DSC 10 WEEK 9 DISCUSSION WORKSHEET

We will be examining MLS soccer players' salaries to explore the Central Limit Theorem, confidence intervals, and more. There are 693 players in our dataset (full population). The dataset has Total Compensation average of μ and standard deviation of σ. Total compensation will be referred to as the players' salary.

	First Name	Last Name	Club	Position	Base Salary	Total Compensation
0	Sal	Zizzo	Atlanta United	D	$129,999.96	$129,999.96
1	Andrew	Wheeler-Omiunu	Atlanta United	M	$55,654.20	$55,654.20
2	Gordon	Wild	Atlanta United	F	$90,000.00	$120,000.00
3	Romario	Williams	Atlanta United	F	$71,500.00	$71,500.00
4	Brandon	Vazquez	Atlanta United	F	$125,004.00	$145,004.00

Figure 1: First 5 rows of the table

1. We gather MLS salary data from the top 10 most successful clubs in the MLS. We then use this data to estimate the average salary for the league. This this an acceptable way to sample? Why or why not?

2. What is the shape of the distribution of sample average salaries of sample size 100? What will the mean and SD of this distribution be?

3. What is the shape of 10 samples of average player salary of size 100?

4. A random sample of 100 players in the MLS have an average salary of $330,000. The SD of the sample is $700,000. What is our 95% confidence interval of the average salary?

5. Interpret a 95% confidence interval in context of this dataset.

6. Assuming a sample standard deviation of $700,000 and 99.7% confidence interval from $100,000 to $500,000, how big must the sample be?

7. What are the 2 different ways to decrease the size of a confidence interval? (assume the distribution from which you're sampling can not be changed)

8. A 99.7% confidence interval of $250,000 to $350,000 is found through bootstrapping 50 times from a sample of size 100 with SD $600,000. Will the interval get bigger or smaller when bootstrapping 5000 times is used?