Review of Large Sample Result for Sample Proportion

Recall that for large sample (i.e., sample size \( n \) is large, say \( np > 5 \) and \( n(1 - p) > 5 \)), the sample proportion \( \hat{p} \) of successes has an approximate normal distribution:

\[
\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).
\]

Hence, the sample proportion \( \hat{p} \) estimates the population proportion \( p \) with a standard error estimated by

\[
SE = SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.
\]
Estimating Population Proportion Using Intervals

Since, for large sample size $n$,

$$\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

The estimate $\hat{p}$ tend to miss the expected value $p$ by 1 $SD$ which is estimated by 1 $SE$. Hence, approximately 95% of time $\hat{p}$ will fall within 1.96 $SE$ about $p$:

$$|\hat{p} - p| \leq 1.96SE$$

or that (mathematically)

$p$ is inside the interval $\hat{p} \pm 1.96SE$ 95% of the time

That is, with 95% certainty, the interval $\hat{p} \pm 1.96SE$ contains the true value $p$.

95% Confidence Interval for Population Proportion

95% Confidence Interval for $p$: A 95% confidence interval estimate for the population proportion $p$ is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

1.96 $SE$ $= 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is called 95% margin of error $= ME$
Business Graduates Example

If 6 out of 40 students plan to go to graduate school, the proportion \( p \) of all students who plan to go to graduate school is estimated as \( \hat{p} = \frac{6}{40} = .15 \) with a

\[ (95\%) \text{ margin of error } ME = 1.96 \sqrt{\frac{.15 \times .85}{40}} \approx 0.11 \]

and hence a 95% confidence interval for \( p \) is

\[ 0.15 \pm 0.11 = (0.15 - 0.11, 0.15 + 0.11) = (0.04, 0.26) \]

If 54 out of 360 students plan to go to graduate school, then a 95% confidence interval for \( p \) is (note: \( \hat{p} = 54/360 = .15 \))

\[ 0.15 \pm 1.96 \sqrt{\frac{.15 \times .85}{360}} = 0.15 \pm 0.037 = (0.113, 0.187) \]

Calculation of a 95% confidence interval for the true proportion

Given a sample of size \( n \), the number of successes \( X \) is observed. A 95% confidence interval for the true proportion \( p \) is constructed as follows:

1. Calculate the (point estimate) \( \hat{p} = \frac{X}{n} \).
2. Calculate the standard error

\[ SE = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}. \]

3. Calculate the margin of error \( ME = 1.96 \times SE \).
4. Construct the confidence interval

\[ (\text{point estimate} - ME, \text{point estimate} + ME) = (\hat{p} - ME, \hat{p} + ME). \]

5. If a 95% confidence interval for the true percentage is requested, convert the interval above to percentages by shifting the numbers 2 decimal places to the right.
Interpretation of a 95% Confidence Interval

When a sample becomes available (i.e., has been observed), the confidence interval is completely specified. Cautions should be exercised concerning the interpretation of its result. We say that, with 95% confidence, the true proportion $p$ is between $\hat{p} - ME$ and $\hat{p} + ME$ since the interval either contains or misses entirely the true proportion.

For the *Business Graduates Example*, the 95% confidence interval has been obtained: $(0.113, 0.187)$. We are 95% confident that the *true proportion* of business students planning to attend graduate school is between 0.113 and 0.187. Or that we are 95% confident that the *true percentage* of business students planning to attend graduate school is between 11.3% and 18.7%.

Interpretation of a 95% Confidence Interval

continued

Suppose that a 95% confidence interval for the true population proportion is $(0.23, 0.47)$ then the correct interpretation is

*With 95% confidence, the true population proportion is between 0.23 and 0.47.*

The following gives an example list of incorrect interpretations:

- There is 95% *chance* that true population proportion is between 0.23 and 0.47.
- We are 95% confident that the true *sample* proportion is between 0.23 and 0.47.
Using Confidence Interval

Suppose that it is desired to check if a conjectured value $p_0$, say, for the population proportion is plausible. A sample is taken. If the 95% confidence interval for the true population proportion contains $p_0$, then it is judged plausible. Otherwise, it is judged implausible.

For example it is conjectured, before the survey, that the true proportion of business students planning to attend graduate school is 0.10. The 95% confidence interval for the true population proportion from the survey yielded (0.113,0.187). The conjectured value is implausible.

iClicker Question 12.1

A survey of $n = 1500$ American adults was conducted to check if they believe in astrology. It was conjectured that the true proportion of American adults believing in astrology is $p_0 = 0.32$. The survey showed that $X = 405$ adults believe in astrology. Consequently, a 95% confidence interval for the true proportion is (0.248,0.292). Is the conjectured value plausible according to the confidence interval?

A. Yes.
B. No.
C. Insufficient information to judge.

Note: $\hat{p} = X/n = 405/1500 = 0.27$,
$SE = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{0.27(1 - 0.27)/1500} = 0.011$,
$ME = 1.96 \times SE = 1.96 \times 0.011 = 0.022$. Hence, a 95% confidence interval for the true proportion is $(0.27 - 0.022, 0.27 + 0.022) = (0.248, 0.292)$. 

iClicker Question 12.2

A survey of $n = 1500$ American adults was conducted to check if they believe in astrology. The survey showed that $X = 405$ adults believe in astrology. Consequently, a 95% confidence interval for the true proportion is $(0.248, 0.292)$. Which of the following is true?

A. We are 95% confident that the sample proportion is between 0.248 and 0.292.
B. The chance that the true proportion is between 0.248 and 0.292 is 95%.
C. We are 95% confident that the true proportion is between 0.248 and 0.292.
D. The chance that the true proportion is between 0.248 and 0.292 is 5%.
E. None of the previous.

Sample Size Determination
to ensure estimation accuracy—Election Poll Example

Consider $p$ the proportion of votes a candidate will get in a presidential election. An election poll is installed and a 99.7% reliability in estimation is to be ensured with a margin of error (ME) for the estimation of proportion at 0.02. What is the minimum sample size $n$?
Recall that \( \hat{p} \approx N \left( p, \sqrt{\frac{p(1-p)}{n}} \right) \).

According to the empirical rule, with 99.7% chance the estimate \( \hat{p} \) will be

within \( 3 \sqrt{\frac{p(1-p)}{n}} \) about the mean \( p \).

That is, to get a margin of error at \( ME=0.02 \) with the said reliability, the sample size is to be set so that

\[
3 \sqrt{\frac{p(1-p)}{n}} = ME \quad \text{or} \quad n = \frac{3^2 p(1-p)}{ME^2}.
\]

Two ways to resolve the paradox in coming up with a sample size for the estimation of the unknown population proportion \( p \) using

\[
n = \frac{3^2 p(1-p)}{ME^2} \quad \text{is to replace} \ p \ \text{by}:
\]

- a preliminary estimate \( p^* \)
- the conservative estimate \( \frac{1}{2} \)

If the second way is used, then to get a margin of error of
ME=0.02 with 99.7% reliability, the sample size should be set at

\[
n = \frac{3^2 \times \frac{1}{2} \times \frac{1}{2}}{ME^2} = \frac{9}{4ME^2} = \frac{9}{4 \times (0.02)^2} = 5625.
\]