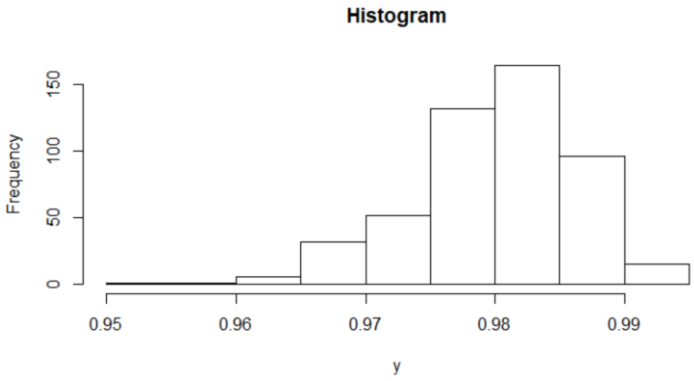
ST 311 Practice Exam

Multiple Choice: For each, fill in the circle for one answer only.

1. Does a small Interquartile Range (IQR) always indicate that the data has low variability?
    - ○ Yes, because 50% of the data points do not vary much from one another.
    - ○ Yes, because the IQR is representative of the entire data set.
    - ○ No, because the IQR is the difference between the maximum and the minimum values.
    - ○ No, because the lower 25% and upper 25% of the data points can still vary greatly.

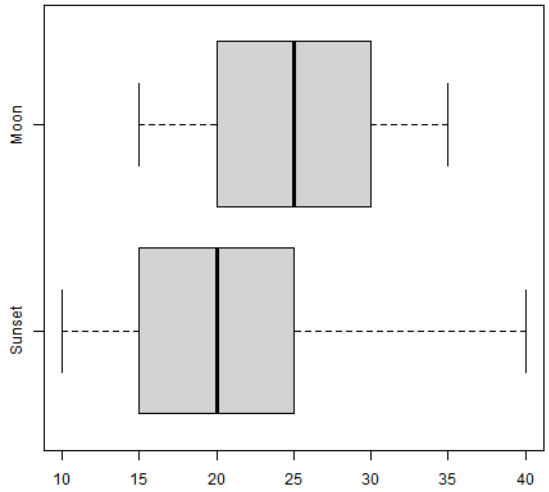
2. Which of the following is true about this histogram?



    - ○ The mean is less than the median
    - ○ The mean and median are the same
    - ○ The mean is greater than the median
    - ○ It is impossible to tell about the mean and median from the histogram

3. Joe hosts a weekly local radio podcast in Raleigh and talks about various health benefits from running. He wants to host a 5k run for awareness, but is afraid nobody will show up and he will be lonely. To try to gauge interest from Raleigh citizens, he decides to ask listeners of his show one day if they would be interested in a 5k occurring exactly three weeks later. He gets a total of 20 calls, all stating they would be interested in this 5k. Due to this, he decides that there would be overwhelming interest in this run from Raleigh citizens and plans the run. On the day of the run, less than 15 people show up and he is very confused. Which of the following is **NOT** a sampling error that Joe made?
    - ○ Undercoverage
    - ○ Data Entry
    - ○ Using a Volunteer Response Sample
    - ○ He made all of the above errors

4. A researcher is interested in the relationship between study time and exam scores for college students. The investigator collects the number of study hours per week for students from Sunset College and Moon College. The result is presented in the boxplots:



Which of the following could we conclude from the graph?

- ○ The mean of two colleges is the same.
- ○ The range of two colleges is the same.
- ○ Sunset College has a larger interquartile range than Moon College.
- ○ The mean of Sunset College is larger than its median.

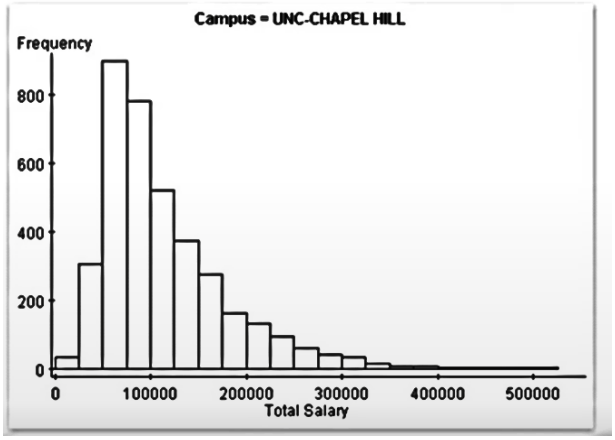
5. The distribution of the population of interest is Normal with mean μ and standard deviation σ. We take a random sample from the population. Which of the following is **NOT** true about the distribution of the sample mean (i.e. the sampling distribution)?

- ○ The sampling distribution allows us to find probabilities about the sample mean.
- ○ The true mean of the sampling distribution is also μ.
- ○ The standard deviation of the sampling distribution will be greater than or equal to the standard deviation of the population.
- ○ The distribution of the sample mean is also a Normal distribution.

6. Anna works at a party center. She is buying party poppers for upcoming events and finds two different vendors, PartyPlus and HooraySupplies. PartyPlus states that p = 95% of their poppers will pop when the string is pulled. HooraySupplies states that p = 90% will pop. If Anna purchases 200 poppers from each vendor, which vendor's distribution will have a **larger standard deviation** of the sample proportion of poppers that pop?

   - ○ PartyPlus has a larger standard deviation because its proportion is closer to 1.
   - ○ HooraySupplies has a larger standard deviation because its proportion is closer to 0.5.
   - ○ They will have the same standard deviation because the sample size is the same.
   - ○ It is impossible to tell because we do not know the sample proportions.

7. Which of the following could **NOT** be used to describe the histogram below:

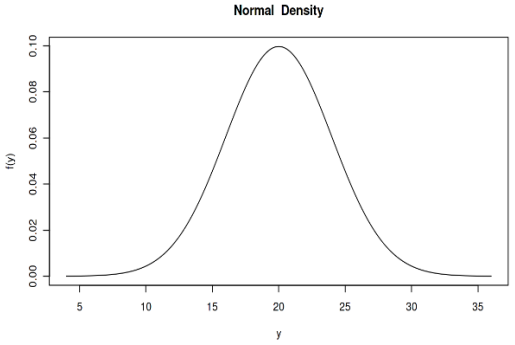


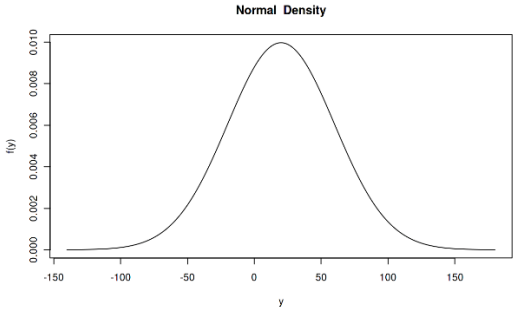
   - ○ Skewed right
   - ○ Normal Distribution
   - ○ Unimodal
   - ○ Non-symmetric

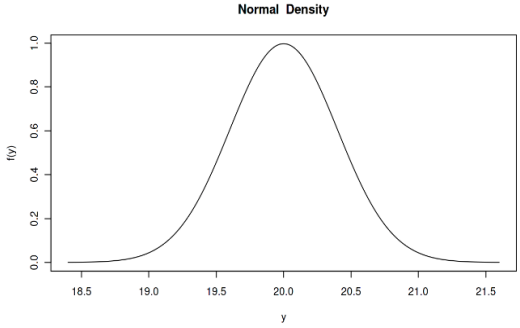
8. Suppose that the number of students that use Macs in ST311 every semester is normally distributed with a mean of 653 students with a standard deviation of 123 students. Between which two values do approximately 68% of the number of students that use Macs fall between every semester?

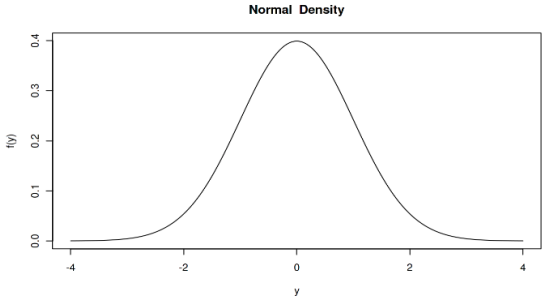
   - ○ 0 students and 653 students
   - ○ 123 students and 776 students
   - ○ 530 students and 776 students
   - ○ 407 students and 899 students

9. The number of minutes that it takes for students to commute to class during the week follows a normal distribution with mean 20 and standard deviation 4. I go around the nation taking samples of 100 people at a time and asking how long they take to get to class. For each group of 100 people, I take the mean of the times I recorded. I do this until I feel confident that I can estimate the sampling distribution of the mean amount of time it takes for people to get to class. Which graph below is the best representation of what we would expect the **<u>sampling distribution</u>** of the sample mean to look like?

○ 

○ 

○ 

○ 

10. Three friends Lenny, Karl, and Edna all compete in the Springfield Marathon and finish with the following times 221, 207, and 241 minutes, respectively. After the race, the three get into a friendly debate over which performed better relative to their age group. Edna ran in the 50-60-year-old race and while Lenny and Karl ran in the 40-50-year-old race. The 50-60-year-old group finished the race with a mean time of 261 minutes and a standard deviation of 5 minutes and the 40-50-year-old group finished with a mean time of 242 minutes and standard deviation of 7 minutes. What is the correct ranking of performances for the three:

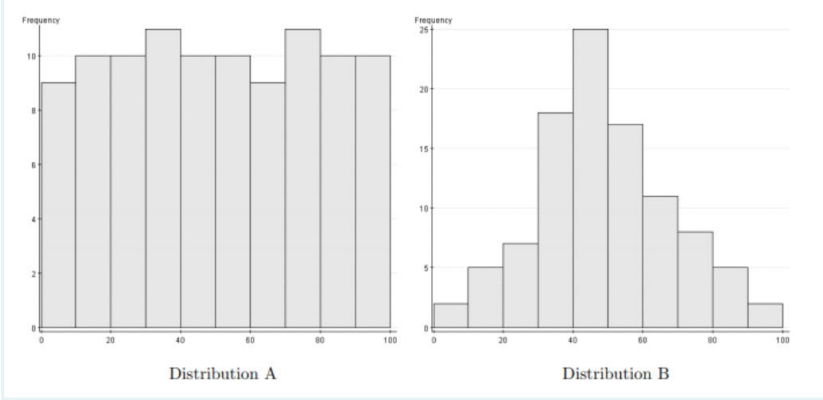
- ○ Lenny was the fastest, followed by Edna, then Karl
- ○ Edna was the fastest, followed by Karl, then Lenny
- ○ Edna was the fastest, followed by Lenny, then Karl
- ○ Karl was the fastest, followed by Edna, then Lenny

11. Suppose a high school principal wants to gauge student interest in having mandatory nap time in the middle of the day. The high school consists of grades 9-12 with the students distributed as follows:
    - 300 in grade 9
    - 240 in grade 10
    - 420 in grade 11
    - 330 in grade 12

    The principal wants to be sure all grades are included in the sample. What sampling technique should the principal use to get an accurate understanding of student opinions regarding mandatory nap?

    - ○ Cluster Sample: The principal randomly selects 5 students from each grade and gauges each selected student's opinions
    - ○ Cluster Sample: The principal randomly selects one grade level and gauges each student's opinion from the grade
    - ○ Stratified Sample: The principal randomly selects 5 students from each grade and gauges each selected student's opinions
    - ○ Stratified Sample: The principal randomly selects one grade level and gauges each student's opinion from the grade
    - ○ Simple Random Sample: The principal randomly selects 20 students and gauges each selected student's opinions

12. Consider the two distributions of data below. Distribution A consists of data randomly sampled from population A, while distribution B consists of data randomly sampled from population B.



    Which of the following statements is true?

    - ○ Individuals are more likely to be more than 1 standard deviation from the mean in population B than population A.
    - ○ Distribution A's mean will be significantly smaller than its median.
    - ○ Distribution A has a larger interquartile range than distribution B.
    - ○ Distribution A has a smaller standard deviation than distribution B.

13. Sonja reads in an article that 33% of Americans have at least one tattoo. She takes a random sample of 25 people and finds that 16 of them have at least one tattoo. Is this outcome unusual?
    - ○ Yes – $\hat{p} = \frac{16}{25} = 0.64$ is more than 2 standard deviations above $p = 0.33$.
    - ○ Yes – $\hat{p} = \frac{16}{25} = 0.64$ is higher than $p = 0.33$.
    - ○ No - $\hat{p} = \frac{16}{25} = 0.64$ is greater than 0.05.
    - ○ It's difficult to say – the distribution of $\hat{p}$ is skewed so we can't use the Normal distribution.

14. Suppose you are a data scientist at the social network company Mata and your director wants to know the user's attitude on the new logo of Mata. Ethnicity is an important factor and there are 7 groups of users at Mata: Asian and Pacific Island, Black, Latino, Native American, Other race, Two or More race, White. To design a sampling scheme that represents user opinions from each group, which of the following strategies is BEST (select one):
    - ○ Simple random samples: randomly survey 10,000 users from the Mata network.
    - ○ Stratified samples: randomly survey 1500 users within each ethnic group.
    - ○ Cluster samples: randomly select 5 ethnic groups, then survey all users of those groups.
    - ○ Multi-stage samples: combine stratified with cluster samples by randomly select 5 ethnic groups from each state and surveying all users of those groups.

15. Suppose the ages of ST 311 students are: 20, 24, 17, 18, 21, 19, 25, 20, 26, 19, 43, 20. If we add 5 to each age, which measurement would change?
    - ○ IQR
    - ○ Standard deviation
    - ○ Median
    - ○ Range

16. It is assumed that 75% of people have a deviated septum (most are minor). In a random sample of 14 people, how many would we expect to have a deviated septum and with what variability?
    - ○ We would expect 10.5 people to have a deviated septum, with a standard deviation of 0.116.
    - ○ We would expect 10.5 people to have a deviated septum, with a standard deviation of 1.62.
    - ○ We would expect 11 people to have a deviated septum, with a standard deviation of 0.116.
    - ○ We would expect 11 people to have a deviated septum, with a standard deviation of 1.62.
    - ○ We are unable to answer because we do not know if the distribution is Normal.

Short Answer: Show all your work to receive credit.

17. According to a study conducted in 2015, daily caffeine consumption for adults in the U.S is approximately normal with a mean of 250 mg and a standard deviation of 65 mg.
    a) The U.S. Food and Drug Administration considers 400 mg a safe amount of caffeine for healthy adults to consume daily. We want to know the probability that a randomly selected US adult consumes more than 400 mg in one day. Set up the probability statement, draw the distribution, and shade the corresponding area under the curve that represents this probability.

    b) Is it likely that a random US adult consumes too much caffeine in a day? Explain.

    c) How many milligrams (mg) of caffeine correspond to an amount that lies one standard deviation below the mean consumption?

(Continued from previous) A dietician believes that adults in the U.S. today consume more caffeine than in recent years. She designs a new study to compare current consumption to that in 2015. She measures caffeine intake for a random sample of 287 U.S. adults.

d) The dietician found a sample mean of 258 mg in her study. If nothing has changed since 2015, is it unusual to observe a sample mean of 258 mg or more?

e) Does the study provide evidence that the dietician is correct in believing that adults consume more caffeine now than they did in 2015? Explain in 2-3 sentences using your results from part (c) above, as though you were talking to someone who has never taken statistics.

18. Interest rates for a 30-year fixed-rate mortgage on a particular day from lenders in Seattle, WA had a mean of 4.671 with a standard deviation of 0.436. Describe the probability that the mean interest rate for such a mortgage from a random sample of 10 Seattle lenders is below 4.5 by writing the probability statement and shading the distribution, OR explain why it cannot be found.

19. The 2020 Trustwave Global Security Report shows that spam messages account for 28.5% of all email traffic worldwide. Suppose you are going to randomly select 200 emails and look at the sample proportion of spam message.
    a) What is the standard deviation of sample proportion (i.e. the standard error of the sample proportion)?

    b) Set up a probability statement for the probability that the proportion of these 200 emails is more than 31.5%, draw the distribution, and shade the corresponding area under the curve that represents this probability.

    c) Which probability most closely estimates this probability?
        - ○ 0.315
        - ○ 0.9375
        - ○ 0.15
        - ○ 0.85

20. The transit authority in City A wants to obtain input from bus users regarding a new payment system. Two proposals for obtaining such feedback, below, are presented.

    Proposal 1: The transit authority holds a meeting in the town hall to discuss the changes and get feedback from the community.
    Proposal 2: On each bus, a flyer with information about the new system is posted. Bus riders may complete a survey online, the link to which is shared at the bottom of the flyer.

    a) What issue(s) may arise if Proposal 1 is employed? Explain using a complete sentence.

    b) What issue(s) may arise if Proposal 2 is employed? Explain using a complete sentence.

    c) If you were asked to write a third proposal, what would you suggest the transit authority do to collect a random sample?