

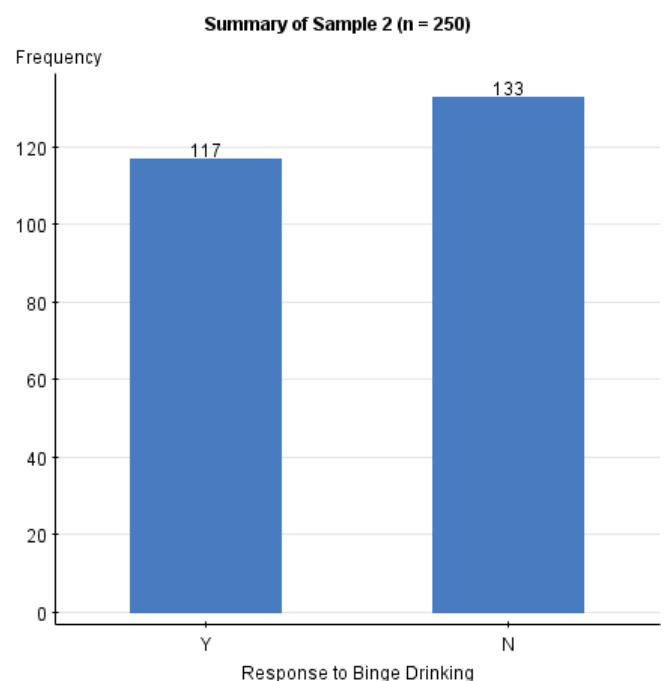
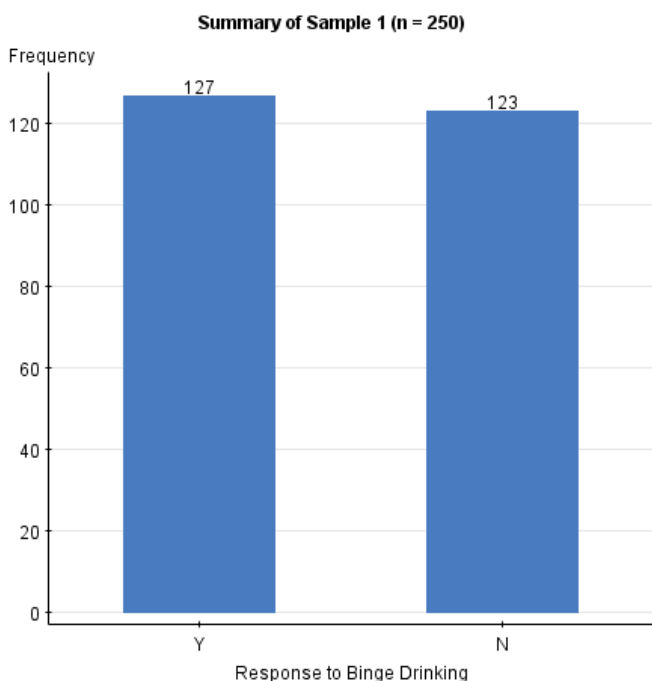
Packet 3: Sampling Distribution of the Sample Proportion

Textbook pages: 399 – 406

After completing this material, you should be able to:

- explain what the symbols \hat{p} , p , $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$ represent.
- describe the sampling distribution of the sample proportion by discussing its shape, mean, and standard deviation.
- find probabilities associated with various sample proportions based on the sampling distribution.
- make inferences from the probability and explain the reasoning.

A university is concerned with the percentage of its students who binge drink. Two different campus offices take samples in order to investigate the severity of the problem. Students in each sample were asked whether or not they had engaged in binge drinking (5 drinks at a sitting for men, 4 for women) in the past month. Results from the two surveys are summarized in the relative frequency bar graphs below:

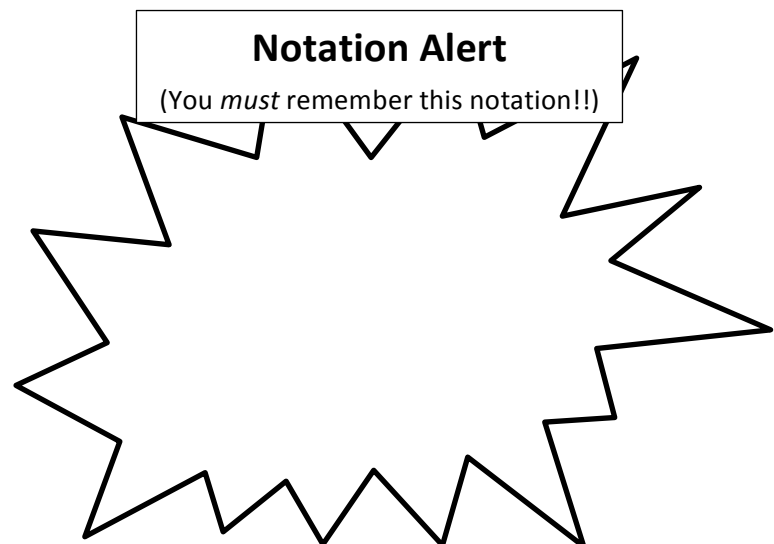


For each of the two samples, determine the sample proportion who responded that they had engaged in binge drinking over the past month.

Sample 1:

Sample 2:

Why are the two sample proportions different?



What exactly is a **sampling distribution** and why is it important?

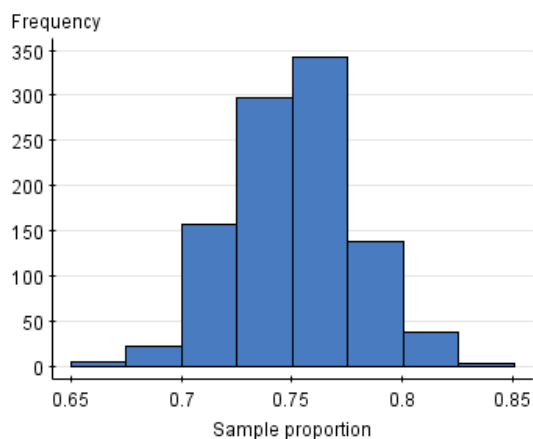
Example: According to a 2005 newspaper report about financial aid for college students, 75% of all full-time degree-seeking students receive some form of financial aid. Given the recent financial crisis, an economist conjectures that the percentage receiving some form of financial aid has increased. In order to test his conjecture, he plans to sample 265 full-time degree-seeking students to determine the proportion that receive financial aid.

Two numbers are given in the example. Assign the appropriate notation (based on the previous page) to these values.

What is the conjecture that the economist is trying to find support for?

Suppose he takes a sample of 265 college students. Which would give more support to his conjecture – finding that 204 students received financial aid or that 215 students received financial aid? Explain your choice.

Instead of taking a single sample of 265 students, suppose that 1000 different samples of 265 students were taken. A sample proportion from each sample was computed and summarized in the graph below.



What seems familiar about the shape of this graph?

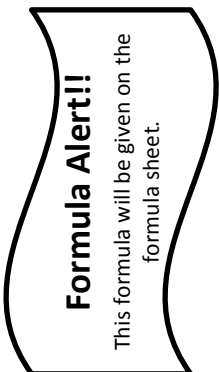
When asked to **describe** the sampling distribution of the sample proportion, the following three characteristics must be addressed:

-
-
-

Why are we using the notation $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$ instead of just using μ and σ ??

Back to the example: Use these characteristics to describe the sampling distribution of the sample proportion of full-time degree-seeking students who receive some form of financial aid.

The fact that the sampling distribution of the sample proportion is *normally distributed* is important – we know how to find probabilities from the normal distribution. To do this, though, we will need to modify our formula for the z-score to reflect that we are now dealing with sample proportions:



Let's go back to the example and see how this formula is used ...

Assuming the newspaper's claim is correct, find the **probability** of observing 204 or more students on financial aid and the **probability** of finding 215 or more students on financial aid. Of the two probabilities, which one gives more support to the conjecture that the proportion of students on financial aid has increased? Explain.

Using a probability to make a decision (rule of thumb):

Example 2: According to an article on webmd.com, 28.6% of Kentucky residents smoked in 2000. After significant advertising campaigns by the American Cancer Society, a researcher would like to know if the proportion of smokers has decreased. A random sample of 672 Kentucky residents is taken, and each is asked whether or not they smoke.

What is the conjecture that we would like to find evidence to support?

Completely **describe the sampling distribution** for the sample proportion of Kentucky residents that smoke when samples of size 672 are taken.

Suppose that a sample of 672 residents is taken, and 27.9% smoke. Assuming the advertising campaigns have had no effect, what is the probability of observing a smoking percentage of 27.9% or less in this sample?

Based on the probability, what conclusion would be reached concerning the conjecture which was made?

Example 3: According to census estimates in 2010, approximately 72% of all US citizens were registered to vote. In the recent presidential election, a push was made to encourage those who were not registered to vote to do so. To see if this has had an effect on voter registration, a random sample of 300 American adults is taken, and the proportion who are registered to vote is recorded.

What variable was recorded? Is this a categorical or quantitative variable?

What is the conjecture that we would like to find evidence for?

Completely **describe the sampling distribution** for the sample proportion of potential voters that are registered to vote if samples of size 300 are taken.

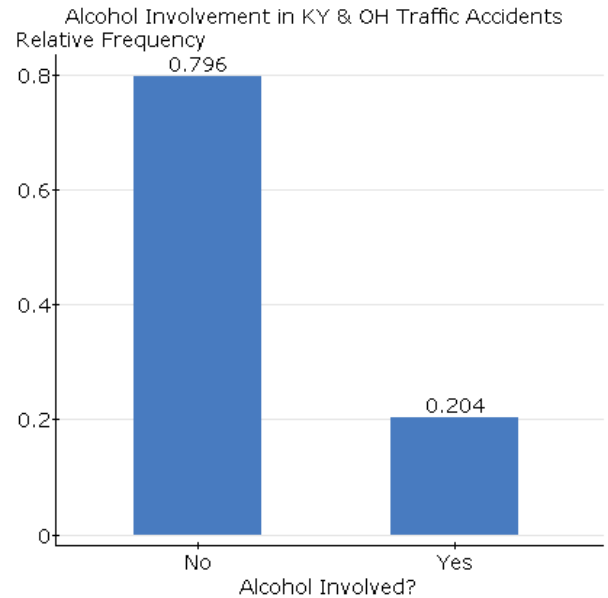
Suppose that a sample of 300 potential voters is taken, and 228 registered voters are found. Assuming the claim is true, what is the probability of obtaining a sample where more than 228 individuals are registered voters?

Based on the probability, what conclusion would be reached about the conjecture which was made?

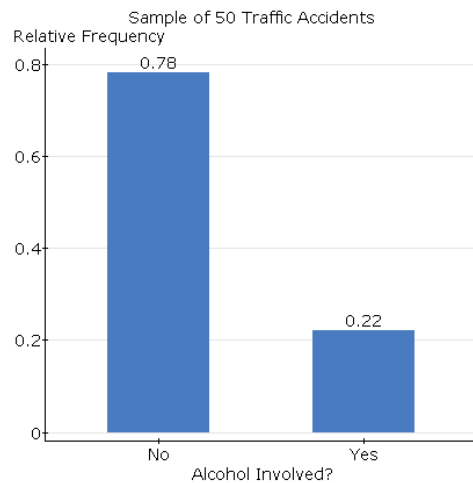
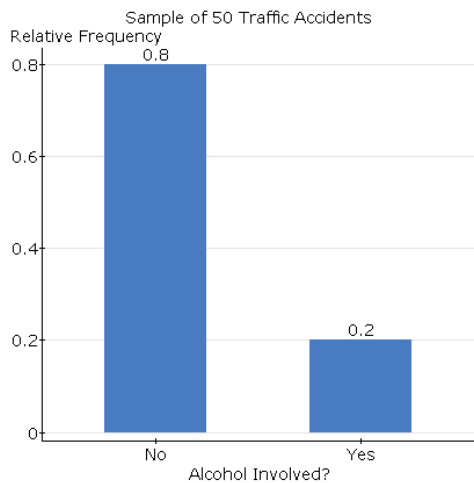
Video Notes – Sampling distributions are a difficult concept. Intuitively, we understand that different samples will lead to different results (which means different statistics), but it is more difficult to think about describing the pattern we expect to see in those statistics.

Example: According to the National Highway Traffic Safety Administration, there were a total of 3444 traffic crashes in Kentucky & Ohio in 2013. A variety of data was recorded for each accident, including whether or not alcohol was involved. The data collected for this population of all traffic crashes is summarized at the right.

What notation is assigned to the value 0.204? Explain why this notation is used.

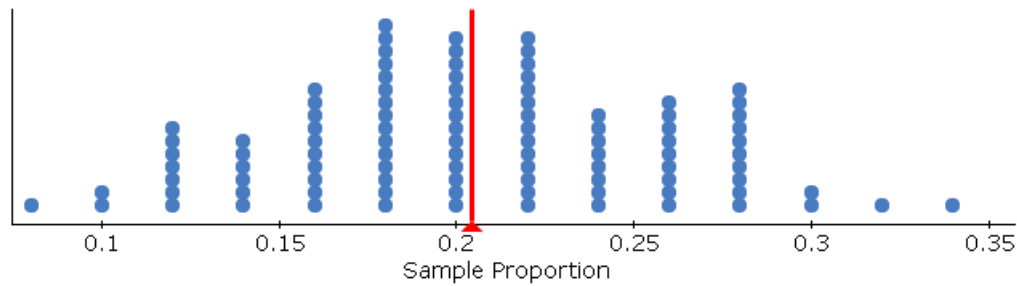


From the population of 3444 traffic accidents, two different samples of 50 traffic accidents are sampled. The results from the samples are summarized in the bar graphs below:



- Why do we expect the bar graphs summarizing these two samples to look similar (but not exactly the same) as the population?
- What notation is assigned to the values 0.2 and 0.22? Explain why this notation is used.
- Why are the sample proportions, 0.2 and 0.22, not the same values?

If we were to continue taking more samples of size 50 from the population, we could use the sample proportions to construct the sampling distribution – this distribution will describe what values of \hat{p} we expect to obtain from samples of size 50. The sampling distribution generated is below:



- What shape does the sampling distribution of the sample proportion have?

- What value is the sampling distribution centered over? Does this agree with the formula we had in class?

Example: The National Safety Council released a report in 2010 stating 25% of all traffic crashes involve drivers using cell phones and texting. The council would like to determine if this percentage has increased over the past three years. A random sample of 250 traffic crashes occurring in 2015 was selected, and it was determined whether or not the driver was using a cell phone during the accident.

- What variable will be recorded for each traffic crash sampled?

- What conjecture has been made by the National Safety Council?

- Describe the sampling distribution for the proportion of accidents involving a texting driver when samples of size 250 traffic crashes are taken.

- Using the description of the sampling distribution, draw and label the model which describes what values of the sample proportion we expect to observe.

- When the random sample of 250 traffic crashes was taken, it was found that 76 of these accidents involved the use of a cell phone. Based on the distribution describe above, what is the probability of finding 76 or more accidents involving the use of a cell phone?

- Based on the probability found above, write a conclusion which addresses the conjecture that has been made.