AP Statistics
Unit 05 – Distributions
Day 02 Notes – Sample Proportions (7.2)

Name Key
Period

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 sample proportion varies in all possible samples form the population.

Candy Machine Activity! www.rossmanchance.com (Reese's Pieces Applet)

Facts about the sampling distribution of $\hat{p}$:

- 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 $\sqrt{p(1-p)/n}$ 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 gets smaller as the sample size n gets larger. Because of the square root, a sample four times larger is needed to cut the standard deviation in half.

10% Condition: When taking an SRS of size n from a population of size N, we can use the above formula for calculating standard deviation as long as $n \le \frac{1}{10}N$

sampling without replacement

- When the sample size n is large, the sampling distribution of $\hat{p}$ is close to a Normal distribution with mean p and standard deviation $\sqrt{p(1-p)/n}$. In practice, use this **Normal approximation** when both $np \geq 10$ and $n(1-p) \geq 10$ (the *Large Counts condition*).

Large Counts Condition: As a rule of thumb, we will use the Normal approximation when n is so large that $\boxed{np \geq 10 \text{ and } np \geq 10}$.

Key Concepts:

Sampling Distribution of a Sample Proportion

Choose an SRS of size n from a population of size N with proportion p of successes. Let $\hat{p}$ be the sample proportion of successes. Then:

- 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

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

as long as the 10% condition is satisfied: $n \leq \frac{1}{10}N$

- As n increases, the sampling distribution of $\hat{p}$ becomes **approximately Normal.** Before you perform Normal calculations, check that the Large Counts condition is satisfied:

$$np \geq 10 \text{ and } n(1-p) \geq 10.$$

See page 443 Starnes 5th edition for detailed explanation of mean and standard deviation calculation.

Sometimes it makes sense to sample more than 10% of the population because larger random sample give better information. The formula for calculating a more accurate standard deviation in these cases is called a ***finite population correction (FPC)*** *but we will avoid situations that require this.*

Example:

Bag Check! Thousands of travelers pass through the airport in Guadalajara, Mexico, each day. Before leaving the airport, each passenger must pass through the Customs inspection area. Customs agents want to be sure that passengers do not bring illegal items into the country. But they do no have time to search every traveler's luggage. Instead, they require each person to press a button. Either a red or a green bulb lights up. If the red light shows, the passenger will be searched by Customs agents. A green light means "go ahead." Customs agents claim that the proportion of all travelers who will be stopped (red light) is 0.30, because the light has probability 0.30 of showing red on any push of the button. To test this claim, a concerned citizen watches a random sample of 100 travelers push the button. Only 20 get a red light.

a. Assume that the Customs agents' claim is true. Find the probability that the proportion of travelers who get a red light is as small as or smaller than the result in this sample. Show your work.

State:

We want to find the probability that 20% or fewer travelers get a red light.

$\mu_{\hat{p}} = p = 0.30$

$\sigma_{\hat{p}} = \sqrt{\frac{0.30(0.70)}{100}} = 0.0458$ because 100 travelers is less than 10% of ALL travelers (10% condition)

red lights

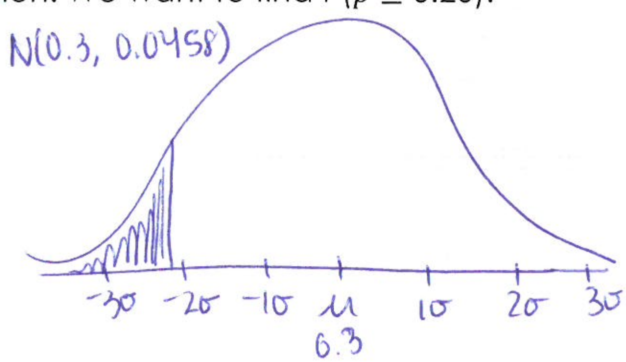
Plan:

10% Condition: satisfied above

Large Counts: $np = (0.30)(100) = \_\_\_\_ \geq 10$ and $n(1-p) = (0.70)(100) = \_\_\_\_ \geq 10$

Because this condition is satisfied, the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution. We want to find $P(\hat{p} \leq 0.20)$.

Do:



$z = \frac{x-\mu}{\sigma} = \frac{0.2 - 0.3}{0.0458} = -2.18$

So, $P(Z \leq -2.18) = 0.0145$

Conclude:

There is a 1.45 % probability that 20% or fewer of the travelers get a red light.

b. Based on your results in (a), do you believe the Customs agents' claim? Explain.

Because the probability is so small (<5%), there is convincing evidence against the custom agents' claim. It is unlikely to get a sample proportion of travelers with a red light this small by chance alone.

Practice Problems:

1. Suppose a large candy machine has 45% orange candies. Use figures 7.11 and 7.12 to help answer the following questions.

Figure 7.11
$\pi = 0.45$
sample size = 25
# of samples = 400

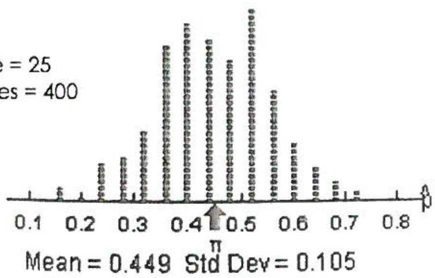
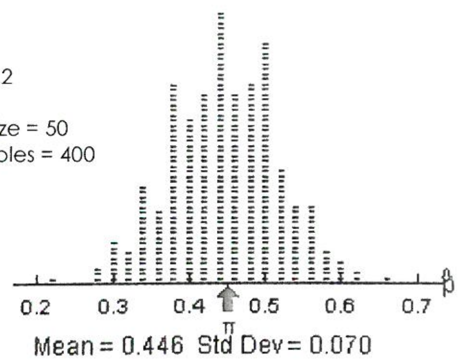


Figure 7.12
$\pi = 0.45$
sample size = 50
# of samples = 400



a. Would you be surprised if a sample of 25 candies from the machine contained 8 orange candies (that's 32% orange)? How about 5 orange candies (20% orange)? Explain.

We would not be surprised to find 8 orange candies because this happened fairly often in the simulation. 5 orange candies would be more surprising because it rarely occurred.

b. Which is more surprising: getting a sample of 25 candies in which 32% are orange or getting a sample of 50 candies in which 32% are orange? Explain.

A sample of 50 because we expect our values to be closer to 0.45 in larger samples, which have smaller standard deviations.

2. Suppose a large candy machine has 15% orange candies. Use figure 7.13 to help answer the following questions.

Figure 7.13a
$\pi = 0.15$
sample size = 25
# of samples = 400

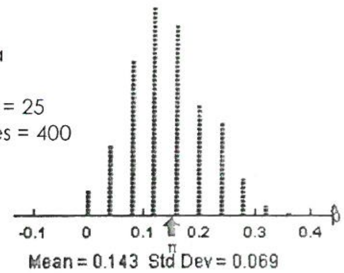
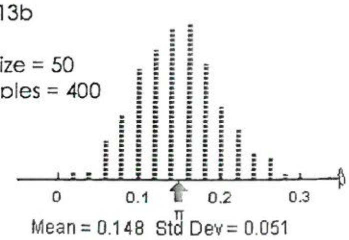


Figure 7.13b
$\pi = 0.15$
sample size = 50
# of samples = 400



a. Would you be surprised if a sample of 25 candies from the machine contained 8 orange candies (that's 32% orange)? How about 5 orange candies (that's 20% orange)? Explain. we would see 8 orange candies show up much less than 5 orange candies because $p = 0.15$, which is closer to 0.2 than 0.32.

b. Which is more surprising: getting a sample of 25 candies in which 32% are orange or getting a sample of 50 candies in which 32% are orange? Explain.

A sample of 50 candies would be more surprising because we expect our values to be closer to 0.15 with the larger sample size (due to a smaller standard deviation).

3. Suppose a large candy machine has 45% orange candies. Imagine taking an SRS of 25 candies from the machine and observing the sample proportion $\hat{p}$ of orange candies.

a. What is the mean of the sampling distribution of $\hat{p}$? Why?

$\mu_{\hat{p}} = p = 0.45$ because $\hat{p}$ is an unbiased estimator of p.

b. Find the standard deviation of the sampling distribution of $\hat{p}$. Check to see if the 10% condition is met.

✓ 10% condition: $n = 25$

$25 \cdot 10 = 250 \leq$ all candies inside the machine

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.45)(0.55)}{25}} = 0.0995$$

c. Is the sampling distribution of $\hat{p}$ approximately Normal? Check to see if the Large Counts condition is met. Yes! Because ...

✓ Large counts: $np = 25 \cdot 0.45 = 11.25 \geq 10$ ✓
$nq = 25 \cdot 0.55 = 13.75 \geq 10$ ✓

d. If the sample size were 100 rather than 25, how would this change the sampling distribution of $\hat{p}$? The standard deviation would decrease to $\sigma_{\hat{p}} = \sqrt{\frac{(0.45)(0.55)}{100}} = 0.0497$
(be cut in half)

4. Suppose a large candy machine has 15% orange candies. Imagine taking an SRS of 25 candies from the machine and observing the sample proportion $\hat{p}$ of orange candies.

a. What is the mean of the sampling distribution of $\hat{p}$? Why?

$\mu_{\hat{p}} = p = 0.15$ because $\hat{p}$ is an unbiased estimator of p.

b. Find the standard deviation of the sampling distribution of $\hat{p}$. Check to see if the 10% condition is met.

✓ 10% condition: $25 \cdot 100 = 2500 \leq$ all candies in the machine

$\sigma_{\hat{p}} = \sqrt{\frac{(0.15)(0.85)}{25}} = 0.0714$

c. Is the sampling distribution of $\hat{p}$ approximately Normal? Check to see if the Large Counts condition is met. NO! Because ...

X Large counts: $np = 25 \cdot 0.15 = 3.75 \leq 10$ X
$nq = 25 \cdot 0.85 = 21.25 \geq 10$ ✓

d. If the sample size were 225 rather than 25, how would this change the sampling distribution of $\hat{p}$? The sampling distribution would become approximately Normal and the standard deviation would decrease (1/3 of original).

✓ Large Counts: $np = 22.5 \geq 10$ ✓
$nq = 191.25 \geq 10$ ✓

$\sigma_{\hat{p}} = \sqrt{\frac{(0.15)(0.85)}{225}} = 0.0238$