CH 7  Confidence Interval

Part 1  CI for proportion

We will often want to make a conclusion or inference about a population parameter based on a single sample.

One of the most common types of inference is to construct what is called a confidence interval, which is defined as “an interval of values computed from sample data that is almost sure to cover the true population value.”

By Central Limit Theorem, if $n$ is sufficiently large, then $\hat{p}$ has approximately a normal distribution with

$$E(\hat{p}) = p$$
and

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

What's the values of $a$ and $b$ such that

$$P(a \leq \hat{p} \leq b) = 0.95$$

Recall from the Empirical Rule, for the normal distribution 95% of observations will fall within 2 (or 1.96) standard deviations of the mean
Thus, there is a 95% probability that the sample proportions \( \hat{p} \) will lie within \( 1.96 \times \sqrt{p(1-p)/n} \) of \( p \) (the true population proportion).

Therefore, there is (approximately) a 95% probability that \( p \) will lie within \( 1.96 \times \sqrt{\hat{p}(1-\hat{p})/n} \) of \( \hat{p} \).

Hence an approximate 95% confidence interval for \( p \) is given by:
\[
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.
\]

Large Sample Confidence Interval for \( p \)

The standard error of \( \hat{p} \) is \( SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \).

The approximate (1-\( \alpha \))100% Confidence Interval for \( p \) is:
\[
\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.
\]

This formula is valid if (1) the data can be regarded as a random sample from a large population, (2) the observations are independent and (3) \( n \) is large: \( n\hat{p} \geq 10 \) and \( n(1-\hat{p}) \geq 10 \).

Z\( \alpha \) Notation

- We will use \( Z_{\alpha} \) to denote the value on the z-axis for which \( \alpha \) of the area under the z curve lies to the right of \( Z_{\alpha} \).

General Formula for Confidence Intervals

During this course, we will be looking at confidence intervals for different population parameters, e.g., proportions and means.

The general formula are the same for any parameters:

\[
\text{Estimate} \pm \text{Table Value} \times \text{Standard Error}
\]

OR

\[
\text{Estimate} \pm \text{Margin of Error}
\]
Table Values for the CI for $p$

- Researchers are typically interested in 90%, 95% or 99% confidence intervals.
- The table values ($Z_{\alpha/2}$) for these common intervals are:

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>$Z_{\alpha/2}$</td>
<td>1.645</td>
<td>1.96</td>
<td>2.58</td>
</tr>
</tbody>
</table>

Confidence Level and $\alpha$

- The proportion of the time we fail to cover the parameter with a CI is denoted $\alpha$.
- i.e. If we have a 90% CI, then $\alpha$ is 0.10.
- We can write the confidence in terms of $\alpha$: $100(1-\alpha)$%.
- $Z_{\alpha/2}$ is the area of the tail.

Example: Bumpers

Recall that for the Bumpers example, $n = 30$ and $\hat{p} = 16/30 = 0.5333$.

Let's calculate a 95% confidence interval for $p$ (the true population proportion of bumpers with no visible damage).

Step 1: calculate the standard error:

Step 2: look for the table value. For a 95% confidence interval $\alpha$=0.05, we need $Z_{0.025} = 1.96$.

Step 3: get the CI
So, we can be 95% confident that the proportion of all bumpers (from the same manufacture) that will not show visible damage is between 0.354 and 0.712.

Large Sample CI for p using Rcmdr

• Start by creating a dataset that gives the list of “successes” and “failures”.
• The choose Statistics -> Proportions -> Single-Sample Proportion Test.
• Check the confidence level and make sure “Normal Approximation” selected and hit OK.
• Note: We will come back and discuss the other options when we talk about hypothesis testing.

CI for p with summarized data

• In the script window type:
  prop.test(16, 30, correct=FALSE)
  # “successes” = 16  n = # trials = 30  No continuity correction
• The output will be identical to that on the next slide.

Example: Bumpers

<table>
<thead>
<tr>
<th>Damage</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
<td>14</td>
</tr>
</tbody>
</table>

1-sample proportions test without continuity correction

data: rbind(.Table), null probability 0.5
X-squared = 0.1333, df = 1, p-value = 0.715
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.3614230 0.6976761
sample estimates:
  p
0.5333333
Exact CI for p using Rcmdr

For “Type of Test” choose “Exact Binomial”.

Exact binomial test
number of successes = 16, number of trials = 30, p-value = 0.8555
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval: 0.3432552 0.7165819
sample estimates:
probability of success 0.5333333

As confidence level increases, will the width of the confidence interval increase or decrease?

INCREASE

As sample size increases, will the width of the confidence interval increase or decrease?

DECREASE

Finding the Sample Size

“When planning an experiment, it is wise to consider in advance whether the estimates generated from the data will be sufficiently precise. It can be painful indeed to discover after a long and expensive study that the standard errors are so large that the primary questions addressed by the study cannot be answered.”

• The margin of error (and width of the confidence interval) depends on the confidence level, sample size and \( \hat{p} \).
• Margin of Error = \( ME = Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)
• Sample size = \( n = \frac{Z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{ME^2} \)
• Before we conduct the experiment, we don’t know the value of \( \hat{p} \).
• However, we know the margin of error will be maximized when \( \hat{p} = 0.5 \).
Sample Size to Use for Estimating a Population Proportion

For experiment planning purposes, the suggested sample size \( n \) in order to keep the margin of error less than or equal to \( d \) is:

\[
n = \left( \frac{Z_{\alpha/2}}{2d} \right)^2
\]

If the value of \( n \) is not an integer, we always round up to the next integer.

Ex 7.51: Interest Charges

The financial manager of a large department store chain has selected a random sample of 200 of its credit card customers and found that 136 had incurred an interest charge during the previous year because of an unpaid balance.

Compute a 90% CI for the true proportion of credit card customers who incurred an interest charge during the previous year.

If the desired width of the 90% confidence interval is 0.05, what sample size would be required?