# Chapter 7: Sampling Distributions

## Section 7.2

## Sample Proportions

**The Practice of Statistics, 4th edition – For AP***

**STARNES, YATES, MOORE**

# Chapter 7
# Sampling Distributions

- 7.1 What is a Sampling Distribution?
- **7.2 Sample Proportions**
- 7.3 Sample Means

## Section 7.2
## Sample Proportions

**Learning Objectives**

After this section, you should be able to…

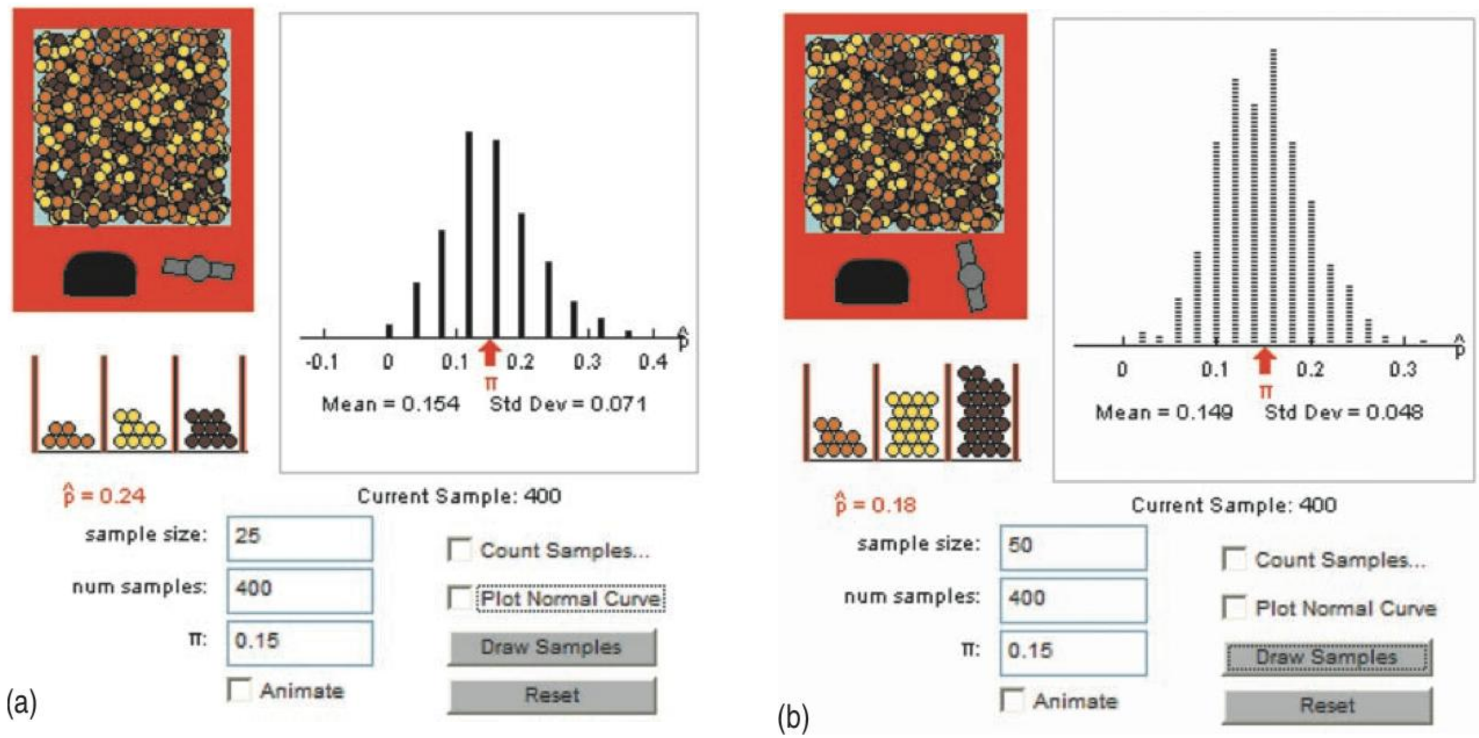
- ✓ FIND the mean and standard deviation of the sampling distribution of a sample proportion
- ✓ DETERMINE whether or not it is appropriate to use the Normal approximation to calculate probabilities involving the sample proportion
- ✓ CALCULATE probabilities involving the sample proportion
- ✓ EVALUATE a claim about a population proportion using the sampling distribution of the sample proportion

## The Sampling Distribution of $\hat{p}$

How good is the statistic $\hat{p}$ as an estimate of the parameter $p$? The sampling distribution of $\hat{p}$ answers this question.

Consider the approximate sampling distributions generated by a simulation in which SRSs of *Reese's Pieces* are drawn from a population whose proportion of orange candies is either 0.45 or 0.15.

What do you notice about the shape, center, and spread of each?



## The Sampling Distribution of $\hat{p}$

What did you notice about the shape, center, and spread of each sampling distribution?

**Shape**: In some cases, the sampling distribution of $\hat{p}$ can be approximated by a Normal curve. This seems to depend on both th sample size $n$ and the population proportion $p$.

**Center**: The mean of the distribution is $\mu_{\hat{p}} = p$. This makes sens because the sample proportion $\hat{p}$ is an unbiased estimator of $p$.

**Spread**: For a specific value of $p$, the standard deviation $\sigma_{\hat{p}}$ gets smaller as $n$ gets larger. The value of $\sigma_{\hat{p}}$ depends on both $n$ and $p$.

There is and important connection between the sample proportion $\hat{p}$ and the number of "successes" $X$ in the sample.

$$\hat{p} = \frac{\text{count of successes in sample}}{\text{size of sample}} = \frac{X}{n}$$

Sample Proportions

## The Sampling Distribution of $\hat{p}$



In Chapter 6, we learned that the mean and standard deviation of a binomial random variable *X* are

$$\mu_X = np \qquad \sigma_X = \sqrt{np(1-p)}$$

Since $\hat{p} = X/n = (1/n)\cdot X$, we are just multiplying the random variable $X$ by a constant $(1/n)$ to get the random variable $\hat{p}$. Therefore,

$$\mu_{\hat{p}} = \frac{1}{n}(np) = p$$

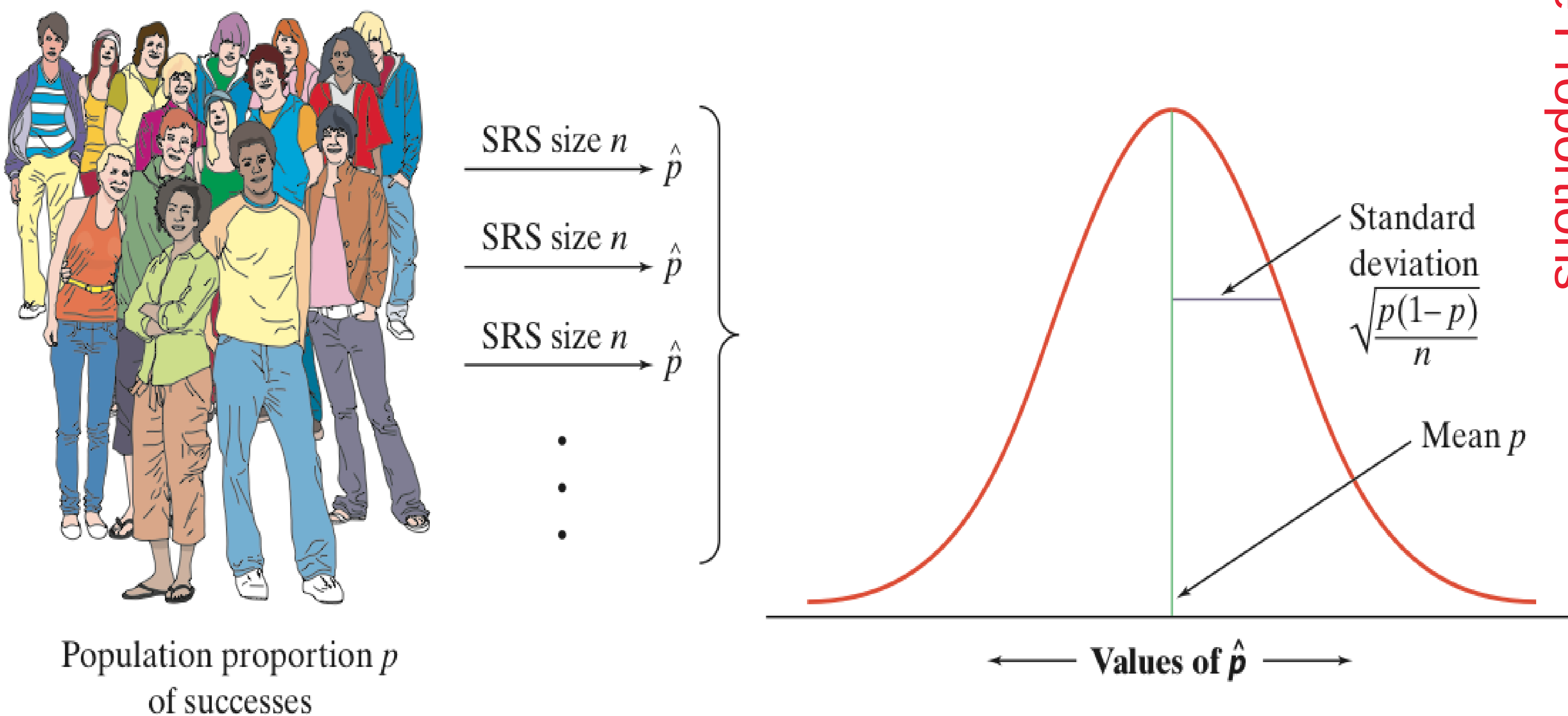
$\hat{p}$ is an unbiased estimator of $p$

$$\sigma_{\hat{p}} = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}}$$

As sample size increases, the spread decreases.

## The Sampling Distribution of $\hat{p}$

We can summarize the facts about the sampling distributior of $\hat{p}$ as follows:



## Using the Normal Approximation for $\hat{p}$

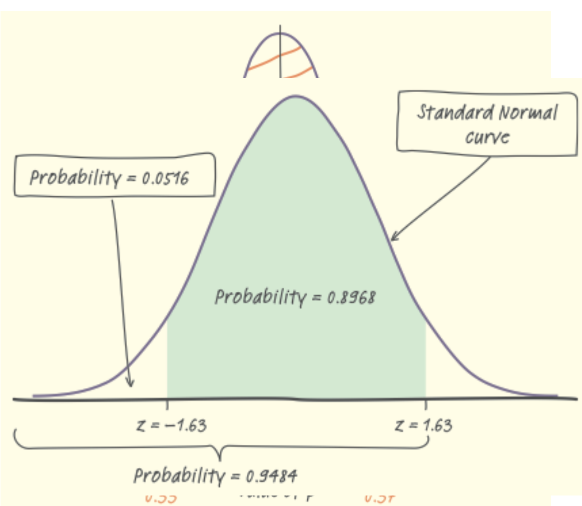
Inference about a population proportion $p$ is based on the sampling distribution of $\hat{p}$. When the sample size is large enough for $np$ and $n(1-p)$ to both be at least 10 (the Normal condition), the sampling distribution of $\hat{p}$ is approximately Normal.

STEP 4

A polling organization asks an SRS of 1500 first-year college students how far away their home is. Suppose that 35% of all first-year students actually attend college within 50 miles of home. What is the probability that the random sample of 1500 students will give a result within 2 percentage points of this true value?

**STATE**: We want to find the probability that the sample proportion falls between 0.33 and 0.37 (within 2 percentage points, or 0.02, of 0.35).

**PLAN**: We have an SRS of size *n* = 1500 drawn from a population in which the proportion *p* = 0.35 attend college within 50 miles of home.




$$\mu_{\hat{p}} = 0.35 \qquad \sigma_{\hat{p}} = \sqrt{\frac{(0.35)(0.65)}{1500}} = 0.0123$$

**DO**: Since $np$ = 1500(0.35) = 525 and $n(1-p)$ = 1500(0.65)=975 are both greater than 10, we'll standardize and then use Table A to find the desired probability.

$$z = \frac{0.33-0.35}{0.123} = -1.63 \qquad z = \frac{0.37-0.35}{0.123} = 1.63$$

$$P(0.33 \le \hat{p} \le 0.37) = P(-1.63 \le Z \le 1.63) = 0.9484 - 0.0516 = 0.8968$$

**CONCLUDE**: About 90% of all SRSs of size 1500 will give a result within 2 percentage points of the truth about the population.

# Alternate Example – Planning for College



The superintendent of a large school district wants to know what proportion of middle school students in her district are planning to attend a four-year college or university. Suppose that 80% of all middle school students in her district are planning to attend a four-year college or university. What is the probability that a SRS of size 125 will give a result within 7 percentage points of the true value?

**STATE**: We want to find the probability that the proportion of middle school students who plan to attend a four-year college or university falls between 73% and 87%. That is, $P(0.73 \le p \le 0.87)$.

**PLAN**: Because the school district is large, we can assume that there are more than 10(125) = 1250 middle school students so

$$\mu_{\hat{p}} = 0.80 \qquad \sigma_{\hat{p}} = \sqrt{\frac{(0.80)(0.20)}{125}} = 0.036$$

We can consider the distribution of $\hat{p}$ to be approximately Normal since $np = 125(0.80) = 100 \ge 10$ and $n(1-p) = 125(0.20) = 25 \ge 10$.

**DO**: $$P(0.73 \le \hat{p} \le 0.87) = normalcdf\,(0.73, 0.87, 0.80, 0.036) = 0.948$$

*(Note: To get full credit when using normalcdf on an AP exam question, students must explicitly state the mean and standard deviation of the distribution as in the Plan step above.)*

**CONCLUDE**: About 95% of all SRSs of size 125 will give a sample proportion within 7 percentage points of the true proportion of middle school students who are planning to attend a four-year college or university.

# Section 7.2
# Sample Proportions

## Summary

In this section, we learned that…

- ✓ When we want information about the population proportion $p$ of successes, we often take an SRS and use the sample proportion $\hat{p}$ to estimate the unknown parameter $p$. The **sampling distribution** of $\hat{p}$ describes how the statistic varies in all possible samples from the population.
- ✓ The **mean** of the sampling distribution of $\hat{p}$ is equal to the population proportion $p$. That is, $\hat{p}$ is an unbiased estimator of $p$.
- ✓ The **standard deviation** of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ for an SRS of size $n$. This formula can be used if the population is at least 10 times as large as the sample (the 10% condition). The standard deviation of $\hat{p}$ gets smaller as the sample size $n$ gets larger.
- ✓ When the sample size $n$ is larger, the sampling distribution of $\hat{p}$ is close to a Normal distribution with mean $p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- ✓ In practice, use this Normal approximation when both $np \geq 10$ and $n(1 - p) \geq 10$ (the Normal condition).

+

# Looking Ahead…

**In the next Section…**

We'll learn how to describe and use the sampling distribution of sample means.

We'll learn about

- ✓ **The sampling distribution of** $\bar{x}$
- ✓ **Sampling from a Normal population**
- ✓ **The central limit theorem**