What do I know already? (Do not look up the answers to these questions. The purpose is to assess your current level of knowledge on these topics.)

A. According to the empirical rule, approximately what percent of data is captured between $\mu - 2\sigma$ and $\mu + 2\sigma$?
95%_____

B. The empirical rule assumes the shape of the distribution is known. What shape is assumed? It is given different names: bell-shaped, mound shaped and symmetric, or normal._____

C. The empirical rule provides us with the approximate percent of data within some given interval. What does Chebyshev's theorem give us? Chebyshev's theorem gives us a lower bound for the amount of data that can be found in an interval that is symmetric with respect to the mean._____

D. Describe a z score in your own words. What sort of z score would indicate that a measurement was unusual? A z score represents the number of standard deviations a value is from the mean for the set of measurements the value belongs to. Typically, a value is considered unusual if the absolute value of its z score is greater than 2. It depends on the distribution, but Chebyshev's theorem tells us that at most 25% of values will have a z score with an absolute value greater than 2._____

Learning Objectives: (Click the learning objectives below for a short clip on the topic.)

Recall the Empirical Rule (1)
Use Empirical Rule to Determine the Approx. Percentage of Data Inside an Interval (3)
Form an Interval that Captures Approximately X% of the Data (3)
Contrast the Use of Chebyshev's Theorem and the Empirical Rule (2)
Explain the Z-score Formula (2)
Calculate and Interpret Z-scores (as a measure of unusualness) (3)
Compare Data Points from Different Sets Using Z-scores (4)

Exercises:

1. According to a report by Common Sense Media, teens are spending an average of 9 hours per day in front of a screen for entertainment. The standard deviation for the time teens spend in front of a screen for entertainment is 2.2 hours. Would it be unusual for a teen to spend 2 hours or less in front of a screen for entertainment each day? Why or why not? (Hint: use the z-score to support your answer.) Yes, because the z score for these times would be less than or equal to -3.18.

2. Some writers earn very large annual salaries, but this sort of financial success is unusual. What measure of the center would you recommend to describe the typical earnings for writers?
   A. Mean
   B. Median
   C. Mode
   D. Range
   E. Mean Absolute Deviation

3. NPR (National Public Radio) claims that their average listener listens to NPR programming for 3.2 hours every day. To test this claim a newspaper conducted a poll of 200 random NPR listeners. Each of the polled listeners were asked to estimate the amount of time they listened to NPR each day. The results of the poll were summarized and expressed as a z score by using the claimed value of 3.2 hours as the population mean. The resulting z score was -0.46. Based on the z score, does NPR's claim appear to be plausible?
   Yes, because -0.46 is not extreme or unusual. Usually, if the claimed mean is correct, the sample mean will be very close to the claimed mean value. This in turn will produce a z score that is typically small in absolute value (i.e., |z| is between 0 and 1.2).

4. True or false: The empirical rule can be used to determine the minimum percentage of data contained in an interval of the form $(\mu - k\sigma, \mu + k\sigma)$ where k is greater than 1. This describes Chebyshev's rule.

5. Two CrossFit athletes, Julie and Samantha, are working together to complete a pair of workouts during a competition. Julie finished her workout in just 5.1 minutes. Samantha finished hers in 7.0 minutes. The workout Julie finished had an average completion time of 6.5 minutes and a standard deviation of 1.1 minutes. The workout Samantha completed had an average completion time of 9 minutes with a standard deviation of 1.3 minutes. Who performed relatively better? Julie's time has a z score of -1.27, and Samantha's time has a z score of -1.54. That indicates Samantha actually finished faster than Julie relatively speaking because Samantha's time is more below average when compared to her peers. Note below average here implies faster because it took less time.

6. Each patient undergoing an annual checkup at a local physician's office is weighed. The weights are recorded for each patient. This set of measurements is an example of: (select all that apply)
A. continuous data
B. discrete data
C. qualitative data
D. ordinal level data
E. quantitative data

7. A study in 2014 indicated that the average height for women in Guatemala was 58.7 inches (a little under 4ft 11in). The standard deviation for these heights is 1.9 inches. It is safe to assume these heights are normally distributed. What percent of Guatemalan women are between 54.9 inches and 60.6 inches tall?
Approximately 81.5%

8. True or false: Chebyshev's theorem can only be used with normally distributed data. The theorem can be used for any distribution.

9. Dutch men are now the tallest men in the world. The average height for Dutch males is 71.9 inches (That's almost 6 feet tall!). The standard deviation for their heights is 2.6 inches. Assuming that these heights have a bell-shaped distribution, what can be said about the percentage of Dutch men that are taller than 77.1 inches tall? (i.e.-taller than 6ft 5.1 inches)
Approximately 2.5%

10. Each patient undergoing an annual checkup at a physician's office in Canada is weighed. The weights are recorded for each patient in kilograms (kg). What unit of measurement would the standard deviation and variance of these weights have respectively?
A. kg and kg²
B. kg² and kg
C. kg for both
D. kg² for both
E. √kg and kg

11. The average weight of cars on U.S. roads is 4,009 pounds. If we assume the standard deviation for the weights of the cars is 919 pounds, create an interval that would capture the weights of at least 88.9% of all cars on U.S. roads.
(1,252 pounds to 6,766 pounds) using k=3 with Chebyshev's theorem

12. Final exam scores for STA3123 have an average of 68.3. The median score is 72, and the most common score (the mode) for the class is 75. Is the distribution likely left skewed, right skewed, or symmetric?

13. True or false: a measurement's z score provides us with the number of standard deviations the measurement is away from the mean for the set of measurements.

14. The lengths of time it takes Spirit airlines to board flights traveling from FLL in Florida to BOS in Boston are normally distributed (i.e.-bell shaped) with a mean of 25 minutes and a standard deviation of 4.2 minutes. Form an interval that contains approximately 95% of these times. Using the empirical rule (16.6 min to 33.4 min)

15. According to a report by Common Sense Media, teens are spending an average of 9 hours per day in front of a screen for entertainment. The standard deviation for the time teens spend in front of a screen for entertainment is 2.2 hours. What is the maximum percentage of teens that spend less than 3 hours in front of a screen for entertainment each day?   At most 13.4% will spend less than 3 hours in front of a screen for entertainment each day

16. Two friends want to compare their performances on the final exam in STA 2122. Juan earned 72 points in a class with a mean grade of 65 and a standard deviation of 12.8. Lana earned 70 points in a class with a mean grade of 60 and a standard deviation of 14.6. Who performed relatively better on their course final? Lana's points have a z score of 0.68. Juan's points have a z score of 0.55, so Lana did relatively better.

17. Responding to criticism from nutritionists, the CEO of a company that operates a chain of movie theatres claims that typically customers only eat half of the bag of large popcorn served at its theatres. If true, this would equate to an average of 600 calories of popcorn consumed by customers ordering the large tub of popcorn. Researchers weighed the contents of a random sample of customers' large popcorn containers before and after their movie. The sample results were summarized into a z score by using the claimed value of 600 calories as the population mean. The resulting z score was 3.74. Based on this result, does the CEO's claim seem credible?
No, the z score is unusually large, so the claim seems implausible. When the claimed mean is not correct, the sample mean will typically be far from the claimed mean value. As a result, the data will produce a z score that is typically unusually large in absolute value (i.e., |z| is between 2.0 and ∞).