Sample Proportions

Section 9.2

Are You Smarter Than a 5th Grader?

What proportion of US teens know the year that Columbus sailed to America?

A Gallup Poll found that 210 out of a random sample of 501 American teens knew that it was 1492. So what is the sample proportion, $\hat{p}$?

$\hat{p} = \frac{210}{501} \approx .419$

This proportion is the statistic that we use to gain information about the unknown population parameter, *p*.

Sampling Distribution of a Sample Proportion

Choose an SRS of size n from a large population with population proportion 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 of $\hat{p}$ is $\mu_{\hat{p}} = p$
- The **standard deviation** of the sampling distribution of $\hat{p}$ is

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

on formula's sheet

Think Back to the Bullseyes…

Because the mean of the sampling distribution of $\hat{p}$ is always equal to the population parameter p, we can say that the sample proportion $\hat{p}$ is an unbiased estimator of p.

- What happens to the standard deviation as the sample size *n* increases? Why? it's in the denominator

it decreases ; denominator

- That means that $\hat{p}$ is less variable in larger samples.

- Also, since *n* is under the square root sign, what would we have to do to the sample size in order to cut the standard deviation in half?

multiply by 4 (quadruple it)

Why Not Just Ask?

The formula for the standard deviaon of the sampling distribuon of $\hat{p}$ doesn't apply when the sample is a large part of the populaon, say asking 50 people out of 100. In pracce, we usually only take a sample when the populaon is very large. If the populaon is small, why not just ask everyone in the populaon?

So When IS the Population Large?

Rule of Thumb 1 → Check for Independence

Use the recipe for the standard deviation of $\hat{p}$ only when the population is at least 10 times as large as the sample; that is, when $N \geq 10n$

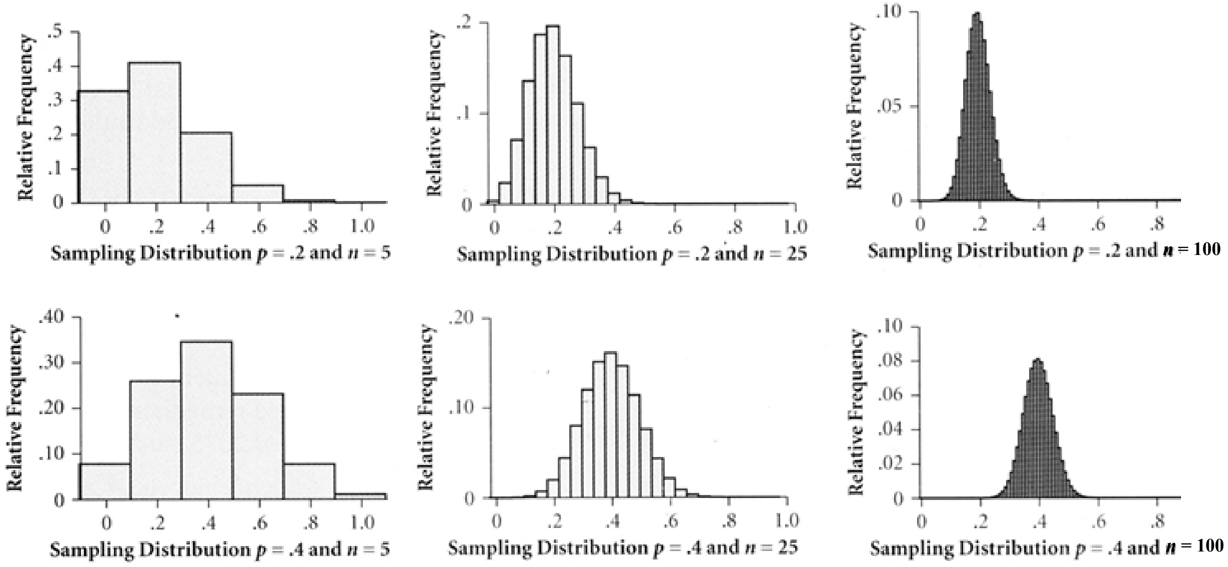
N = population size

n = sample size

- In Section 9.1 we saw that the sampling distribution of $\hat{p}$ is approximately Normal and the accuracy of the Normal approximation improves as the sample size increases.

- Also, for a fixed sample size, *n*, the Normal approximation is most accurate when *p* is close to 0.5 and least accurate when *p* is close to 0 or 1.

What That Looks Like...



Display 5.32 Sampling distributions of $\hat{p}$ for $p = .2$ and $p = .4$ ($n = 5$, 25, and 100).

So When Can We Use the Normal Approximation?

Rule of Thumb 2 → Check for Normality for props.

We will use the Normal approximation to the sampling distribution of $\hat{p}$ for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$

Did You Cover All Your Bases?

One way of checking the effect of undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known facts about the population.

Applying To College?

A polling organization asks an SRS of 1500 first-year college students whether they applied for admission to any other college. In fact, 35% $= p = .35$ of all first-year students applied to colleges besides the one that they are attending. What is the probability that the random sample will give a result within 2 percentage points of this true value?

$P(.33 < \hat{p} < .37)$

Rule of Thumb 1:

$N \geq 10n$

$N \geq 10(1500)$

$N \geq 15000$ ✓

Rule of Thumb 2:

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

$1500(.35) \geq 10$ and $1500(.65) \geq 10$

$525 \geq 10$ ✓ and $975 \geq 10$ ✓

$$P(.33 < \hat{p} < .37) = P\left(\frac{.33 - .35}{\sqrt{\frac{.35(1-.35)}{1500}}} < z < \frac{.37 - .35}{\sqrt{\frac{.35(.65)}{1500}}}\right)$$

$$P(-1.624 < z < 1.624) \approx .896$$

It Doesn't Matter if You're Black or White

About 11% $=p$ of American adults are African American. A national sample of 1500 American adults gives a sample proportion of only 9.2% $=\hat{p}$ African Americans. Should we suspect that the sampling procedure is biased?

$$P(\hat{p} \le .092) =$$

$$P\left(z \le \frac{.092 - .11}{\sqrt{\frac{.11(1-.11)}{1500}}}\right)$$

$$P(z \le -2.228)$$

$$\approx .013$$

B/c the prob. = .013 is < 0.05, I think the sampling procedure was biased.

Rule of Thumb 1:

$N \ge 10n$

$N \ge 10(1500)$

$N \ge 15{,}000$ ✓

Rule of Thumb 2:

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

$1500(.11) \ge 10$ and $1500(.89) \ge 10$

$165 \ge 10$ and $1335 \ge 10$

✓ ✓

Assignment: pg. 588 #19-24 all & bring 50 pennies by Wed/Thurs.