STAT 630, HW 5
Due: Thursday, September 30

Directions: Please submit your completed assignment to Blackboard. For the concept questions, your solutions may be typed, or handwritten and then scanned. The data analysis questions should be completed using R Markdown and then rendered to HTML, PDF, or Word format.

Concept Questions

Please refer to the lecture 5 notes.

Exercise 1. Suppose you select a sample of size of size $n = 35$ from a moderately skewed distribution with mean $\mu = 250$ and standard deviation $\sigma = 50$.

(a) What is the the mean and standard deviation of the sampling distribution for $\bar{X}$?

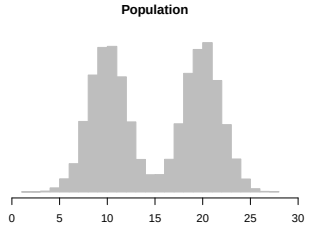
(b) What distribution does the sample mean $\bar{X}$ follow? Explain.

(c) What is the probability that the sample mean is greater than 240?

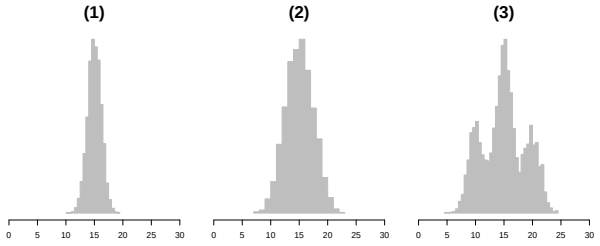
(d) What is the 20^{th} percentile of the distribution of the sample mean?

Exercise 2. The baggage limit for an airplane is set at 100 pounds per passenger. Thus, for an airplane with 200 passenger seats, there would be a limit of 20,000 pounds. The weight of the baggage of an individual passenger is a random variable with a mean of 95 pounds and a standard deviation of 35 pounds. If all 200 seats are sold for a particular flight, what is the probability that the total weight of the passengers' baggage will exceed the 20,000-pound limit?

Exercise 3. Suppose data are sampled from a bimodal population distribution that has the following shape:



The three histograms shown below are sampling distributions for the mean, which were each generated by repeatedly taking random samples of size n from this population. Match the sample size n with the sampling distribution, and give a brief explanation.
(a) $n = 2$, (b) $n = 5$, (c) $n = 20$



Data Analysis Questions

Please refer to lab 5.

The San Francisco Controller's Office maintains data on the salary and benefits paid to city employees[1]. Run the following command to load in the vector of salaries for over 45,000 employees:

```
Salaries <- readRDS(url("https://ericwfox.github.io/data/Salaries.rds"))
```

This unique data set contains the salary (in 1000's of dollars) of every city employee in San Francisco during the year 2017. This will serve as our population of interest for the following questions which investigate sampling distributions.

Exercise 4. Make a histogram and normal QQ plot of the salaries for the entire population of city employees. Also compute the population mean, μ, and population standard deviation, σ, of the salaries. Describe the shape of the population distribution.

Exercise 5. Take a single sample of 30 salaries and store the sample in a vector. Compute the mean of the sample of 30 salaries that you selected. How does your sample mean compare to the population mean for salaries? Does it seem like an accurate estimate?

Exercise 6. Use R to draw 5000 samples of size $n = 30$ from this population. Compute the mean of the salaries in each sample; store the 5000 sample means you generated in a vector. Then make a histogram and QQ plot of the 5000 sample means. Describe the shape of the sampling distribution.

Exercise 7. Compute the mean and standard deviation of the 5000 sample means generated in Exercise 6. How close are these values to μ and $\sigma/\sqrt{n}$, where μ and σ are the population mean and standard deviation of the salaries of SF city employees?

Exercise 8. According to the Central Limit Theorem, what distribution does the sample mean follow when the sample size n is sufficiently large? Does the CLT appear to hold true for the sampling distribution you generated in Exercise 6 with $n = 30$?

Exercise 9. Repeat Exercise 6, but this time use a smaller sample size $n = 5$. How does decreasing the sample size affect the shape of the sampling distribution?

[1]Data Source: `https://data.sfgov.org/City-Management-and-Ethics/Employee-Compensation/88g8-5mnd`