Chapter 19: Confidence Intervals for Proportions

One of the two principal results of Chapter 18 was that
\[ \hat{p} \sim N(p, \sqrt{pq/n}), \]  
provided that the randomization, success/failure, and 10% conditions were met. This chapter develops the standard method of addressing the accuracy of \( \hat{p} \) as an estimator of \( p \). The method is based on the normal approximation (formula 1).

To address accuracy, interval estimate for \( p \) is computed in addition to the point estimate of \( p \). The interval estimate is a set of values for \( p \) that agree with the sample.

- An example below reports the point estimate \( \hat{p} = .709 \) and the interval estimate \([.659, .759] \). Any value of \( p \) between the lower bound (.659) and the upper bound (.759) is a reasonable value for \( p \) in the sense that it may have led to the observed data. It’s highly improbable to have observed the sample if \( p \) is any value outside of the interval.

The confidence interval method can be motivated by the concept of repeatedly sampling the population.

- The empirical rule states that if \( Y \sim N(\mu, \sigma) \), then 95% of all realizations of \( Y \) will be between \( \mu - 2\sigma \) and \( \mu + 2\sigma \).

- If \( n \) is sufficiently large, then an analogous statement applies to \( \hat{p} \): approximately 95% of all random samples of size \( n \) will produce \( \hat{p} \) between \( p - 2\sigma(\hat{p}) \) and \( p + 2\sigma(\hat{p}) \).

- More concisely,
\[ .95 = P[p - 2\sigma(\hat{p}) \leq \hat{p} \leq p + 2\sigma(\hat{p})]. \]  
This expression is not exactly what is wanted. The objective is to find an interval that brackets \( p \) (not \( \hat{p} \)) so that there is a .95 probability that \( p \) will be bracketed by (or contained in) the interval. In other words, the objective is compute lower and upper bounds from \( \hat{p} \) (\( l \) and \( u \)) so that
\[ .95 = P(l \leq p \leq u). \]

- The desired interval can be derived from formula (2) as follows:
\[ .95 = P[p - 2\sigma(\hat{p}) \leq \hat{p} \leq p + 2\sigma(\hat{p})] \]
\[ = P[-p + 2\sigma(\hat{p}) \geq -\hat{p} \geq p - 2\sigma(\hat{p})] \]
\[ = P[2\sigma(\hat{p}) \geq p - \hat{p} \geq -2\sigma(\hat{p})] \]
\[ = P[\hat{p} + 2\sigma(\hat{p}) \geq p \geq \hat{p} - 2\sigma(\hat{p})] \]
\[ = P[\hat{p} - 2\sigma(\hat{p}) \leq p \leq \hat{p} + 2\sigma(\hat{p})]. \]
The lower bound of the interval is \( l = \hat{p} - 2\sigma(\hat{p}) \) and the upper bound is \( u = \hat{p} + 2\sigma(\hat{p}) \).

To construct an interval, it’s necessary to estimate \( \sigma(\hat{p}) \) by plugging \( \hat{p} \) into the formula \( \sigma(\hat{p}) = \sqrt{pq/n} \). The estimate is called the standard error of \( \hat{p} \), and it is

\[
\hat{\sigma}(\hat{p}) = \sqrt{\hat{p}\hat{q}/n}
\]

where \( \hat{q} = 1 - \hat{p} \). De Veaux et al. sometimes use the notation \( \hat{\sigma}(\hat{p}) = \text{SE}(\hat{p}) \).

**Example:** In a 1997 study of traffic stops by Philadelphia police officers\(^1\), of 262 traffic stops, 207 of the drivers were African Americans. At that time, 42.2% of the Philadelphia population was African American. These data can be used to determine whether there is statistical evidence of racial profiling by Philadelphia police officers.

Let \( p \) denote the proportion of all traffic stops involving African American drivers. If there were no profiling, then \( p \) ought to be .422. However, the sample estimate of \( p \) is

\[
\hat{p} = \frac{207}{262} = .790.
\]

While \( \hat{p} = .790 \) strongly suggests that \( p \) is larger than .422, it must be recognized that the data are a sample of traffic stops, and that sampling variability is present. To resolve uncertainty attributable to sampling variability, a confidence interval is computed that brackets all values of \( p \) that are consistent with the sample.

If the interval contains .422, then the data do not support the contention that Philadelphia police officers were profiling African Americans. If .422 is not contained in the interval, then data support the contention that Philadelphia police officers were profiling African Americans.

- The standard error of \( \hat{p} \) is

\[
\hat{\sigma}(\hat{p}) = \sqrt{\hat{p}\hat{q}/n} = \sqrt{.790 \times .210/262} = .0251.
\]

- The lower bound of the interval is

\[
l = \hat{p} - 2\hat{\sigma}(\hat{p}) = .709 - 2 \times .0251 = .659.
\]

and the upper bound is \( u = \hat{p} + 2\hat{\sigma}(\hat{p}) = .759 \).

\(^1\)http://aclu/pubs
The confidence interval is [.659, .759]. It is very unlikely to have obtained a sample with 207 stops of African Americans out of 262 total stops if \( p = .422 \). It is concluded that the data strongly support the contention that Philadelpia police officers were profiling African Americans.

It remains for the researchers to argue that the data are a random sample of all traffic stops, or at least representative of all traffic stops.

This example illustrates statistical inference. From a sample of 262 traffic stops, a conclusion has been drawn about all traffic stops (that’s an inference), and it has been supported by a statistical statement (the confidence interval). The confidence interval objectively identifies all values of \( p \) that agree (or are consistent with) with the sample.

Some terminology and details remain to be clarified.

- The *margin of error* for a particular confidence interval is the quantity that is added and subtracted from the estimate.
- In this example, the margin of error is \( 2\hat{\sigma}(\hat{p}) = 2 \times .0251 = .0502 \).
- Opinion polls usually report compute the margin of error as \( 2\hat{\sigma}(\hat{p}) \). The confidence level associated with this margin of error is 95%.
- The *confidence level* is the percentage of intervals that bracket the parameter. That is, if the method (collect the sample and compute the interval) is used many times, then the confidence level is, approximately, the percentage of intervals containing \( p \).
- Different confidence levels are sometimes used besides 95%. A greater confidence level implies that the probability of capturing the parameter is greater. The interval is wider, and so is somewhat less informative. This topic will be expanded on below.
- Commonly used confidence levels are 90%, 95% and 99%. An interval based on any chosen level between 0 and 100 is created from the equation

\[
1 - 2\alpha = P[\hat{p} - z^*\hat{\sigma}(\hat{p}) \leq p \leq \hat{p} + z^*\hat{\sigma}(\hat{p})]
\]

where \( 100(1 - 2\alpha)\% \) is the *confidence level* and \( -z^* \) is the 100\( \alpha \) percentile of the standard normal distribution.

- Start by choosing a confidence level, say 90% and identify \( \alpha \). Set 90 = 100(1 - 2\( \alpha \)). Then, \( \alpha = \frac{1}{2} \left(1 - \frac{90}{100}\right) = .05 \).

\(^2\)Since .422 is not in the interval.
• Next, the critical value \(-z^*\) is determined. \(-z^*\) is the 100\(\alpha\) percentile of the standard normal distribution and \(z^*\) is the 100\((1 - \alpha)\) percentile of the standard normal distribution. The area between \(-z^*\) and \(z^*\) is the confidence level expressed in decimal form.

• For example, for a 95\% confidence level, look for the 2.5 (or 97.5) percentile in a normal table. The 2.5 and 97.5 percentiles are \(-1.96\) and 1.96.

• Critical values for the commonly used confidence levels are given in the table to the right.

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Critical value (z^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>1.96</td>
</tr>
<tr>
<td>99%</td>
<td>2.576</td>
</tr>
</tbody>
</table>

• The general form for a confidence interval for a population proportion \(p\) based on a random sample of size \(n\) is

\[
\hat{p} \pm z^*\hat{\sigma}(\hat{p}) = \hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}},
\]

*Example:* Data on extramarital and premarital sex are discussed in Thorne and Collard.\(^3\) They report that of 360 married men, 60 had engaged in extramarital sex and of 615 married women, 61 had engaged in extramarital sex. Assuming these data to be a random sample of married men and women, 90\% confidence intervals for the population proportions engaging in extramarital sex are obtained from the statistics

\[
\begin{align*}
\hat{p}_m & = \frac{60}{360} = .166 \text{ and } \hat{\sigma}(\hat{p}_m) = .0197, \\
\hat{p}_w & = \frac{61}{615} = .099 \text{ and } \hat{\sigma}(\hat{p}_w) = .0120.
\end{align*}
\]

90\% confidence intervals for the population proportions \(p_m\) and \(p_w\) are

\[
\begin{align*}
\hat{p}_m \pm z^*\hat{\sigma}(\hat{p}_m) & = .166 \pm 1.645 \times .0197 = [.134, .198] \\
\text{and } \hat{p}_w \pm z^*\hat{\sigma}(\hat{p}_w) & = .099 \pm 1.645 \times .0120 = [.079, .119].
\end{align*}
\]

*Reporting:* The estimated proportion of males engaged in extramarital sex is \(\hat{p} = .166\). I am 90\% confident that the proportion of all males engaged in extramarital sex is between .134 and .198.

As the male and female intervals contain no common values, it can be concluded with confidence that males are more likely than females to engage in extramarital sex. Said another way, there are statistically significant differences between the proportion of males and females engaging in extramarital sex.\(^4\)


\(^4\)This comparison is informal and confidence level cannot be set. A formal approach will be developed in Chapter 22.
Necessary conditions: confidence intervals are based upon the normal approximation of the sampling distribution of \( \hat{p} \). The accuracy of an interval depends on three conditions being met. The conditions are familiar:

1. **Randomization condition**: the data are a random sample (or at least representative) of the population.
2. **10% condition**: \( n \) is less than 10% of the population size.
3. **Success/failure condition**: \( np > 10 \) and \( nq > 10 \).

The meaning of confidence level, and why not say the confidence interval has a 100\((1 - 2\alpha)\) probability of having capturing \( p \)?

- The method of constructing the confidence interval successfully captures the parameter with probability \( 1 - 2\alpha \). The method is: collect a random sample of size \( n \) and compute the upper and lower bounds of the interval.
- Once the interval is computed, it either does or does not contain \( p \). The probability that \( p \) is in the interval is either 0 or 1, but it’s not known which value is correct for the computed interval.
- An analogy: I’ve learned how to walk into a completely dark room and throw a hoop onto a peg. (The theory of hoop tossing has provided the necessary knowledge). Next, I demonstrate the technique to an audience by walking into a dark room containing the peg and the audience, throw the hoop, and walk out (still in the dark). I don’t know if I was successful, but I’m confident that I succeeded because I know my success rate (which describes my method, not the toss) is .95.

- Example: One hundred 90% confidence intervals are shown below. These were constructed by a computer simulation of randomly sampling populations of men and women. The male population consisted of 16,667 ones and 83,333 zeros. The female population was composed of 9,900 ones and 90,100 zeros.

The figure below shows the results. The vertical lines represent the confidence intervals. A blue line represents an interval that correctly brackets \( p_m \); a red line represents an interval that correctly brackets \( p_w \). Black and green lines represent intervals that incorrectly bracket the parameter. In brief, 89% percent of male confidence intervals captured \( p_m \), 89% of the female intervals captured \( p_w \), and 25% of the male and female intervals overlap.
Misinterpretations of confidence intervals

1. In repeated sampling, the interval $[.079, .119]$ contains the true proportion of women that have engaged in extramarital affairs 90% of the time. \textit{Not true}: $p$ does not vary and it’s either between .079 and .119 or not. Furthermore, the interval will change from sample to sample in repeated sampling.

2. There is a 90% chance that $p$ is between .079 and .119. \textit{Not true}: $p$ does not vary (it’s not random), so there’s no meaningful probability related to the question of whether $p$ is between .079 and .119.

The connection between confidence interval width and the confidence level

- In general, the confidence level is $100(1 - 2\alpha)\%$ and $1 - 2\alpha$ is the probability that $p$ will be bracketed between the lower and upper bounds. If the confidence level is large (near 100%), then the interval must be wide to insure that $p$ will be bracketed with a probability near 1.

- Let $z^*_{\alpha}$ denote the critical value associated with a $100(1 - 2\alpha)\%$ confidence interval. For example, for a 95% confidence level, set $\alpha = .025 \Rightarrow z^*_{.025} = 1.96$.

Using this notation, a $100(1 - 2\alpha)\%$ confidence interval for $p$ is

$$
\left[ \hat{p} - z^*_{\alpha} \sqrt{\hat{p}\hat{q}/n}, \hat{p} + z^*_{\alpha} \sqrt{\hat{p}\hat{q}/n} \right].
$$
• The width of the interval is an important attribute of an interval. Narrow intervals are more informative than wide intervals. The precision of a sample proportion \( \hat{p} \) as an estimator of \( p \) is directly related to the width of an interval computed from \( \hat{p} \): a precise estimator leads to a narrow interval.

• The width of an interval is the distance between the upper and lower bounds

\[
 w = \hat{p} + z^*_\alpha \sqrt{\hat{p}(1-\hat{p})/n} - \left( \hat{p} - z^*_\alpha \sqrt{\hat{p}(1-\hat{p})/n} \right) \\
= 2z^*_\alpha \sqrt{\hat{p}(1-\hat{p})/n}.
\]

• Two terms in this equation that affect the width are in the control of the researcher: \( n \) and \( \alpha \).

• Increasing \( n \) will reduce the width. Example: doubling \( n \) will make the interval narrower by the factor \( \frac{1}{\sqrt{2}} \); more generally, increasing \( n \) by a factor of \( k \) will make the interval narrower by the factor \( \frac{1}{\sqrt{k}} \).

• If someone is planning a study to estimate \( p \), then it is possible to determine a minimal necessary sample size if the level of precision and confidence levels can be identified.

1. The minimal necessary sample size also depends on \( p \). The worst case outcome in terms of requiring a large sample size occurs when \( p = .5 \). So, it’s best to suppose that \( \hat{p} = .5 \) when computing a minimal sample size.

2. The minimal necessary sample size can be found by solving for \( n \) in the equation

\[
 w = 2z^*_\alpha \sqrt{\hat{p}(1-\hat{p})/n}.
\]

3. First, the equation is simplified using \( \hat{p} = 1/2 = \hat{q} \):

\[
 w = 2z^*_\alpha \sqrt{\hat{p}(1-\hat{p})/n} \\
= 2z^*_\alpha \sqrt{\frac{1}{4n}} \\
= z^*_\alpha \sqrt{\frac{1}{n}}.
\]

4. Solving for \( n \) leads to the formula for the minimal necessary sample size:

\[
 n = \left( \frac{z^*_\alpha}{w} \right)^2.
\]

5. Example: suppose that the desired confidence level is \( 100(1 - 2\alpha)\% = 95\% \) and that the maximum acceptable width for the interval is \( .15 \). Then

\[
 n = \left( \frac{1.96}{.15} \right)^2 \\
= 170.7,
\]

and a minimal sample size is \( n = 171 \). (Always round up).
6. Example: De Veaux et al. state that for opinion polls, the width should not be greater than .05. Assuming the desired confidence level is 95%,

\[ n = \left( \frac{1.96}{.05} \right)^2 = 1536.64. \]

Rounding up, the minimal sample size is \( n = 1537. \)

- The confidence level, or equivalently, \( \alpha \) also affects the width of the interval (but to a much lesser extent than \( n \)). Recall that \( z_{\alpha}^* \) is the \( 1 - \alpha \) percentile of the \( N(0,1) \). Two common confidence levels are

\[
\begin{align*}
90 &= 100(1 - 2\alpha) \Rightarrow \alpha = .05 \Rightarrow z_{.95}^* = 1.645 \\
95 &= 100(1 - 2\alpha) \Rightarrow \alpha = .025 \Rightarrow z_{.975}^* = 1.96.
\end{align*}
\]

Suppose that \( w = .1 \) is acceptable and either 90 or 95% confidence levels are acceptable. The minimal necessary sample sizes are

\[
\begin{align*}
95\% &\Rightarrow n = \left( \frac{1.96}{.1} \right)^2 = 385 \\
\text{and } 90\% &\Rightarrow n = \left( \frac{1.645}{.1} \right)^2 = 271.
\end{align*}
\]