Inferences about Population Proportions

Our discussion of statistical inference to this point has concerned making inferences about population means. But we often want to answer questions about the proportion of some outcome in the population. Here are some examples.

- What proportion of U.S. adults are unemployed right now?
- What proportion of teenagers has a computer with Internet access in their bedroom?
- What proportion of college students pray on a daily basis?
- What proportion of preteens has a cell phone?
- What proportion of Californians approve of President Bush's handling of the situation in Iraq?

Recall the conditions for the sampling distribution of $\hat{p}$.

1. $M_{\text{p-hat}} = p$

2. Spread: $\sigma_{\text{p-hat}} = \sqrt{p(1-p)/n}$ if $N \geq 10n$

3. Shape: p-hat is approximately Normal if $np \geq 10$ and $n(1 - p) \geq 10$

Inferences about population proportions are based on this sampling distribution.

> Conditions for Inference about a Proportion:
>
> 1. SRS
> 2. Normality $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$
> 3. Independence

Example 20: a) Describe the population of interest and explain in words what the parameter p is. (b) Give the numerical value of the statistic $\hat{p}$ that estimates p. (c) Determine whether each of the conditions is met for calculating a confidence interval for the population proportion p.

Tonya wants to estimate what proportion of the students in her dormitory like the dorm food. She interviews an SRS of 50 of the 175 students living in the dormitory. She finds that 14 think the dorm food is good.

a) Students in Tonya's dorm

p is % of students who think dorm food is good

b) $\hat{p} = .28$

c) SRS is stated

Normality yes

$.28(50) = 14$

$.72(50) = 36$

Ind: NO population needs to be at least 500

Confidence Interval for a Population Proportion:

CI for p is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

Example 21: Alcohol abuse has been described by college presidents as the number one problem on campus, and it is an important cause of death in young adults. How common is it? A 2001 survey of 10,904 U.S. college students collected information on drinking behavior and alcohol-related problems. The researchers defined "frequent binge drinking" as having five or more drinks in a row three or more times in the past two weeks. According to this definition, 2486 students were classified as frequent binge drinkers. That's 22.8% of the sample. Based on these data, what can we say about the proportion of all college students who have engaged in frequent binge drinking? Construct and interpret a 99% confidence interval for the population proportion. $\hat{p} = .228$ $\sigma_{\hat{p}} = \sqrt{\frac{.228(.772)}{10904}} \approx .0040$

Population: College students in US

Parameter: The true % who participate in frequent binge drinking

SRS - Assume + proceed with caution

Ind: Assume more than 109,040 college students

Normality: $.228(10904) = 2486.112$ so $\hat{p}$ is approx Normal

$.772(10904) = 8417.888$

$.228 \pm 2.576\left(\sqrt{\frac{.228(.772)}{10904}}\right)$

We are 99% confident the true % of college students is

$.228 \pm .0103$ (.2177, .2383) 21.77 to 23.83%

Remember that the margin of error in this confidence interval includes only random sampling error! There are other sources of error that are not taken into account. What are they?

Nonresponse

Voluntary response

Undercoverage

Example 22: The National Survey of Student Engagement found that 87% of students report that their peers at least "sometimes" copy information from the Internet in their reports without citing the source.[16] Assume that the sample size is 430,000.

(a) Find the margin of error for 95% confidence.

$1.96\sqrt{\frac{.87(.13)}{430000}}$

MOE = .001

(b) Here are some items from the report that summarizes the survey. More than 430,000 students from 730 four-year colleges and universities participated. The average response rate was 43% and ranged from 15% to 89%. Institutions pay a participation fee of between $3000 and $7500 based on the size of their undergraduate enrollment. Are these issues part of the margin of error you found in (a)? What impact might these issues have on the survey results?

None are part of MoE. MoE only accounts for sampling variability!

Nonresponse rates are huge in some cases

Fees could cause some universities to Not Participate

Choosing the sample size

In planning a study, we may want to choose a sample size that will allow us to estimate the parameter within a given margin of error. We saw earlier how to do this for a population mean. The method is similar for estimating a population proportion.

The margin of error in the approximate confidence interval for p is

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Here z^* is the critical value for the level of confidence we want. Because the margin of error involves the sample proportion of successes $\hat{p}$, we need to guess this value when choosing n. Call our guess p^*. Here are two ways to get p^*:

1. Use a guess p^* based on a pilot study or on past experience with similar studies. You should do several calculations that cover the range of $\hat{p}$-values you might get.

2. Use $p^* = 0.5$ as the guess. The margin of error m is largest when $\hat{p} = 0.5$, so this guess is conservative in the sense that if we get any other $\hat{p}$ when we do our study, we will get a margin of error smaller than planned.

Once you have a guess p^*, the formula for the margin of error can be solved to give the sample size n needed. Here is the result.

Sample Size for desired Margin of Error:

$$Z^* \sqrt{\frac{p^*(1-p^*)}{n}} \leq \text{MoE}$$

p^* is guess for $\hat{p}$

Which method for finding the guess p^* should you use? The n you get doesn't change much when you change p^* as long as p^* is not too far from 0.5. So use the conservative guess $p^* = 0.5$ if you expect the true p to be roughly between 0.3 and 0.7. If the true $\hat{p}$ is close to 0 or 1, using $p^* = 0.5$ as your guess will give a sample much larger than you need. So try to use a better guess from a pilot study when you suspect that $\hat{p}$ will be less than 0.3 or greater than 0.7.

Example 23: When trying to hire managers and executives, companies sometimes verify the academic credentials described by the applicants. One company that performs these checks summarized their findings for a six-month period. Of the 84 applicants whose credentials were checked, 15 lied about having a degree.

(a) Find the proportion of applicants who lied about having a degree, and find the standard error.

$\hat{p} = 15/84 = .179$

$$SE_{\hat{p}} = \sqrt{\frac{.179(.821)}{84}}$$

$= .0418$

(b) Consider these data to be a random sample of credentials from a large collection of similar applicants. Calculate and interpret a 90% confidence interval for the true proportion of applicants who lie about having a degree.

$.179 \pm .0688$ Pop are all applicants, parameter % who lie

$(.1102, .2478)$ SRS is Assumed, Normality) $.179(84) = 15$ $.821(84) = 69$ So $\hat{p}$ is N.

IND ?? ≥ 840 applicants

$.179 \pm 1.645(.0418)$

Example 24: A national opinion poll found that 44% of all American adults agree that parents should be given vouchers good for education at any public or private school of their choice. The result was based on a small sample. How large an SRS is required to obtain a margin of error of 0.03 (that is, ±3%) in a 95% confidence interval?

(a) Answer this question using the previous poll's result as the guessed value p^*.

$$1.96\sqrt{\frac{.44(.56)}{n}} = .03$$

$$.9729 = .03\sqrt{n}$$

$$32.43 = \sqrt{n}$$

$$1051.74 = n$$

1052 adults

(b) Do the problem again using the conservative guess $p^* = 0.5$. By how much do the two sample sizes differ?

$$1.96\pm\sqrt{\frac{.5(.5)}{n}} = .03$$

$$.98 = .03\sqrt{n}$$

$$32.67 = \sqrt{n}$$

$$1067.11 = n$$

1068

differ by 16 adults!

Calculators and CI for Proportions

The TI-83/84/89 can be used to construct a confidence interval for an unknown population proportion.

Of $n = 439$ teens surveyed, $X = 246$ said they thought young people should wait to have sex until after marriage.

To construct a confidence interval:

Press STAT, then choose TESTS and A:1-PropZInt

When the 1-PropZInt screen appears, enter $x = 246$, $n = 439$, and confidence level 0.95.

```
1-PropZInt
 x:246
 n:439
 C-Level:.95
 Calculate
```

Highlight "Calculate" and press ENTER. The 95% confidence interval for p is reported, along with the sample proportion $\hat{p}$ and the sample size, as shown here.

```
1-PropZInt
 (.51393,.60679)
 p̂=.5603644647
 n=439
```