Exam #2 STA-209 Sections 03, 05

Name:_____

Directions

- Several questions have a *suggested* number of sentences for your answer. This is to help indicate the scope of solution I am looking for (i.e., you do not always need every single detail) and to discourage you from "information dumping"
- Information that is included with your answer that is not relevant to the problem will not help you but *will still be graded for correctness*. In other words, including more information than is asked for can generally only hurt you
- You **do not** need to write in complete sentences: bullet points are completely acceptable and even preferred

Formulas

$$\overline{X} \sim N(\mu, \sigma/\sqrt{n})$$

$$\hat{p} \sim N\left(p,\ \sqrt{\frac{p(1-p)}{n}}\right)$$

Question #1 (conceptual)

Part A: In 1-2 sentences, explain the meaning of "95% confidence" in the term "95% confidence interval." Avoiding using the words "confident" or "confidence" in your explanation.

Part B: In 2-3 sentences, explain the relationship between a *population* and a *sample* and how each of them are used within the statistical framework.

Part C: In 2-3 sentences, explain what is meant by a "sampling distribution." Why might we want to know this, and where does variability in the sampling distribution come from?

Part D: Write out the formula for a t-statistic (similar to a z-score). Explain how each term in the formula changes the value in the t-statistic

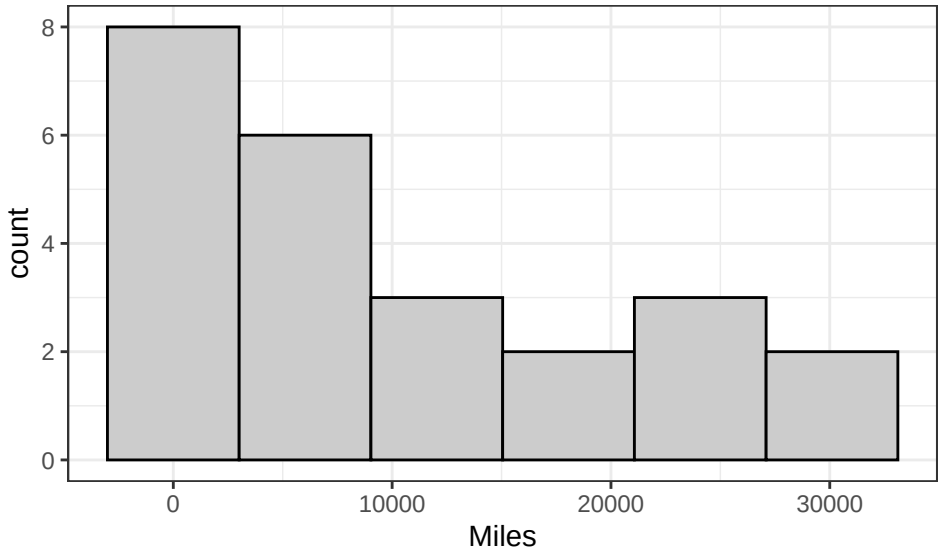
Part D: Write out the formula for a confidence interval for the mean using the Point Estimate $\pm$ Margin of Error method and explain what impact each of the terms has on the width of the confidence interval.

Part F: In 2-3 sentences, explain how the central limit theorem justifies the use of the 68-95-99 rule for the sampling distribution of the mean.

Question #2

The data below includes the passenger miles flown by commercial airlines in the US from 1937 to 1960. As with the Grinnell Rain data, we are interested in finding and computing the *centrality ratio*, defined as

$$CR = \frac{\text{mean}}{\text{median}}$$

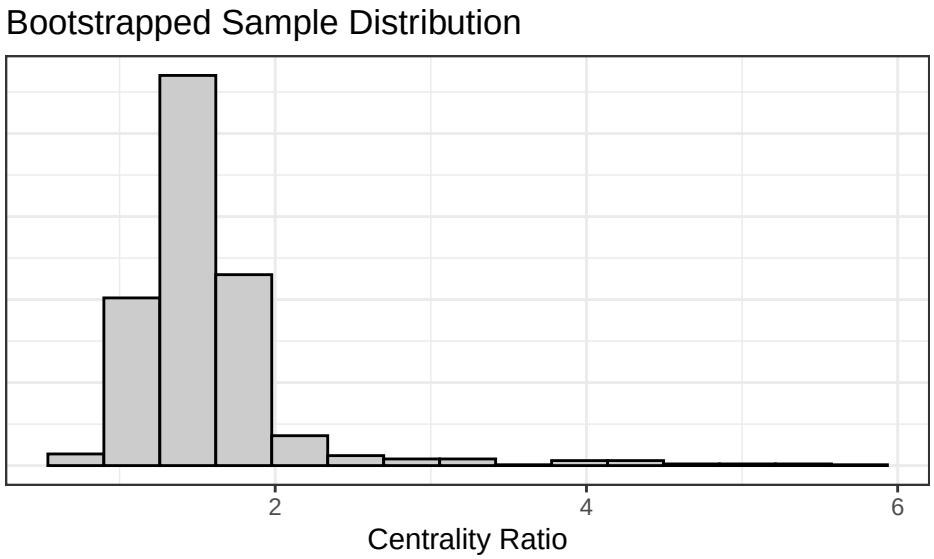



```
data.frame(Miles = airmiles) %>% summarize(Mean = mean(Miles),
                                           Median = median(Miles),
                                           SD = sd(Miles),
                                           N = n())
```

```
##    Mean Median    SD  N
## 1 10528   6431 10033 24
```

Part A: Using the summary information provided, provide the *point estimate* for the centrality ratio.

Part B: Given below is a histogram of the bootstrapped sampling distribution for the centrality ratio. Based on this, would the 69-95-99 rule be applicable here for finding a 95% confidence interval? Why or why not?



Part C: Given below are the deciles of the bootstrapped sample distribution. Using these, construct an 80% confidence interval for the value of the centrality ratio.

10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1.0939	1.2272	1.3414	1.4225	1.5011	1.5601	1.6287	1.7227	1.8974	5.745

Part D: Based on your results for Part C, what would you conclude about the skew of the US commercial airline mile data?

Question #3

Critical ICU patients are sometimes put on mechanical ventilators to assist or replace the function of breathing. A serious side-effect of this procedure is ventilator-associated pneumonia (VAP). A new type of ventilator was introduced claiming that the rate of VAP on the new machines was 5%. To test this, in a study of patients who were mechanically ventilated using the new machines in the ICU, researchers found that 63 out of 472 patients using the new machine developed VAP.

Part A: State the null hypothesis regarding the incidence of VAP using the new ventilator machines.

Part B: In the study conducted, what proportion of patients developed VAP?

Part C: Using the formula on the front page, compute a 95% confidence interval for the proportion of patients who will develop VAP using the new machines (Note: $C = 1.96$)

Part D: Based on what you did for Part C, what is the Type I error rate of this test?

Part E: Based on your results in Part C, what conclusions do you draw with respect to the null hypothesis?