Stat 218 - Day 9
Sampling distribution of sample mean
Central Limit Theorem

Recall from last time:

Key Result 1: Suppose that the proportion of a population having some characteristic is denoted by p, and suppose that a random sample of size n is taken from the population. Then the sampling distribution of the sample proportion $\hat{p}$ is approximately normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. This approximation is generally considered to be valid as long as $np \geq 10$ and $n(1-p) \geq 10$.

Today we will encounter a similar result for the sampling distribution of a sample *mean* $\bar{y}$. In other words, today we will consider a quantitative variable, whereas last time we considered a categorical variable. We will consider two cases:

- The population follows a normal distribution.
- The population does not follow a normal distribution.

Example: Birthweights (cont.)
Recall that birthweights vary according to a normal distribution with mean 3250 grams and standard deviation 550 grams. Suppose that we take a random sample of birthweights, calculate the average weight, and do this repeatedly.

Key question:
(a) Do you expect that the sample averages will vary more, less, or the same as individual values? [For another example, think about how heights of students across campus vary, and then think about how the average heights per classroom vary.]

We will examine a Minitab worksheet containing:

- 1000 birthweights
- 1000 average birthweights from random samples of size n=2
- 1000 average birthweights from random samples of size n=4
- 1000 average birthweights from random samples of size n=20

(b) Compare the shapes, centers, and spreads of these distributions.

Key Result 2: Suppose that a population follows a normal distribution with mean μ and standard deviation σ, and suppose that a random sample of size *n* is taken from the population. Then the sampling distribution of the sample mean $\bar{y}$ is also normal, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Notes: This result can be very challenging to understand. Be careful to note that there are three different uses of "mean" here: the mean value of the population (μ), the sample mean ($\bar{y}$) that varies from sample to sample, and the mean of the distribution of those $\bar{y}$'s, which turns out to equal the population mean μ. There are also two different uses of "standard deviation" here: the standard deviation of the values in the population (σ) and the standard deviation of the distribution of the $\bar{y}$'s, which turns out to equal the population standard deviation divided by the square root of the sample size.

We found earlier that the probability that a birthweight is less than 2500 grams is .0863.

(c) Determine the probability that the average birthweight in a random sample of size *n*=2 babies is less than 2500 grams. [*Hint*: Determine the std dev of this average, and then draw a sketch, standardize, and use the normal table as always.]

(d) Is this probability smaller, larger, or the same as .0863? Explain why this makes sense.

(e) Repeat for the average birthweight in a random sample of size *n*=4.

(f) Determine the probability that a newborn weighs between 3000 and 3500 grams.

(g) Would you expect this probability to be larger, smaller, or the same concerning the average birthweight in a sample of n=20 babies? Explain.

(h) Calculate this probability, and comment on how it compares to the probability for a single baby's birthweight.

Now we will consider the case in which the population does not follow a normal distribution.

Example: Sampling Words (cont.)

We will use the "Sampling Words" applet to take repeated random samples from the population of words in the Gettysburg Address.

(a) How would you describe the shape of the population distribution?

(b) The mean word length is 4.29 and the standard deviation is 2.12. Are these parameters or statistics? What symbols do we use for them?

(c) Take one random sample of size n=1. Then take another. Then take 498 more for a total of 500 samples. Describe the shape of the sampling distribution of sample means, and report the mean and standard deviation of the sample means. Are they close to the population values?

(d) Now take a random sample of size n=5. Then take another. Then take 498 more for a total of 500 samples. Describe the shape of the sampling distribution of sample means, and report the mean and standard deviation of the sample means. Are they close to the population values?

(e) Now take a random sample of size n=20. Then take another. Then take 498 more for a total of 500 samples. Describe the shape of the sampling distribution of sample means, and report the mean and standard deviation of the sample means. Are they close to the population values?

Key Result 3 (Central Limit Theorem): Suppose that the population distribution of a quantitative variable has mean μ and standard deviation σ, and suppose that a random sample of size n is taken from the population. Then the sampling distribution of the sample mean $\bar{y}$ is approximately normal, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, provided that either the population itself is normal or the sample size is large ($\geq$30 as a rule-of-thumb).

Note: Key Result 3 includes Key Result 2 as the special case where the population itself follows a normal distribution.

Example: Potato chip bag weights
Suppose that the advertised weight of bag of potato chips is 12 ounces. Suppose further that individual bag weights vary according to a normal distribution and that the standard deviation of these weights is 0.4 ounces. Suppose that a consumer group takes a random sample of n=40 bags and finds an average weight of 11.8 ounces. How surprising would such a small sample mean be, if the population mean were really 12 ounces? Does this sample cast doubt on the manufacturer's claim? How important is the assumption that bag weights vary according to a normal distribution? Explain, and support your conclusion with an appropriate calculation.