Assignment #2: Sampling, chi-squared, F-distributions

Scenario: One supplier provides upholstery fabric with an average durability of 74,283 DR and a standard deviation of 4676 DR (variance = 21,864,976). Another supplier provides fabric with a lower average durability of 74,200 DR and a lower standard deviation of 4500 DR (variance = 20,250,000). All measurements are based on a sample of n=61. You must decide which supplier to purchase from based on the variance of their fabric.

1. Complete the following to conduct an F-test (using a 0.01 significance level) comparing the variances in durability of the fabric from the two suppliers.

   Null hypothesis: _____ Alternative hypothesis: _____

   Calculate the F-statistic comparing the larger sample variance to the smaller sample variance:

   F =

   Numerator degrees of freedom = _____ Denominator degrees of freedom = _____

   The F-statistic you calculated from the two samples of n=61 observations is one of an **infinite** number of F-statistics you *could have* calculated (from different random samples from each population). Assuming the null hypothesis is true, sketch the distribution of all possible F-statistics you could have calculated from different random samples of size n=61.

   _____

   Distribution of all possible F-statistics if null hypothesis is true

   Locate your observed F-statistic (the one you calculated earlier) on that distribution. Use a computer (online or on your calculator) to estimate the p-value. Interpret that p-value (or draw an appropriate conclusion).

   p = _____ Interpretation: _____

   _____

   What assumptions, if any, are you making when you conduct this F-test?

   Assumptions: _____

Scenario: In class, we took a look at the hours St. Ambrose students report studying each week. The distributions for both freshmen and sophomores were heavily skewed to the right. We also saw:

528 freshmen reported spending an average of 10.6 hours per week studying (with a std. deviation of 8.4
234 of those same students reported studying 13.7 hours per week (with a standard deviation of 10.0)

2. Test the claim that sophomores have a greater variance in hours spent studying per week. Write out your hypotheses, sketch the sampling distribution, estimate the p-value, and write out any conclusions you can make.

3. To use this F-test, we need to make some assumptions. Which assumption has been violated in this example?

Scenario: The reported hours studying per week from both freshmen and sophomores have been combined into a datafile you can download at: http://www.bradthiessen.com/html5/data/hours2.csv

Your goal will be to construct a 95% confidence interval for the population standard deviation (of hours studying per week).

4. Let's construct our first confidence interval using bootstrap methods. To do this:

- Copy the data at http://www.bradthiessen.com/html5/data/hours2.csv
- Go to this bootstrap applet and choose to construct a **Bootstrap Dotplot of StDev**
  Applet: http://lock5stat.com/statkey/bootstrap_1_quant/bootstrap_1_quant.html
- Click **edit data** and paste the data into the pop-up box

You should now see the data graphed on the right. Record the sample size and standard deviation below:

Sample size = n = _____ Sample std. deviation = stdev = _____

- Generate at least 10,000 bootstrap samples and find the limits of the 95% confidence interval.

95% confidence interval for the population standard deviation = _____

5. This time, let's construct a confidence interval using the chi-squared distribution. This is the formula we used in class. (Oh, and use the sample size and standard deviation you recorded in the previous question).

95% confidence interval for the population standard deviation = _____

6. Record some information about each of the confidence intervals you just constructed:

| Method | Interval | Width of the interval | Center of the interval |
|---|---|---|---|
| Bootstrap | _____ | _____ | _____ |
| Chi-squared | _____ | _____ | _____ |

Based on this information, which interval (bootstrap or chi-squared) do you think it most appropriate for this scenario? Briefly explain:

7. Now construct a 90% confidence interval for the population standard deviation. You can use either bootstrap methods or the formula with chi-square. How does the width of this interval compare to the width of the 95% CI?

90% confidence interval for the population standard deviation = _____

This interval is **WIDER** **MORE NARROW** than the 95% confidence interval

8. Go back to question #2 in this assignment (comparing variances in hours studying for freshmen and sophomores). Briefly explain how you could use bootstrap methods to construct a 95% confidence interval for the ratio of the two variances.