Sampling Distribution for the Sample Proportion

In this module we deal with categorical variables to calculate and estimate proportions of samples.

Categorical Variable:

Category of Interest:

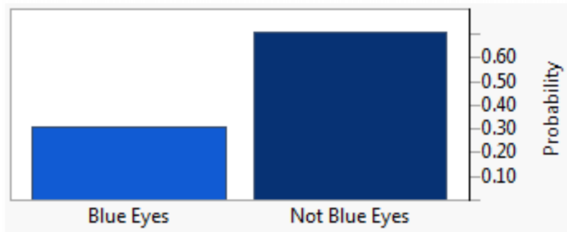
So an example could be that we are interested in studying voting patterns in the most recent presidential primary. Our category of interest could be whether or not people voted in the democratic primary rather than the republican primary.

	Population **Parameter**	Sample **Statistic**
Notation	p	$\hat{p}$
Description	The true proportion in the population. Usually this quantity is unknown and can only be described using words.	The proportion calculated from the sample… can be computed as $\hat{p} = \frac{x}{n}$ where n is the total number of members in the sample and x is the number of sample members in the category of interest.

Some examples of parameters might include:

-
-
-

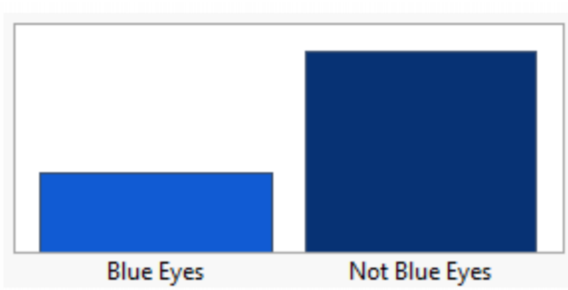
The ____ is the distribution of the categorical variable in the **population**. This distribution is usually ____________________. Suppose 30% of all people have blue eyes. Therefore, the distribution of the categorical variable of eye color in the **population** is as follows:



Some examples of proportion statistics would be:

-
-
-

The ____ is the distribution of the categorical variable in the **sample**. This distribution generally ____________________ be calculated from the sample data. Suppose that in a randomly selected sample of 500 people, 140 have blue eyes. Therefore the distribution of the categorical variable of eye color in this **one** sample of 500 is given below:



Frequencies

Level	Count	Prob
Blue Eyes	140	0.28000
Not Blue Eyes	360	0.72000
Total	500	1.00000

N Missing 0
2 Levels

Sampling Distributions

Recall: the ________________________________ is the variability we expect to see between samples. Different samples have different individuals included so we expect to see different sample proportions.

Sampling Distribution for the Sample Proportion:

Creating a Sampling Distribution of Sample Proportions:

1.

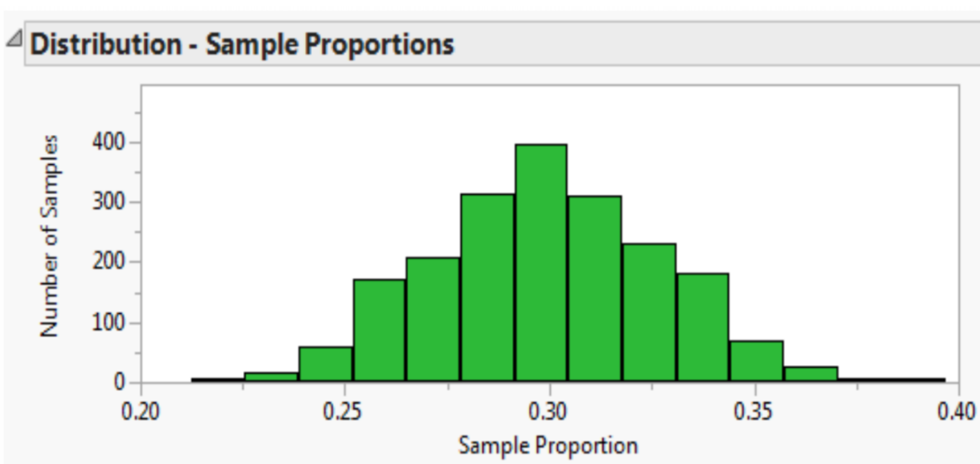
2.

3.

A sampling distribution allows us to see how statistics vary from sample to sample.

It's important to remember that a Sampling Distribution is not the same as the Sample Distribution.

Example *Suppose 30% of people have blue eyes. If we were to obtain a sample of* $n = 250$ *we obtain a sample distribution shown below:*

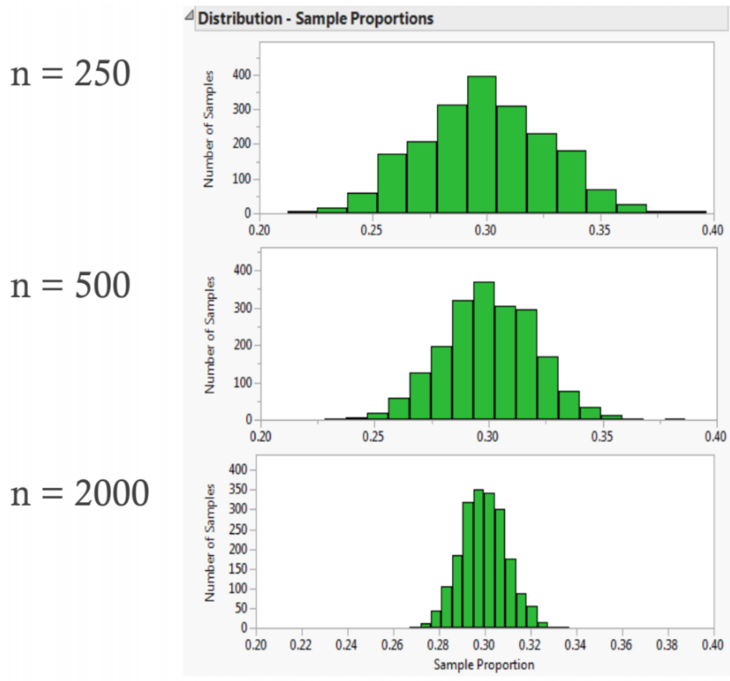


- *Shape:*

- *Mean:*

- *Standard Deviation:*

As we increase the sample size from 250 to 2000, we see a new pattern emerge.



What happened when we increased the sample size?

Characteristics of Sampling Distribution

Center:

Spread/Variability:

Shape: The shape of a sampling distribution follows a Normal Distribution so long as the following conditions are met:

1. Randomization Condition:

2. Succes/Failure Condition:

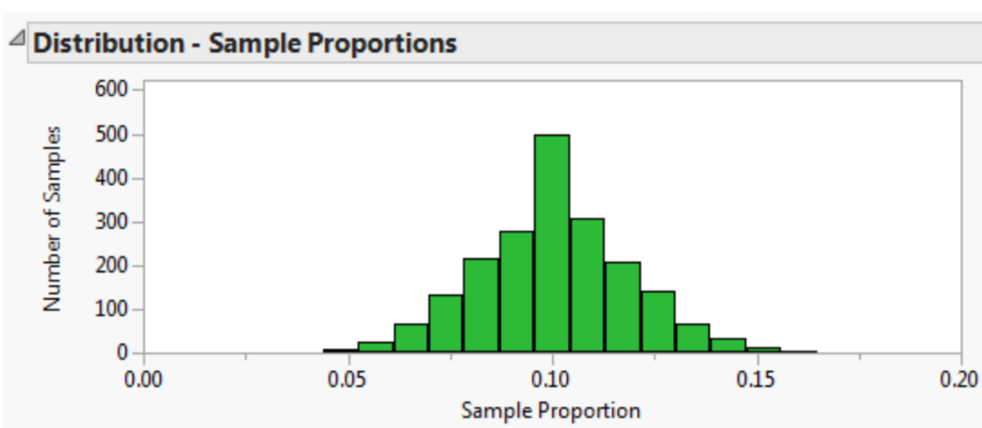
3. 10% Condition:

When asked to determine the sampling distribution for a sample proportion:

First:

Then:

Example *Suppose 10% of people are left handed. Now consider the case where we have random samples of 250 people with the following sample distribution:*



Summary Table - Sample Proportions

Mean:	0.1006
Std Dev:	0.0187

Check the Normality conditions:

Find the sampling distribution of $\hat{p}$:

Since the sample proportions follow a Normal Distribution we can use our knowledge from Chapter 5 to answer the following questions:

1. *What proportion of samples will have a sample proportion of left-handed people greater than 13.5%?*

2. *What is the probability of getting a sample proportion of left-handed people less than 6% in a sample of 250 people?*

3. *95% of all samples of size 250 will have a sample proportion of left-handed people between which two percents?*