FFM 8/19/14:

1.) Among various ethnic groups, the population standard deviation of heights is known to be approximately 3 inches. We wish to construct a 95% confidence interval for the average height of men from Sweden. 48 Swedish men are surveyed. The sample mean is 71 inches. What will happen to the width of the confidence interval obtained if 1000 Swedish men are surveyed instead of 48? Why?

a. It will be wider because ________

b. It will be narrower because __________

c. no change because _________

2.) In a study of the length of time that students take to earn bachelor's degrees, a sample of 1,200 students took an average of 5.9 years to complete their degree. Assume that the population standard deviation is known to be 2.3 years. What would happen to the width of the confidence interval if 300 students were surveyed instead of 1,200? Why?

a. It would be narrower because __________

b. It would be wider because ___________

c. No change because _________

Unit 1, Task 1 – Sample Proportions with Simulation

I. Reviewing some basic definitions:

a. Think about a single bag of M&M's. Does this single bag represent a *sample* of M&M's or the *population* of M&M's?

Sample

b. We use the term *statistic* to refer to measures based on samples and the term *parameter* to refer to measures of the entire population.

If there are 62 M&M's in your bag, is 62 a statistic or a parameter? Statistic $n=62$

If Hershey claims that 24% of all M&M's are blue, is 24% a statistic or a parameter? parameter

proportion

$P=0.24$

Not Blue: 76%

$q=1-p=1-.24$

$q=.76$

II. How many blue candies should I expect in a bag of M&M's?

First, **random samples of 10 candies** were taken last year from bags of M&M's. The count and proportion of each color in the samples were recorded by each Math IV student in a table like this:

Binomials

	Blue "success"	Not Blue "failure"
Count	3	7
Proportion	$\frac{3}{10} \Rightarrow \hat{p} = .3$	$\hat{q} = .7$

The average of the sample proportions obtained by each class for sample size 10 was recorded and complied.

a. Do you think their averages ($\mu_{\hat{p}}$) are closer or farther from the true proportion of blue M&Ms than each individual student found in his/her sample? closer Why do you think that? average → center of data.

Next, **random sample of 25 candies** were taken and recorded by each Math IV student. The average of the sample proportions obtained by each class for sample size 25 was recorded and complied.

Finally, **random samples of 40 candies** were taken and recorded by each Math IV student. The averages of the sample proportions obtained by each class for sample size 40 was recorded and compiled. :

b. Do you think the averages of the 40 candies are closer or farther from the true proportion of blue M&Ms than each individual student found in his/her sample of size 10 and 25? closer Explain: n ↑ smaller error, estimate is more accurate.

Here is the data obtained in one of the class periods:

6th period			
TOTALS:	n= 10	n= 25	n= 40
count:	46	105	165
# of samples:	170	425	680
$\hat{p}$ sample proportion:	27.06%	24.71%	$\boxed{24.26\%}$

c. What is your initial reaction to the three average sample proportions, i.e, which do you think is more accurate? Why? ___________

as n increased, $\hat{p}$ got closer to $P = 0.24$.

III. Sampling Distribution of $\hat{p}$

We have been looking a number of different *sampling distributions* of $\hat{p}$, but we have seen that there is great variability in the distributions. We would like to know that $\hat{p}$ is a good estimate for the true proportion of blue M&M's. However, there are guidelines for when we can use the statistic to estimate the parameter. This is what we will investigate in the next section.

First, however, we need to understand the center, shape, and spread of the sampling distribution of $\hat{p}$.

We know that if we are counting the number of M&M's that are blue and comparing with those that are not blue, then the counts of blues follow a binomial distribution (given that the population is much larger than our sample size). Formulas for the mean and standard deviation of a binomial distribution were discussed in $\boxed{\text{Math III.}}$ $\mu_x = np$ $\sigma_x = \sqrt{npq}$.

Given that $\hat{p} = \frac{X}{n}$, where X is the count of blues and n is the total in the sample, how might we find $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$? Find formulas for each statistic. ____

This leads to the following statement of the characteristics of the sampling distribution of a sample proportion:

The Sampling Distribution of a Sample Proportion: N is 10 times n (at least)

Math IV Choose a simple random sample of size n from a $\boxed{\text{large population}}$ with population parameter p having some characteristic of interest. Let $\hat{p}$ be the proportion of the sample having that characteristic. Then:

- The mean of the sampling distribution is $\mu_{\hat{p}} = p$
- The standard deviation of the sampling distribution is $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

a. Let's look at the standard deviation a bit more. What happens to the standard deviation as the sample size increases?

____

$\sqrt{\frac{pq}{4n}} = \sqrt{\frac{1}{4}}\sqrt{\frac{pq}{n}} =$

b. **Show** a few examples to verify your conclusion. Then use the formula to **explain** why your conjecture is true:

c. If we wanted to cut the standard deviation in half, thus decreasing the variability of $\hat{p}$, what would we need to do in terms of our sample size? Use the formula to explain.

$$\frac{1}{2}\sigma_{\hat{p}} = \sqrt{\frac{pq}{4n}} \qquad \sqrt{\frac{1}{4}} = \frac{1}{2}$$

***Caution:** We can only use the formula for the standard deviation of $\hat{p}$ when the population is at least 10 times as large as the n.

d. For each of the samples taken in part 3, determine what the population of M&M's must be for us to use the standard deviation formula derived above. Is it safe to assume that the population is at least as large as these amounts? Explain.

n	$\times 10$	N
10		100
25		250
40		400

logical thinking

check pop. size

CLT for Proportions:

* IF SRS, size n, from large population ($N \geq 10n$, $\boxed{np \geq 10, \; nq \geq 10}$) with population parameter, p, having some characteristic of interest
* THEN Sampling distribution of sample proportion, $\hat{p}$,
  - is approximately normal
  - has $\mu_{\hat{p}} = p$
  - has $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$ where $q = 1 - p$

$\sigma = \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ if $n \geq 30$.

* NOTE: This approximation becomes more and more accurate as the sample size, n, increases.
* ALSO NOTE: The CLT allows us to use normal calculations to determine probabilities about sample proportions obtained from populations that are not normally distributed.

front: #1, 2, 3, 5