We have previously studied the sampling distribution of a sample proportion (as with the Reese’s Pieces). Today we investigate the sampling distribution of a sample mean. In other words, we shift gears today to analyze quantitative rather than categorical variables.

Example 20–1: Gettysburg Address (cont.)
Recall that we took random samples from the population of 268 words in the Gettysburg Address. We determined the number of letters in each word of the sample, and we calculated the mean (average) number of letters per word in the sample.

a) Identify the observational units and variable. Also classify the variable as categorical or quantitative.

b) Is the mean number of letters in your sample a parameter or statistic? Explain, and indicate the symbol used to represent it.

c) In the population of all 268 words, the mean number of letters per word is 4.295. Is this a parameter or statistic? Explain, and indicate the symbol used to represent it.

d) Use the “Sampling Words” applet to look at the distribution of the population word lengths. Describe the shape of this distribution. Also record its mean and standard deviation, along with appropriate symbols.

e) Use the applet to take 1000 random samples of size 1 word each. What does the distribution of sample means look like in this case? Why does this make sense?

f) Use the applet to take 1000 random samples of size 5 words per sample. Look at the distribution of sample means. Describe two ways in which this distribution differs from the population. Also comment on one aspect in which these two distributions are similar.
g) Increase the sample size to 25, and predict how the distribution will differ and how it will be similar. Then examine the distribution of 1000 sample means. Describe two ways in which this distribution has changed (from when the sample size was 5).

Central Limit Theorem (CLT) for Sample Mean: Suppose that the population distribution of a quantitative variable has mean $\mu$ and standard deviation $\sigma$, and suppose that a random sample of size $n$ is taken from the population. Then the sampling distribution of the sample mean $\bar{x}$ is approximately normal, with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$, provided that either the population itself is normal (in which case the sampling distribution of $\bar{x}$ is exactly normal) or the sample size is large ($\geq 30$ as a rule-of-thumb).

h) Draw a well-labeled sketch that illustrates this CLT result in general. Then draw sketch specific to this example with a sample size of 25.

Example 20–2: Study Times
Suppose that the distribution of weekly studying times among the population of all Cal Poly students has a mean of 20 hours and a standard deviation of 8 hours.

a) Suppose that you select a simple random sample of 75 Cal Poly students and calculate the sample mean studying time. Describe what the CLT says about how this sample mean would vary from sample to sample. Include a well-labeled sketch to illustrate this sampling distribution.

b) In a random sample of 75 students, determine the (approximate) probability that the mean (average) study time would be more than 21.5 hours.
c) In a random sample of 75 students, what is the probability that the average study time is between 18.5 and 21.5 hours?

d) Are your answers to (a) – (c) valid even if the distribution of studying times in the population is skewed? Explain.

e) Would these probabilities (in (b) and (c)) increase, decrease, or stay the same if the sample size were larger? Explain.

f) In order for the probability in (b) to be less than .02, how many students would have to be sampled?

g) Does the given information allow you to determine the probability that an *individual* student studies for more than 21.5 hours per week? Explain.