Library Sampling Project Analysis Plan | v.o.91 | FA19 name: _____

1. Populate the following data table with the results of your sampling

	Library sub-section 1	**Library sub-section 2:**
Call Number prefix:		
Name of sub-section:		
Sampled variable:		
Statement of inquiry question:		
Sample size n=		
Min:		
Q1		
Median		
Q3		
Max:		
Average:		
Standard Deviation (sample):		
Bootstrap sample: 10,000 draws:	mean: standard error:	mean: standard error:
80% Confidence interval of pop parameter:	(,) width of interval:	(,) width of interval:
	Do the 80% confidence intervals overlap? If yes, by how much?	
95% Confidence interval of pop parameter:	(,) width of interval:	(,) width of interval:
	Do the 95% confidence intervals overlap? If yes, by how much?	
99% Confidence interval of pop parameter:	(,) width of interval:	(,) width of interval:
	Do the 99% confidence intervals overlap? If yes, by how much?	
Hypothesis testing (for page number sampling)		
Ho (null)		
H1 (alternative)		
P-value		
Interpretation of p-value		

Library sample analysis questions (please use separate paper or edit document digitally):
0. Please upload a well-formed spreadsheet containing your raw data to the OneDrive librarySample directory. Prefix your files with your first name and the letters of your sub-section prefix. You may perform your analysis in the spreadsheet, or by hand, but please include a digital copy of your results from the data organizer in this online repository.

1. Based on your samples of the two library sub-sections, summarize the differences the data suggest about the difference in page length of books in each. How do the mean and median measures compare between sections?

2. Using similar methods, write a brief overview analysis of your second sampled variable. Address both measures of center.

3. Compare the variation among values sampled for both of your variables. Which sub-section's sample suggests more variation among members of the population? Interpret the standard deviation of of samples for both variables to arrive at a conclusion.

4. How did you compute your confidence intervals around the population estimate? What assumption(s) MUST hold true about the population of books you've sampled for this data to be meaningful?

5. Analyze the pattern in the width of the confidence intervals for the 80%, 95%, and 99% confidence intervals. Do both of your sub-sections display similar patterns in the size of these intervals? Using your evolving knowledge of sampling, explain the patterns in the confidence interval sizes as our demanded confidence increases. Write your explanation in terms that would be meaningful to a thoughtful non-stats person.

5b. Consider the overlapping or non-overlapping nature of your confidence intervals at various levels. Explain the significance of checking for interval overlap in terms of your ability to compare differences between sections. Remember that even though we have estimated a single population mean for each sub-section using our sample, we must adjust that mean by a factor of a number of standard deviations plus or minus that point estimated mean.

6. Create a visual representation of your results which shows your confidence intervals around your point estimate for the true population's average page number in each sub-section. Include each of our three confidence levels in your visualization in a way that makes them easily comparable (by, say, stacking the graphics atop one another on the same number line). This is not a box-and-whisker plot, but rather an illustration of your own design depicting the use of your sampling to estimate a population parameter.

6. **Culminating question:** The director of the library has asked you to recommend one of your two sections to be digitized to save shelf space and hence operating costs, which section would you recommend digitizing? Use numbers to back up your claim. How confident are you in your recommendation? (And what do you mean by your confidence level?)

7. In conducting a hypothesis test in which the null hypothesis is that the average page count of books in the two sub-sections are NOT different, what computed p-value relates to these hypotheses? Interpret this p-value in language accessible to a thoughtful non-expert.

8. If you conducted a re-sampling of a sub-section already completed by students of the past term, how do your results compare to previous results?