Econ 184b, Econometrics, Assignment 2

The raw .rmd file for this assignment are available at: https://raw.githubusercontent.com/f1kidd/Econ184/main/Assignment/Assignment_2.Rmd

Note: For the math problems, you can either (1) type your solution and compile it with RMarkdown; (2) type your solution use another word-processing software and convert it into a PDF file; (3) write your solution on paper, scan it and make it into a legible PDF file. TAs can decide, at their discretion, that the submitted file is illegible and thus give zero credit to a question or the entire problem set. If you type your solutions (options 1 and 2 above), you will get **a bonus equal to 5% of your performance** on the problem set. No credit will be given if you only report the final answers without showing intermediate steps whenever appropriate.

For the programming problems, you need to submit both the compiled pdf/html file and the RMarkdown (.rmd) file on LATTE. You will get **an additional bonus equal to 5% of your performance** if your RMarkdown file is completely reproducible with minimal altercation (install packages, etc.). That is, our team (or anyone else) should be able to recompile your Rmarkdown file and reach *the same* result. You need to explicitly write out your answers - just showing programming outputs will receive zero credit.

You will be receiving a total of 15% bonus if you submit one single pdf/html compiled directly from Rmarkdown, along with the .rmd file.

Question 1:

Compute the following probabilities. Write your final answer in terms of Φ (the cdf for the standard normal distribution). For example, if your final answer is $1 - \Phi(0.3254)$, then that's it and you do not need to find exactly what number is $\Phi(0.3254)$ from the Z-table / using qnorm. Of course, you still need to show the intermediate computations, formulas or processes.

a. $Y \sim N(1, 4)$. Find P(Y $\leq$ 3).
b. $Y \sim N(3, 9)$. Find P(Y $>$ 0).
c. $Y \sim N(50, 25)$. Find P(40 $\leq$ Y $<$ 52).
d. $Y \sim N(5, 2)$. Find Pr(6 $\leq$ Y $\leq$ 8).

Question 2:

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Construct a 95% confidence interval for p.
b. Construct a 99% confidence interval for p.

Question 3:

Let $Y_1, ..., Y_n$ be i.i.d. random variables following N(8,4), where *n=27*. Let $\bar{Y}$ be the sample average.

a. Find the probability that $\bar{Y} < 10$.
b. Using 10,000 repetitions, verify your answer in part (a) using R. Your code could look like the following:

```
N = 10000
n = 27
ybar_store = c()
```

```
for(i in 1:N){
  # sample from N(8,4) with a sample size n=27
  # store the sample mean
}
# calculate the proportion that the sample mean is less than 10.
```

Question 4

Use the *Boston* data set in the MASS library, which contains census-tract-level housing values in the suburbs of Boston in 1973. Documentations for the dataset can be found by typing *"Boston"* in the help session after you import the package *"MASS"*.

a. Plot a histogram of median owner-occupied housing value in the area.
b. Calculate the mean, median, minimum, and maximum housing value in the Boston area
c. Are housing values adjacent the Charles River higher? Explain.
d. Plot the bivariate relationship between the property tax rate and the pupil-teacher ratio. Are they correlated?
e. Create an indicator variable showing whether the tract is above or below median crime rate of the city. What is the conditional variance of median housing values for tracts above the median crime rate? Below the median crime rate?
f. If one wish to explore the bivariate relationship between crime rate and housing value in a regression framework. Is the model homoskedastic? Explain.

Your code could look like the following (where plotting is done through the *ggplot2* package:

```
library(MASS)
library(ggplot2)
data = Boston
## a. (note: you need to place variable names inside aes().)
# ggplot(data=data)+geom_histogram(aes(x))

## d. (note: you need to place variable names inside aes().)
# ggplot(data=data)+geom_point(aes(x,y))
```