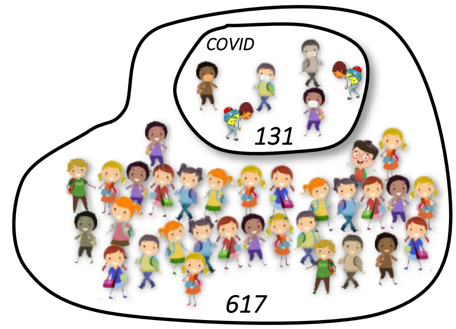
Sec 1.8– Inferences & Conclusions From Data
Population Proportions

Name:

A population proportion, p, is the percentage of a population (written as a decimal) that have some identifiable trait. Let's consider an example. Consider Phoenix High School with exactly 748 students enrolled and we are interested in which students have tested positive for COVID with in the last 3 months. Based on the reports for the school, 131 students had tested positive for COVID-19 in the last 3 months. So, the population proportion can be given by:



Phoenix High School

$$p = \frac{X}{N}$$

Population Proportion: p; Number of Successes: X; Total Population Size: N

$$p = \frac{131}{748} \approx 0.175$$

Normally, it is very difficult to determine the exact population proportion, ***p***, when the population size is large. So, instead we again turn to sampling but a sample will just provide a point estimate, $\hat{p}$, of the actual population proportion. Let's look at a few potential example samples of size 10.

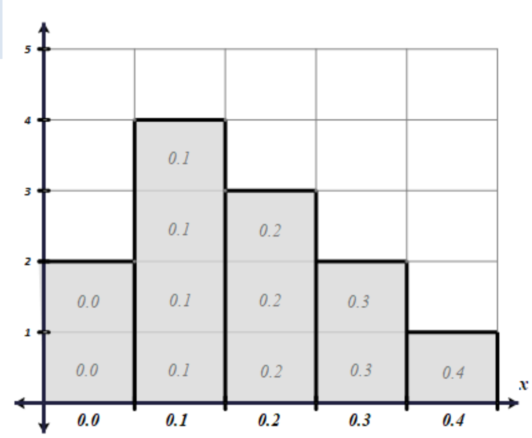
Trial #	10 randomly selected students	Proportion With COVID
1.		$\hat{p}_1 = \frac{x}{n} = \frac{1}{10} = 0.10$
2.		$\hat{p}_2 = \frac{x}{n} = \frac{3}{10} = 0.30$
3.		$\hat{p}_3 = \frac{x}{n} = \frac{1}{10} = 0.10$
4.		$\hat{p}_4 = \frac{x}{n} = \frac{3}{10} = 0.30$
5.		$\hat{p}_5 = \frac{x}{n} = \frac{2}{10} = 0.20$
6.		$\hat{p}_6 = \frac{x}{n} = \frac{2}{10} = 0.20$
7.		$\hat{p}_7 = \frac{x}{n} = \frac{0}{10} = 0.00$
8.		$\hat{p}_8 = \frac{x}{n} = \frac{2}{10} = 0.20$
9.		$\hat{p}_7 = \frac{x}{n} = \frac{0}{10} = 0.00$
10.		$\hat{p}_{10} = \frac{x}{n} = \frac{4}{10} = 0.40$
11.		$\hat{p}_{11} = \frac{x}{n} = \frac{1}{10} = 0.10$
12.		$\hat{p}_{12} = \frac{x}{n} = \frac{1}{10} = 0.10$

It can be mathematically shown that the mean of every possible sample population proportion of size *n* is exactly equal to the population proportion, ***p***. Similarly, we could look at the mean of the sample population proportions that we found at the left:

$$\mu_{\hat{p}} \approx \frac{0+0+.1+.1+.1+.1+.2+.2+.2+.3+.3+.4}{12}$$

$$\mu_{\hat{p}} \approx 0.1667$$

Our calculation above is not exactly the same as, ***p***, because we didn't use every possible sample of size 10. To use every sample of size 10 with replacement would be really difficult because there would be 748^{10} (more than a trillion times a trillion) possible samples. You can see by creating a histogram of our sample population proportions that our distribution is approaching normal. The more of the sample proportions collected and included in the histogram, the more likely the distribution is to be normal. Specifically, when the sample size is less than 5% of the population and $np(1-p) \geq 10$ the distribution is considered to be normal.



Additionally, it can be proven that the standard deviation of the distribution of the means of the sample proportions is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

The proof of $\sigma_{\hat{p}}$ and $\mu_{\hat{p}}$ beyond the scope of this class.

1. New research has suggested that 71% of Americans now get their news content from social media. Describe the sampling distribution if we obtain simple random samples of *n = 60*.

   a. Verify the sampling distribution is considered to be normal.

   b. Determine $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$

   c. If 60 Americans were in a restaurant, what would be the probability that 48 or more of them get their news from social media?

2. Advertising research shows consistently over the last several years 45% of smartphone owners use an Apple iPhone. A classroom has 42 students in it with a smartphone.

   a. Verify the sampling distribution for *n = 42* is considered to be normal.

   b. Determine $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$

   c. In the classroom, 24 of those students had iPhones. What is the probability that many or more had iPhones?

3. A simple random sample survey determined out of 400 California car owners 16 of the owners drove an electric vehicle.

   a. Verify the sampling distribution for *n = 400* is considered to be normal.

   b. Approximate p and $\sigma_{\hat{p}}$

   c. Given a California amusement park parking lot contained 400 cars, what is the approximate probability that there were less than 10 electric vehicles in the parking lot.

4. If a researcher found that 18% of the population in one state uses some form of online grocery shopping. How many people in that state must be included in a sample to ensure the sampling distribution was consider to be normal?

We can also create confidence intervals using population proportions to give a range estimate of ***p*** with a suggested level of confidence. Similar to the way we did for when we considered the distribution of the sampling means in the last section. Consider the confidence intervals we previously used in the last section:

$$\bar{x} - z_{\frac{\alpha}{2}}\,\sigma_{\bar{x}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\,\sigma_{\bar{x}}$$

This formula that we derived in the last section only needs to be slightly modified to:

$$\hat{p} - z_{\frac{\alpha}{2}}\sigma_{\hat{p}} < \boldsymbol{p} < \hat{p} + z_{\frac{\alpha}{2}}\sigma_{\hat{p}}$$

The critical values are still the same as previously:

Confidence Level	90%	95%	99%
Critical value $\left(z_{\frac{\alpha}{2}}\right)$	**1.645**	**1.960**	**2.576**

We can use ***p̂*** is just the proportion of the sample (i.e. $\hat{p} = \frac{x}{n}$) and we have a formula for $\sigma_{\hat{p}}$

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ with the exception we don't usually know ***p*** to use in the formula but since we know ***p̂*** as a point estimate of ***p*** ,we will use it in the formula to approximate $\sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

So the confidence interval for a population proportion could be more explicitly defined as:

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < \boldsymbol{p} < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

5. Given that a simple random survey of 60 students at Central High School found that 45 had at least one other sibling living at home with them. There are approximately 2000 students attending Central High School.
   a. Approximate p and $\sigma_{\hat{p}}$

   b. Show that the distribution of the sample proportions is reasonably normal.

   c. Determine with 95% confidence a range estimate for ***p***.

6. In a local election, a simple random sample of 200 voters was selected. 120 of the voters sampled said they planned on voting for the incumbent, Shawn Bratton. There are approximately 120,000 people in the district that plan to vote.
   a. Approximate p and $\sigma_{\hat{p}}$

   b. Show that the distribution of the sample proportions is reasonably normal.

   c. Determine with 99% confidence a range estimate for ***p***.

   d. If the simple random sample was done correctly, how likely is Shawn to win the election if the law requires a minimum of 50.2% of the vote.

7. A researcher wishes to estimate the proportion of children in the U.S. that have autism. What size sample should be obtained if she wishes the estimate to be within 0.01 and a with 99% confidence and she uses last year's estimate proportion of .023.