

## 9.2 Sample Proportions

sample proportion has a distribution  
with a mean and standard deviation

<http://www.rossmanchance.com/applets/GettysburgSample/GettysburgSample.html>



Draw 100 samples of size 10. Check each mean and standard deviation in the Long and Noun graphs.

<http://www.rossmanchance.com/applets/senators/samplesenators.html>



Draw 100 samples of size 10. Check each mean and standard deviation in the Republican and Female graphs.

large university lecture hall courses  
% who have had dental fillings

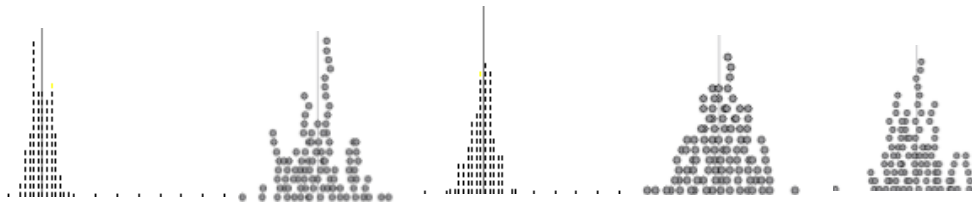
Same % in every auditorium?

No. We usually expect different samples to yield different statistics.



% who have fillings in all groups of 300 students  
graph each % found

probably a bell curve, maybe almost Normal, like these:



sampling distribution of the sample proportion:

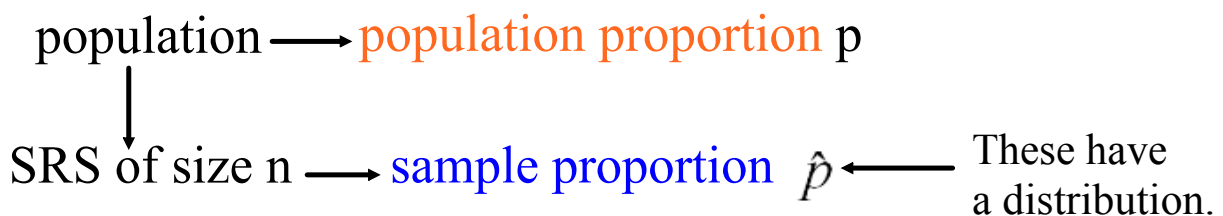
**S** approx. Normal

**O** few, if any

**C**  $\approx$  % who have fillings in population

**S** since  $n=300$ , spread's smaller than if  $n$  had 100 & larger than if  $n$  had been 400

Sampling Distribution of the Sample Proportion is the distribution of sample proportions from all possible samples of this size.



**mean** of the sampling distribution of  $\hat{p}$  is exactly  $p$ .

$$\mu_{\hat{p}} = p$$

**standard deviation** of the sampling distribution of  $\hat{p}$  is

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

as long as the population is at least 10  
times as large as the sample ( $N \geq 10n$ ).

Why  $N > 10n$ ? Because of the methods we use. Larger samples are so much better that our standard deviation formula is no longer a good estimate. If our sample is more than 1/10 the population, the assumptions, conditions, methods, and formulas we have won't work.

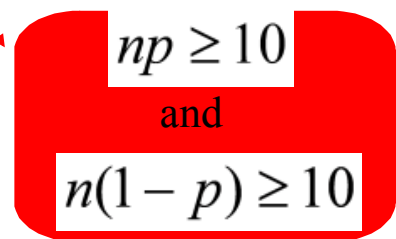
The Success/Failure Condition for Normal Approx.  
If  $np \geq 10$  &  $n(1-p) \geq 10$ , then the sampling distribution of  $\hat{p}$  is approximately Normal.

## Central Limit Theorem

Draw an SRS of size  $n$  from **any population whatsoever** with proportion  $p$ . When  $n$  is **large enough**, the sampling distribution of sample proportion  $\hat{p}$  is close to the

normal distribution  $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ .

Ch. 9  
← %



$np \geq 10$   
and  
 $n(1-p) \geq 10$

$$N\left(np, \sqrt{np(1-p)}\right)$$

Ch. 8  
← counts

Ex. Consider 100 tosses of a coin.

- The mean proportion of tails =  $\mu_{\hat{p}} = p = .5$

- Verify n for the standard deviation:

100 tosses is less than 10% of all possible tosses.  
The standard deviation for the proportion of tails =

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.5)(1-.5)}{100}} = .05$$

- Verify Success/Failure Condition for Normal Approximation:

$$np = 100(.5) = 50 \geq 10 \checkmark$$

$$n(1-p) = 100(1-.5) = 50 \geq 10 \checkmark$$

The sampling distribution of the percent of tails is approximately Normal:

So, by the 68-95-99.7 Rule,

- There's a 68% chance that the proportion of tails will be between 0.45 and 0.55 for 100 tosses.
- There's a 95% chance that the proportion of tails will be between 0.40 and 0.60 for 100 tosses.
- There's a 99.7% chance that the proportion of tails will be between 0.35 and 0.65 for 100 tosses.

It would be unusual to toss a coin 100 times and have under 35% tails or over 65% tails.



Ex. A psychologist studies the existence of ESP by having a participant guess which of 6 cards the psychologist has selected. This is repeated 90 times per participant with many people.

For guessing,  $P(\text{correct guess}) = 1/6$ .

- The mean proportion of correct guesses =  $\mu_{\hat{p}} = p = \frac{1}{6} \approx .1\bar{6} \approx .17$

- Verify  $n$  for the standard deviation:

$n = 90$ , That's under 10% of all a possible guesses.

The standard deviation for the proportion of correct guesses =

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(\frac{1}{6})(\frac{5}{6})}{90}} = .04$$

- Verify Success/Failure Condition for Normal Approximation:

$$np = 90\left(\frac{1}{6}\right) = 15 \geq 10 \checkmark$$

$$n(1-p) = 90\left(\frac{5}{6}\right) = 75 \geq 10 \checkmark$$

The sampling distribution of the percent of correct guesses is approximately Normal:

So, by the 68-95-99.7 Rule,

There's a 99.7% chance that participant will guess between  $\frac{.17}{.05}$  5 % & 29 % of the 90 guesses correctly.

Almost everyone would guess 5 % to 29 % correctly.

If a participant guesses more than 29 % correctly, that is unusually high (fewer than 5 % correct would be unusually low).