STA321: Week #02 Assignment

Due: Sunday, 11:30 PM

Contents

1   HONOR CODE

Integrity is a core value at WCU. All students must familiarize themselves with the Student Academic Integrity Policy and pledge to observe its provisions in all written and oral work. Please review WCU's Undergraduate Student Academic Integrity Policy and understand what Honor Code violations might look like in this class.

This is a project-driven advanced topic course, all submitted work must be yours. You are encouraged to collaborate with your classmates on the course materials, but not the assignments! Here are some suggestions on how to avoid academic dishonesty:

- Read the instruction of each assignment carefully to understand the specific requirements. In this semester, you will work on the same topic on each weekly assignment but with different data sets. All assignments must be completed independently. There will be NO group project for this course.

- Never share your assignments with others. If classmates have a question, try to help them. Sharing your assignment with others will not teach them anything and you might be accused of academic dishonesty since I will check similarities between your submitted work.

- If you paraphrase or summarize what someone else said or from internet blogs and published papers, you must provide me the source and make sure the write-up and code must be yours.

- I will provide code for all analyses in the class notes. You can use the algorithm and logic behind my code but try to avoid copying and pasting my code and only change the name of the data. You are expected to give meaningful names to R objects by yourself.

- If you experience unusual circumstances and cannot meet the deadline, please reach out to me to request an extension for an assignment. Late work without prior arrangements can only receive up to 80% of the credit. Start working on your assignment earlier and Don't wait until the night before to begin an assignment.

2   Data Set Description

This data set includes the **percentage of protein intake** from different types of food in countries around the world. The last couple of columns also includes counts of obesity and COVID-19 cases as percentages of the total population for comparison purposes.

Data can be found at kaggle.com. I also upload this data set on the course web page w02-Protein_Supply_Quantity_Data.csv.

```
url="https://pengdsci.github.io/STA321/ww02/w02-Protein_Supply_Quantity_Data.csv"
protein = read.csv(url, header = TRUE)
var.name = names(protein)
kable(data.frame(var.name))
```

var.name
Country
AlcoholicBeverages
AnimalProducts
Animalfats
CerealsExcludingBeer
Eggs
FishSeafood
FruitsExcludingWine
Meat
MilkExcludingButter
Offals
Oilcrops
Pulses
Spices
StarchyRoots
Stimulants
Treenuts
VegetalProducts
VegetableOils
Vegetables
Miscellaneous
Obesity
Confirmed
Deaths
Recovered
Active
Population

3 Instruction of Assignment for Week #2

- Study the class note and make sure you understand the bootstrap confidence interval.

- Use the above code to read data in R. and explore the variables in the data set. Choose a variable from the data set to construct confidence intervals of the mean of that variable.

- Please go to the D2L discussion Board to tell your class which variable you will use for your assignment. If that variable has been chosen by your classmates, then go back to the data set to choose other variables. I want to make sure everybody will work on the different data set. To go to discussion board: **communications =>Discussions**

- When you finish your analysis and write-up, go to D2L to submit your work: **Assessment => Assignments**.

- The format should be similar to that of the class note. You can use the template of this RMD. That is, you can copy-and-paste the portion of this document above **# HONOR CODE**.

- You MUST submit your RMarkdown and one of HTML, PDF, or WORD via D2L but **Not email!**

4 Assignment Problems

1. Open an R Markdown document, please feel free to use the template in my RMD for the class note.

2. Write a few sentences to describe the variable you are going to use in your analysis.

3. Construct a confidence interval of the mean of the variable use methods you learned in your previous courses. For example, MAT 125, STA 200, STA 319, and/or STA320.

4. Use the Bootstrap method introduced this week (see the class note on the course web page) to construct the bootstrap confidence interval for the mean of the selected variable.

5. Plot the bootstrap sampling distribution of the sample mean.

6. Compare the two confidence intervals and interpret your findings and interpret the intervals.