

Section 8.2: Distribution of the Sample Proportion

Objectives

By the end of this lesson, you will be able to...

1. describe the sampling distribution of a sample proportion
2. compute probabilities of a sample proportion

The Sample Proportion

Consider these recent headlines:

Hispanics See Their Situation in U.S. Deteriorating

Half (50%) of all Latinos say that the situation of Latinos in this country is worse now than it was a year ago, according to a new nationwide survey of 2,015 Hispanic adults conducted by the Pew Hispanic Center. (Source: [Pew Research](#))

Automatic enrollment in 401(k) doesn't take care of everything

Never got around to signing up for the company retirement plan? The boss may have done it for you. Forty-two percent of employers with 401(k) plans automatically enroll new or existing employees in the plans, nearly double the 23 percent from 2006, according to estimates from the 2008 401(k) Benchmarking Survey by the International Foundation of Employee Benefit Plans and Deloitte Consulting. The survey polled 436 employers with workforces of all sizes. (Source: [Chicago Tribune](#))

Stem cell, marijuana proposals lead in Mich. poll

A recent poll shows voter support leading opposition for ballot proposals to loosen Michigan's restrictions on embryonic stem cell research and allow medical use of marijuana. The EPIC-MRA poll conducted for The Detroit News and television stations WXYZ, WILX, WOOD and WJRT found 50 percent of likely Michigan voters support the stem cell proposal, 32 percent against and 18 percent undecided. (Source: [Associated Press](#))

These three articles all have something in common - they're referring to *sample proportions* - 50% of all Latinos, 42% of employers, and 50% of likely Michigan voters, respectively, in the three articles above. Proportions are the number with that certain characteristics divided by the sample size.

In general, if we let x = the number with the specific characteristic, then the **sample proportion**, $\hat{p}$, (read "*p-hat*") is given by:

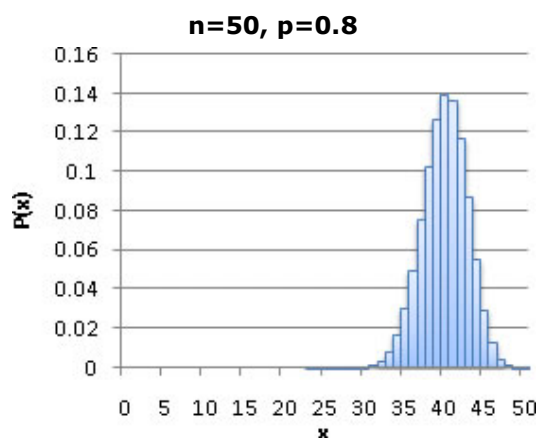
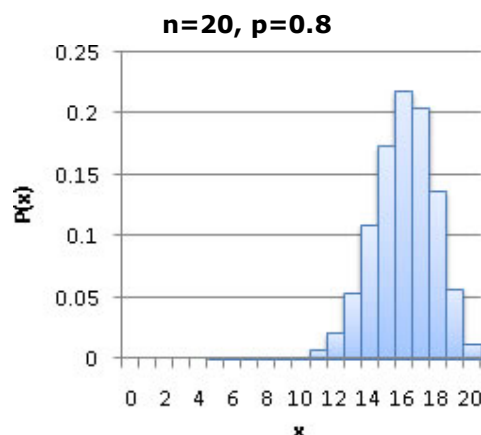
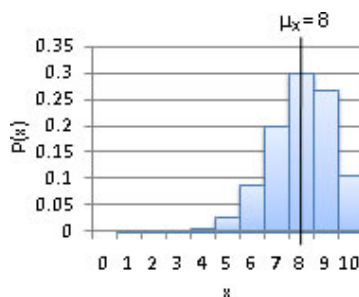
$$\hat{p} = \frac{x}{n}$$

Where $\hat{p}$ is an estimate for the population proportion, p .

Let's focus for a bit on x , the number with that characteristic. If we rephrase that a bit, and consider an individual having that characteristic as a "success", we can see that x follows the [binomial distribution](#).

From [Section 6.2](#), we know that the distribution of a binomial random variable becomes bell-shaped as n increases. The three histograms below demonstrate the effect of the sample size on the distribution shape.

$n=10, p=0.8$



As the number of trials in a binomial experiment increases, the probability distribution becomes bell-shaped. As a rule of thumb, **if $np(1-p) \geq 10$, the distribution will be approximately bell-shaped.**

With our new knowledge of the normal distribution, it appears that if $np(1-p) \geq 10$, x is normally distributed, which implies that $\hat{p}$ is as well. All we need to know, then, is its mean and standard deviation.

Sampling Distribution of $\hat{p}$

For a simple random sample of size n such that $n \leq 0.05N$ (in other words, the sample is less than 5% of the population),

- The shape of the sampling distribution of $\hat{p}$ is approximately normal provided $np(1-p) \geq 10$
- The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$.
- The standard deviation of the sampling distribution of $\hat{p}$ is

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

Now that we know how $\hat{p}$ is distributed, we can use that information to find some probabilities.

Let's see a couple examples.

Example 1

In a typical class, about 70% of students receive a C or better. Out of a random sample of 100 students, what is the probability that less than 60 receive a C or better?

Solution:

Since there are millions of students, 100 is definitely less than 5% of the population.

Because $np(1 - p) = 100(0.7)(1 - 0.7) = 21 \geq 10$, we can say that the distribution of $\hat{p}$ is normal.

We then need to find the mean and standard deviation. From the distribution of the sample proportion, we know

$$\mu_{\hat{p}} = p = 0.7$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{0.7(1 - 0.7)}{100}} \approx 0.046$$

Using StatCrunch, the probability of observing a sample proportion of less than 60/100 = 0.6 is

Mean:	0.7	Std. Dev.:	0.046
Prob(X	<=	0.6	) = 0.014855833
Close		Compute	

So the probability of observing less than 60 out of 100 students is about 0.015 - pretty unusual.

Example 2

A basketball player traditionally makes 85% of her free throws. Suppose she shoots 100 free-throws during practice. Would it be unusual for her to make less than 75?



Source: [stock.xchng](https://www.stock.xchng.com/)

Solution:

Since our player can shoot an unlimited number of free throws, we can assume that 100 is less than 5% of the population. Checking the distribution:

$$np(1 - p) = 100(0.85)(1 - 0.85) = 12.75 \geq 10$$

So the distribution should be normally distributed, with mean and standard deviation

$$\mu_{\hat{p}} = p = 0.85$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} = \sqrt{\frac{0.85(1 - 0.85)}{100}} \approx 0.036$$

Using StatCrunch, with $p = 75/100 = 0.75$:

Mean:	0.85	Std. Dev.:	0.036
Prob(X	<input "="" type="text" value="<="/>	0.75	) = 0.0027366017
		Close	Compute

So the probability that she makes less than 75 out of 100 is about 0.0027. That means she'll make less than 75 out of 100 about 0.27% of the time, which is very unusual.

Here's one for you to try:

Example 3

According to data from the [American Cancer Society](#), about 3.86% of women develop breast cancer between the ages of 40-59.

What is the probability that in a random sample of 500 39-year-old women without breast cancer, more than 20 will develop breast cancer by the age of 60?

[\[reveal answer \]](#)

Since there are about 2 million women aged 39 in the United States, the sample size of 500 is definitely less than 5% of the population. Checking the distribution:

$$np(1-p) = 500(0.0386)(1-0.0386) = 18.6 \geq 10$$

So the distribution should be approximately normal, with mean and standard deviation

$$\mu_{\hat{p}} = p = 0.0386$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.0386(1-0.0386)}{500}} \approx 0.0086$$

Using StatCrunch, the probability of observing a sample proportion of $20/500 = 0.04$ or more is:

Mean:	0.0386	Std. Dev.:	0.0086
Prob(X	=>	0.04	) = 0.43534163
		Close	Compute

So there's a probability of about 0.435 that more than 20 of the 500 39-year-old women will develop breast cancer by age 60.

A note on rounding: We've already seen that rounding is an important topic in this course. In this particular section, rounding can make a significant difference in your calculations. It's important that you take a minute and watch this video regarding the effect of rounding when doing the calculations from this section: [The Dangers of Rounding \(Quicktime or iPod\)](#)