Math 247: Sampling Distributions (Section 7.2)

The purpose of this section and the next is to understand what a **Sampling Distribution** is and what it tells us about **sampling variability**. We'll use that knowledge later to make inferences about populations by using sample data.

Example: Research Question: What proportion of our class considers themselves a "Dog Person" (can't imagine life without a dog; "Dogs are the soul of my being.")

If we draw a sample of 5 students and find out how many of them are a "Dog Person" and use that to say what's true for the entire class (inference!), how reliable will this inference be? This is the simple question with a pretty complicated answer and it will take us a while to get enough tools together to be able to figure this out.

Vocabulary and Notation:

Population:

Variable of Interest:

Parameter:

Sample:

Estimator of the parameter:
(sample statistic)

"Success" = ____

Types of Samples:

If you just asked the 5 students nearest you whether they are a dog person, would this be a random sample? Why or why not?

This is called a ____

To get a Simple Random Sample (SRS) of 5 students, we will first

1. Assign a number to each student in the class
2. Use a Random Number Generator to select 5 numbers.

Explain why this will give us an SRS:

Creating an SRS:

Use the StatCrunch Random Number Generator (Applets, Random Numbers) to get a sample of 5 numbers, identify the people assigned to those numbers then find the proportion of Dog People in the sample.

Random Numbers:

D or N:

Sample Proportion:
("Estimator")

Now that we have a sample proportion ("estimator"), can we confidently say that we know precisely what the proportion of Dog People is for the entire "population" (our class)? Why or why not?

Each table will find a SRS and use the SRS to find the proportion of Dog People in that sample. Record the results here (use proper notation!):

Make a dotplot of the sample <u>proportions</u>:

Discuss Sampling Variability:

The Big Reveal: Class Census Data: Dog:_______ Not Dog:_______ Total: ___________

Population Proportion of Dog People: _____________________

How well did our samples predict what the true population proportion is?

How are the sample proportion values (dots) positioned relative to the true population proportion?

Common sense: Most of the sample values should be near the ____________________________________

If we gathered together all possible samples (without replacement) of size n = 5 from a population of 30, there would be 142,506 of these samples, so 142,506 dots that would go onto our frequency distribution graph!

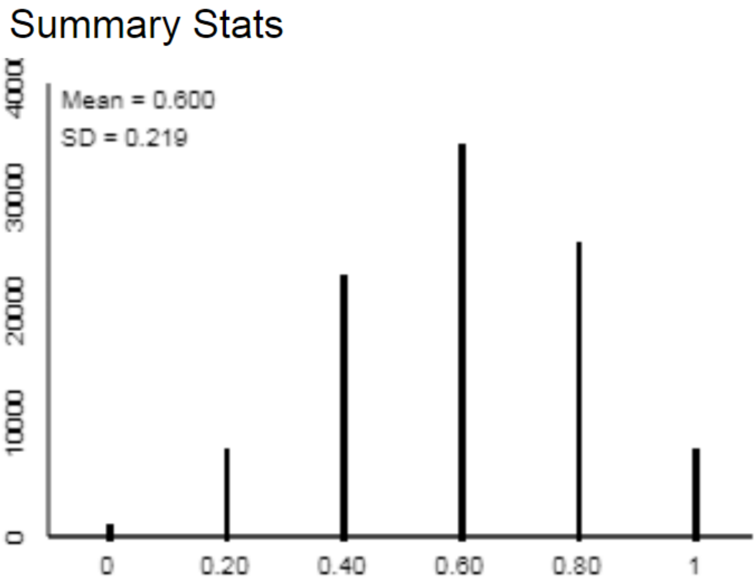
The set of <u>all</u> these sample proportions is called the "**Sampling Distribution of** $\hat{p}$ ".
(Technically, to form the theoretical sampling distribution, we'd have to sample with replacement, in which case there would be 24,300,000 samples of size 5!)

Suppose we knew 60% of the population of interest (this class) considered themselves "Dog People". Using the **<u>Reese's Pieces applet</u>**, find the sampling distribution graph for 100,000 samples of size n = 5.

What is the center of this graph?

How does the center relate to the <u>population proportion</u>?

.



Common sense: Most of the sample values should be near the ______________________________________

Mathematical Bias: This is not bias found by a survey method, but a bias created by the mathematical process itself (like the process of creating a sampling distribution).

Bias is measured as the distance between the **center** (<u>the mean value)</u> of the **estimator** ($\hat{p}$ is the "estimator" in this case) and the actual population **parameter** being estimated (p in this case).

Unbiased Estimator: When the <u>mean</u> of a sampling distribution of a statistic is the same value as the parameter itself, then we say the statistic is an "unbiased estimator".

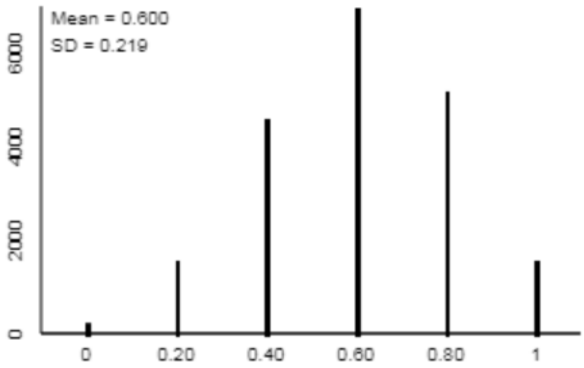
Do you think $\hat{p}$ is an <u>unbiased estimator</u> of p ? Explain.

Measuring Shape and Variability:

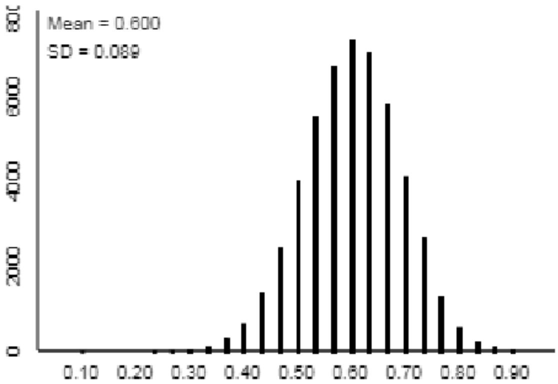
To see how the sample size affects the shape and spread of the Sampling Distribution, we'll use the Reese's Pieces applet to graph 10,000 samples for p = .6, with sample size of 5 and 30 and observe how spread out the distribution is.

Graphs for n = 5 and n = 30 are given:

n = 5



n = 30



Compare the **<u>shapes</u>** of the distributions. What changes in the shape as we <u>increase the sample size</u>?

Compare the **<u>spread</u>**. What changes in the spread of the distribution as we <u>increase the sample size</u>?

What number measures the spread of a distribution?

The *<u>standard deviation</u>* <u>of the</u> *<u>sampling distribution</u>* is called the "**<u>Standard Error</u>**".

The **Standard Error**, SE: $SE = \sqrt{\frac{p(1-p)}{n}}$ Other notation: $\sigma_{\hat{p}}$

Find the Standard Error for the n = 5 and n = 30 (using p = .6) and relate it to the graphs above.