Take a sample to find the percent of people who like fried Twinkies. How likely is it that the population percent is more than 50%?

General Education Statistics
Class Notes
Distribution of the Sample Proportion (Section 8.2)

In the last section, we investigated probabilities that a sample mean was less than or greater than a certain amount. We were justified in using the methods from chapter 7 because these means are normally distributed. Here, we do nearly the same thing but with sample proportions (or percentages). We start with some definitions.

Definition: Population proportion: The proportion or percent of individuals that share a characteristic in the *population* is the **population proportion**. We denote this percentage by p.

Definition: Sample proportion: Suppose that a random sample of size n is obtained from a population in which each individual either does or does *not* have a certain characteristic. (That means two mutually exclusive outcomes.) The **sample proportion,** denoted by $\hat{p}$ (pronounced "p-hat"), is given by $\hat{p} = x/n$ where x is the number of individuals in the sample that share the characteristic. As we know, the sample proportion is used to estimate the population proportion.

The idea of two mutually exclusive outcomes harkens back to the binomial distribution.

expl 1: In a random sample of 125 Americans, 28 of them have tried and like fried Twinkies. Find the sample proportion and label it as $\hat{p}$.

What does this tell us about the whole population?

Since a different random sample would likely yield a different result, we say that the sample proportion is a random variable and is fair game for the statistics we have been performing.

> This means that we can use our normal probability methods.

Sampling Distribution of the Sample Proportion $\hat{p}$:

For a simple random sample of size n with a population proportion p,

1. The shape of the sampling distribution of $\hat{p}$ is approximately normal *if* $np(1-p) \geq 10$.

2. The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$.

3. The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

> Underlying this all is the fact that the sample must be a *random* sample. Also, the sample values must be *independent*. When sampling without replacement, as with a simple random sample, we should verify that the sample size is less than 5% of the population size to ensure independence.

expl 2: Suppose we happen to know the *true* proportion of Americans who have tried and like fried Twinkies is 18%. Answer the following questions.

a.) Is this 18% a value of p or $\hat{p}$? Why?

b.) What is $\mu_{\hat{p}}$?

c.) Calculate $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$ for a sample size of 125.

d.) If we sample 125 Americans, will the sampling distribution of $\hat{p}$ be normal? In other words, verify that $np(1-p) \geq 10$.

> Notice we use p here, not $\hat{p}$.

e.) In a random sample of 125 Americans, what is the probability that the sample proportion is more than 22%?

expl 3: Suppose a town has 40,000 residents. The true proportion of residents who support the installation of a statue of their fine mayor is 14%. Pollsters ask a sample of 1500 residents if they support the statue. Check to see if the sampling distribution of $\hat{p}$ is normal by doing the following.

a.) Check $n \le .05N$.

> Part *a* checks that the sampled individuals are independent.

b.) Check that $np(1-p) \ge 10$.

expl 4: A researcher studying public opinion of proposed Food Stamp policy changes obtains a simple random sample of 50 Americans and asks them whether or not they support the changes. Assume the *true* proportion of Americans who support the changes is 20%. In order for the sampling distribution of $\hat{p}$ to be normal, how many more Americans should the researchers contact?

> We need $np(1-p) \ge 10$.

> To start, ignore the first 50 people sampled and just find n needed to make $np(1-p) \ge 10$.

expl 5: Exit polling is often used to estimate the result of a vote prior to the votes being tallied. Let's investigate its perils. A certain city has a voting population of 50,000. The people of this city are voting on whether or not to build a community center. For this referendum to pass, it needs more than 50% of the vote. An exit poll of 300 people is chosen and it is found that 160 of them (or 53.3%) voted for the community center. If the true vote is 49% (this is the population proportion and it means that the referendum would *not* pass), find the probability that a sample of size 300 yields a result of 160 or more people voting for the measure. Use this to comment on the perils of exit polling. We will do this in steps.

a.) Verify that the sampling distribution of $\hat{p}$ is normal. In other words, check that $np(1-p) \geq 10$. Also, check $n \leq .05N$ to be sure that the exit poll participants can be considered independent.

b.) Find the mean and standard deviation of the sampling proportion $\hat{p}$.

c.) Now, we want to find the probability that a sample of size 300 yields a result of 160 or more people voting for the measure (53.3%) while the true proportion is .49 (which dooms the measure). Or rather, we need to find $P(\hat{p} \geq .533)$. Draw a normal curve and shade it appropriately.

d.) If this probability is *not* unusual, we do *not* want to rely on our sample results too heavily. After all, it would *not* be unusual to get an exit poll result that was opposite the actual vote. What should we conclude?

expl 6: The Pew Research Center recently reported that 18% of women aged 40-44 have never given birth. Suppose a random sample of 250 women aged 40-44 results in 58 indicating that they have never given birth ($\hat{p}$ = .232). Is this evidence that the true proportion of women aged 40-44 who have never given birth has increased since the Pew study?

We take the Pew proportion as *p*. Then, assuming this to be the true proportion, find $P(\hat{p} \geq .232)$.

This probability tells us how likely it is that a sample of this size produces a result that is this far or farther from the assumed "true" percentage of 18%.

If it is **unusual** to get this result, then we have evidence that the true proportion is *not really* 18%.

What must be true of the probability for the event to be considered unusual?

But, it *could* be that the sample *just happened* to get a whole bunch of women who have never given birth. Random happens!

So we will *not* make a definitive statement like "the percentage *has* changed."

Worksheet: Finding Probabilities Involving Sample Means and Proportions:
This worksheet will help you practice the skills from this chapter.