.csv`) provides an annual summary of member state voting records with affinity scores and an ideal point estimate in relation to the United States.
`https://www.kaggle.com/unitednations/general-assembly`

In this assignment, the answers should be derived based on the information given in the third file. However, information in the other two files can help students to understand the information in the third file (as supplementary).

Import the file `votes.csv` into R. Some points to take note about the information given in this file:

- There are many countries in the record, however at the first assembly in 1946, only some countries were the member of UN while other countries were not. The number of members of the UN changes over the years.
- Each assembly has many different resolutions for voting. Each resolution was named by a "vote_id".
- Each country has a state name and a state code for it.

QUESTIONS

Part I: Exploring the information in the data

1. Report the size of the file (number of rows and number of columns).
2. For each resolution, how many outcomes for the vote and what are the outcomes and their meaning?
3. How many assembly were recorded in total (from 1946 to 2015)?
4. How many resolutions were voted in total (from 1946 to 2015)?
5. For the first two assembly sessions (in 1946 and in 1947), how many resolutions were voted in each assembly?
6. In 1948, the third assembly session was held. How many resolutions in this assembly did the United States of America vote "yes" for?
7. Singapore joins the United Nations on 21 September 1965. Which assembly was held in this year? In that assembly session, how many resolutions did Singapore vote "yes" for?
8. Which three countries voted 'yes" the most and the least often, and what are their proportions? (*Hint*: if the outcome of the vote is 8 or 9, then it could be treated as "Abstain")
9. Which assembly session in which year has the largest number of resolutions for voting? Report the number of resolutions in that year.
10. Which assembly session in which year has the smallest number of resolutions for voting? Report the number of resolutions in that year.

Part II: The United State of America (US) and Canada

11. Derive the proportion of voting "yes" to the resolutions of the US, for every assembly session. Name the vector of these proportions as "us". Report this vector. In which year (which assembly session), the US gave the highest "yes" proportion to the resolutions? Report that highest proportion.
12. Derive the proportion of voting "yes" to the resolutions of Canada, for every assembly session. Name the vector of these proportions as "ca". Report this vector. In which year (which assembly session), Canada gave the highest "yes" proportion to the resolutions? Report that highest proportion.

13. Derive the correlation coefficient for the proportion of voting "yes" for the US and Canada. Report the correlation value.

14. Plot a scatter plot for the two vectors, us and ca. Give your comment on the relationship of the US and Canada when giving their support to the resolutions.

END OF ASSESSMENT