STAT 345 Spring 2018
Homework 11 - Confidence Intervals Name: ______________________________

Please adhere to the homework rules as given in the Syllabus.

1. Confidence Intervals for the Mean. Assume that $X_1, X_2 \cdots X_n$ are iid with unknown mean μ. Suppose you calculate $\bar{x} = 50$. **Interpret all of your intervals!**

a) Assume that the variance is known to be $\sigma^2 = 16$ and the sample size is $n = 9$. Compute a 94% two sided confidence interval for μ.

b) Assume that the variance is known to be σ^2 and the sample size is $n = 9$. Compute a 99% upper confidence bound for μ.

c) Assume that the variance is *unknown*, but you calculate $s^2 = 16$. The sample size is still $n = 9$. Compute a 94% two sided confidence interval for μ. Is your answer very different from part a)? Why is this?

d) Repeat parts a) and d) using the same numbers, but this time $n = 144$. Are the intervals very different from each other now? Why is this?

2. Confidence Interval for Population Proportion. In the 2016-2017 season, Stephen Curry shot $n = 790$ three-pointers and made 325 of them. Let p be the true proportion of three-pointers that Curry will make.

a) By realizing that the data x_i are bernoulli trials, we know that p is also the population mean. Moreover, we can determine that the sample variance is $S^2 = 0.24$. *Challenge: Show that this is indeed the sample variance.* Produce a 95% confidence interval for p using the t-procedures.

b) Create a confidence interval for p using the Normal approximation to the Binomial. Which method produces a narrower confidence interval?

c) According to your interval in *b*), do you feel comfortable asserting that Stephen Curry makes more than 40 percent of his three-pointers?

d) If you wanted to know the true proportion within 1% (i.e. 0.01), how many Stephen Curry three-pointers do you have to observe?

3. Data Analysis. Recall the CDI datast which contains demographic information on US counties.The dataset can be found at

`math.unm.edu/~knrumsey/classes/fall16/MiniProjects/data.csv`

We will again consider the variable "PerCapitaIncome". Use R, Excel or any software of your choice to construct a 95% confidence interval for μ, the average per capita income in the United States. Interpret your interval.

4. Exponential Distribution. Assume that $X_1, X_2, \cdots X_{15}$ are iid $Exp(\lambda)$, and suppose we calculate $\sum_{i=1}^{15} x_i = 31$ and $\sum_{i=1}^{15} x_i^2 = 130$.

a) Compute $\bar{x}$ and s^2. *Hint: Use the shortcut method to find* s^2.

b) Set $\mu = \frac{1}{\lambda}$ and compute a small sample 95% CI for μ (using t-procedures).

c) Convert your CI for μ to a CI for λ by noting that $P(a < \mu < b) = P(1/b < \lambda < 1/a)$.

d) Define $T = \sum_{i=1}^{15} X_i$. Recall that T has a $Gamma(15, \lambda)$ distribution (since it is the sum of exponentials). When r is a positive integer, the CDF of a $Gamma(r, \lambda)$ distribution can be written as

$$F(x) = P(T \leq x) = 1 - \sum_{j=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^r}{j!}$$

Define $U = \lambda T$ and find the CDF of U. Use this to argue that $U \sim Gamma(15, 1)$.

e) Using an online calculator find a and b such that $P(a \leq U \leq b) = 0.95$. *Hint: Choose* a *such that* $P(U \leq a) = .025$ *and* b *such that* $P(U \geq b) = 0.025$

g) Convert your interval for U to an interval for λ by "pivoting". How does this interval differ from the one in part c)? *Fun fact: I simulated this data using* $\lambda = 0.5$. *Does your interval capture the true value?*

5. **Challenge Problem.** *The Bootstrap.* Suppose you have n data points $x_1, x_2, \cdots x_n$ and would like to develop a confidence interval for a parameter θ. If you have an estimator $\hat{\theta}$, the Bootstrap algorithm works as follows.

1. Sample from the data points **with** replacement until you have a new data set of length n. Estimate θ using this new dataset and call this estimate $\hat{\theta}^{(1)}$.
2. Repeat step 1. M times, so you have M estimates $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \cdots \hat{\theta}^{(n)}$.
3. For some confidence level $C \in (0,1)$ find the interval (a, b) which contains $100 \times C\%$ of the estimates $\hat{\theta}^{(m)}$. This gives a Bootstrapped CI for θ.

The skew of a distribution was discussed in an earlier HW (homework 6 I believe). Denote the "true" skew of a distribution as γ, and consider the estimator

$$\hat{\gamma} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

where $\bar{x}$ and s are the sample mean and standard deviation respectively.

Using the same data as in problem 3 on this homework, use the Bootstrap to give a 95% CI for the population skewness γ. Set $M = 10,000$.