Homework 6

Instructions: To receive credit, you must submit your assignment to Canvas before **6pm, Friday, March 13**. The file submission must be a knitted .html file, made using RMarkdown. The code you used to answer the questions should be included in your file. You do not need to submit your .rmd file.

1. A simple random sample was taken of 44 water bottles from a bottling plant's warehouse. The dissolved oxygen content (in mg/L) was measured for each bottle, with the results below.

```
library(tidyverse)
doc_data <- tibble(oxygen_content = c(11.53, 8.35, 11.66, 11.54, 9.83,
                                      5.92, 7.14, 8.41, 8.99, 13.81,
                                      10.53, 7.4, 6.7, 8.42, 8.4, 8.18,
                                      9.5, 7.22, 9.87, 6.52, 8.55,
                                      9.75, 9.27, 10.61, 8.89, 10.01,
                                      11.17, 7.62, 6.43, 9.09, 8.53,
                                      7.91, 8.13, 7.7, 10.45, 11.3,
                                      10.98, 8.14, 11.48, 8.44, 12.52,
                                      10.12, 8.09, 7.34))
```

   a. Create a histogram and QQ plot of the data. Is it plausible that the data was drawn from a population that follows a normal distribution? Explain your answer.

   b. Find a 98% confidence interval for the population mean dissolved oxygen content. Show your work.

   c. Interpret the interval you created in (b).

2. Consider the following data that represents a random sample out of a very large population:

```
data_sample <- tibble(x = c(0.43, 4.89, 5.29, 5.16, 0.59, 7.50, 4.48, 6.22))
```

   a. Create a QQ plot of the data. Do you think it is reasonable to assume that the population distribution is normal? Explain your answer.

   b. Denote μ to be the population mean. Regardless of your answer to (a), use R to perform the bootstrap with 2000 resamplings to create a 90% CI for μ. Show the estimated distribution using a histogram. Since answers will differ for this question, it is critical that you show your R code and output to get full credit. Use the code from discussion March 4th/5th.

   c. Regardless of your answer to (a), assume the population distribution is normal and use that fact to create a 90% CI for μ.

   d. Compare your answers to parts (b) and (c). Which one do you think is more correct, and why?

3. Suppose you are in charge of inventory maintenance at a bicycle shop. **One of your jobs is to ensure that the tire pressure in each of the display bicycles is between 65-85 PSI (pounds per square inch)**. If the pressure is too low, then there is a risk of wheel damage when a customer rides out on one. On the other end, if the pressure is too high, there is a (small) risk of the tire exploding! Of course, you don't know the true pressure of any particular tire. Instead, you have the output from your pressure gauge. This will be similar to the true pressure, but not necessarily the same. Laboratory testing of the particular gauge you use has shown that there is a ± 2 PSI error margin, and so to be careful, you decide that you will adjust the pressure on any tire that has a measured pressure above 81 PSI or below 69 PSI, giving you 2x the error margin on either side. Previous testing shows that this procedure will give the following results:

	Measure inside 69-81PSI	Measure outside 69-81 PSI
PSI within 65-85 PSI	5000	105
PSI outside 65-85 PSI	3	216

   a. State (in words) the appropriate null and alternate hypothesis.

   b. Calculate the value of α. What situation does the Type I error rate represent here?

   c. Calculate the value of β. What situation does the Type II error rate represent here?

   d. Calculate the *Power* of this test. What situation does test *Power* represent here?

   e. What is the rejection region for the test you are conducting?

   f. Practically speaking, which error seems more problematic: a Type I error or a Type II error?

   g. If we wanted to decrease the Type I error rate, how would we change the rejection region?

   h. If we wanted to decrease the Type II error rate, should we increase or decrease the rejection region?