1. Consider the following sentence from a statistical report, as presented in De Veaux, Velleman, and Bock (2012).

   Based on meteorological data for the past century, a local TV weather forecaster estimates that the region's average winter snowfall is 23 inches, with a margin of error of ± 2 inches.

   (a) If you lived in this region, would you want the margin of error to be large or small? Explain.

   (b) Why do you think a margin of error is reported?

Investigating Test Scores for Seniors at Maplewood High School

Suppose that our entire population is 100 student test scores taken from the seniors at Maplewood High School (see the *population distribution* below), and we know the population mean test score (out of 100 points) is $\mu = 76.96$ points.
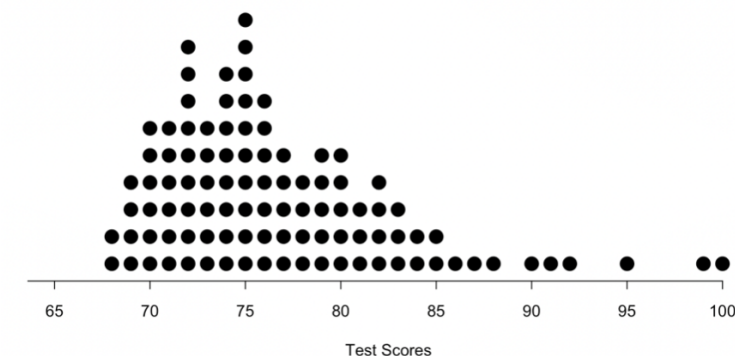
Test Scores

1. You have a set of 100 cards that represent the 100 test scores from the entire population of seniors at Maplewood High School. Randomly draw a sample of **size 10** from these cards.

   (a) Write down the 10 test scores you randomly selected and compute the mean test score of your sample.

   (b) Share your sample mean test score with classmates and compare. Why do you think your sample mean test score is different from others?

2. What happens if we repeat this process? From your set of 100 cards, continue to randomly draw a sample of size 10 and compute the mean test score of your sample. Repeat this process until you have a total of 10 sample means.

   (a) Write the 10 sample mean test scores below.

   (b) As a class, create a dotplot of everyone's sample means and sketch it below. This is the class's *approximate sampling distribution* of sample mean test scores. Describe what one dot in the dotplot represents.

We can continue to perform a simulation by hand with the cards, but technology provides us a more efficient way to do this! Our goal is to see what other sample means we might have gotten from different samples of size 10, and simulation is a tool we can use to see how variable the sample mean test scores might be.

3. Using the population of 100 test scores, use technology to conduct a simulation. Randomly draw a sample of **size 10** and compute the mean test score of your sample. Repeat this process for a total of 500 times.

   (a) Create a dotplot of your 500 sample mean test scores (i.e., another approximate sampling distribution) and sketch it below. Describe what one dot in the dotplot represents.

   (b) Describe the shape, center, and spread of your sampling distribution from Problem 3(a). What is the mean and standard deviation of your sampling distribution? Note that the standard deviation of a sampling distribution is referred to as a *standard error*.

   (c) Compare the shape, center, and spread of your sampling distribution from Problem 3(a) to the population distribution presented on page 1 of the Class Activity.

   (d) Based on your sampling distribution from Problem 3(a), is a sample mean (computed from a random sample of size 10) a good way to estimate the population mean test score? Explain your reasoning.

In Problem 3, we found that a sample mean (computed from a random sample) is usually a really good guess of the population mean. But, not all good guesses are created equally! We also saw in that sampling distribution that some of our good guesses were further from the population mean than others. In statistics, we care about how much variability is present among these sample means (estimated by the standard error you found in Problem 3), and we often report a range of plausible values for a population mean. Problems 4 and 5 guide you through two methods for estimating a margin of error from a unimodal and symmetric sampling distribution (use your sampling distribution from Problem 3 to answer Problems 4 and 5).

4. **Method 1: Counting Dots**

    (a) Based on counting dots in your sampling distribution, the **middle 95%** of the sample mean test scores land between _____ points and _____ points. Explain how you came up with these two values.

    (b) The two values you found above in Problem 4(a) are both approximately _____ points from the mean of your sampling distribution. Explain how you came up with your answer, which is a margin of error with a confidence level of 95%.

5. **Method 2: Empirical Rule**
    A student, Kyle, used the empirical rule instead of counting dots to estimate a margin of error with a confidence level of 95%. They said that using the empirical rule was quicker than counting dots and that they got a similar answer to their friends who counted dots.

    (a) Use the empirical rule and your sampling distribution to estimate a margin of error with a confidence level of 95% (show your work). In other words, how many test score points do you need to go out from the mean of the sampling distribution to capture the **middle 95%** of the sample mean test scores? Compare your answer to your answer in Problem 4(b).

(b) Describe why Kyle might have thought to use the empirical rule to compute a margin of error in this situation.

(c) Explain why it's appropriate for Kyle to use the empirical rule to compute a margin of error in this situation.

6. Luis is another student in the same class as Kyle. He noticed what Kyle did and wonders if it will always work. He asks the teacher if they can always use the empirical rule in this situation to estimate a margin of error. The teacher recognizes that this is an opportunity to help her students understand when these methods can be used to estimate a margin of error and when one method is more convenient over the other. The teacher ask her students the following question:

If you are given a sampling distribution that is approximately unimodal and symmetric, describe how you could estimate a margin of error with a confidence level of 68% and another margin of error with a confidence level of 90%.

(a) Explain how this question can help Luis to understand more about under what conditions the empirical rule is a convenient method for estimating a margin of error from a unimodal and symmetric sampling distribution.

(b) What other confidence levels could the teacher ask about to help students understand under what conditions the empirical rule is (or is not) a convenient method for estimating a margin of error from a unimodal and symmetric sampling distribution? Explain.

7. Use the mean from your first random sample of 10 test scores (see Problem 1(a)) and the margin of error you found using your approximate sampling distribution (see Problem 4 or 5) to write a sentence that describes what you found about the typical test score for seniors at Maplewood High School. Your sentence can be modeled after the sentence describing snowfall in Problem 1 of the Pre-Activity.

1. Consider the two simulated sampling distributions of sample means.
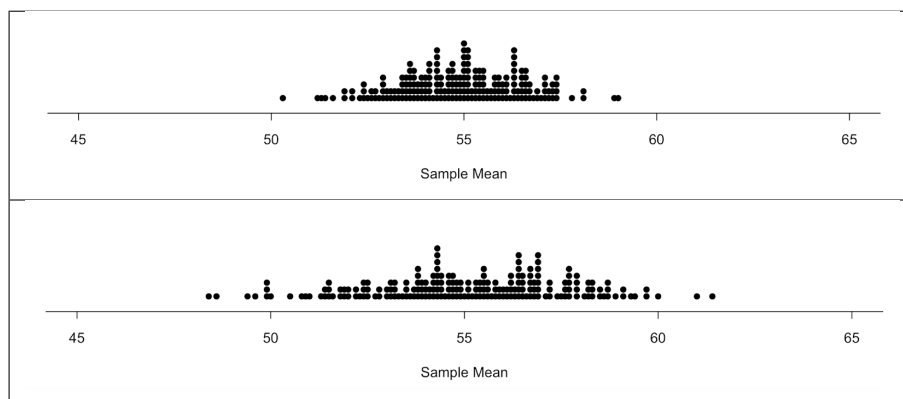
   (a) Assume the sample means came from a random sample. What's your best guess at where the population mean is? Explain.

   (b) Based on the top sampling distribution, what is a reasonable estimate of a margin of error for a confidence level of 95%? Explain your reasoning. (Note that the sampling distribution was created using 200 samples of size 40.)

2. Saskia, Aaron, Gerlie, and Moses are working on the following problem.

   > A survey of 625 randomly selected students was conducted to determine the average amount of time students sleep during a weekday. The survey reported an average of 6.5 hours. The survey estimate had a margin of error of half an hour. A margin of error is reported because
   >
   > A. Sample means vary from sample to sample.
   >
   > B. Students may intentionally respond incorrectly.
   >
   > C. Students may misunderstand the survey questions.
   >
   > D. The people doing the survey may have recorded results incorrectly.

   Each student selects a different reason for why a margin of error is reported.

Choice A	Choice B	Choice C	Choice D
Gerlie	Saskia	Aaron	Moses

   (a) Who selected the correct (and complete) answer and why?

   (b) For each of the three students who selected an incorrect choice, explain what conception they have of margin of error.

3. Recall the context of test scores from the Class Activity. Suppose that you wanted to capture the middle 99% (instead of the middle 95%) of the mean test scores from a sample of size 10. How would your margin of error change? Explain your reasoning.

4. The proliferation of text generated by artificial intelligence has led to questions about how to distinguish passages that are written by humans compared to passages written by artificial intelligence, which leads to the need to examine characteristics of blocks of text. One characteristic of a block of text is the mean word length. Consider the excerpt of Martin Luther King Jr.'s "I Have a Dream" speech.

> And so even though we face the difficulties of today and tomorrow, I still have a dream. It is a dream deeply rooted in the American dream. I have a dream that one day this nation will rise up and live out the true meaning of its creed: We hold these truths to be self-evident, that all men are created equal.

> I have a dream that one day on the red hills of Georgia, the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood.

> I have a dream that one day even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into an oasis of freedom and justice.

> I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character. I have a dream today.

Draw on your experiences from the Class Activity to estimate the mean word length of this passage with some margin of error:

(a) Without going through and calculating the length of every single word in the passage, describe how you would use statistics to construct a "best guess" at the true mean word length.

(b) Describe how you would use simulation to construct a margin of error for the true mean word length?

1. Below is a dotplot of the sample mean weight for 100 different random samples of size 10 from a population of adult Labrador retrievers where the mean weight is 65 pounds.

Sample Mean Weight (pounds)
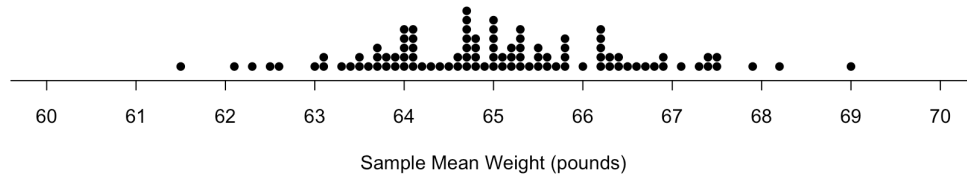
(a) Describe what one dot in the dotplot represents.

(b) Fill in the blanks.

95% of the sample mean weights fall between _____ *and* _____.

Explain how you came up with these endpoints.

(c) Based on your answer in 1(b), estimate a margin of error with a confidence level of 95%. Explain your work.

(d) Would a margin of error with a confidence level of 99% be larger or smaller than the margin of error you estimated in 1(c)? Explain your reasoning.

2. Kyle's and Luis' teacher knows that her students can count dots and use the empirical rule on a unimodal and symmetric sampling distribution to estimate a margin of error. It's because the sampling distribution is symmetric and unimodal that both methods will give approximately the same answer for a margin of error. The teacher wants her students to understand when each method is (or is not) the most useful method for estimating a margin of error in this situation.

   (a) The teacher gives her students a unimodal and symmetric distribution and asks them to estimate a margin of error with a confidence level of 68%. Explain why the teacher uses this prompt to help her students understand when the empirical rule is more useful than the counting dots method to estimate a margin of error associated with a 68% confidence level.

   (b) The teacher then asks her students to use the same sampling distribution from part (a) to now estimate a margin of error with a confidence level of 90%. Explain why this question is useful in helping students understand when the counting dots method is more useful than the empirical rule to estimate a margin of error associated with a 90% confidence level.