Chapter 21

What Is a Confidence Interval?

Review: empirical rule

Recall from previous chapters:

**Parameter**
- fixed, unknown number that describes the population

**Statistic**
- known value calculated from a sample
  - a statistic is used to estimate a parameter

**Sampling Variability**
- different samples from the same population may yield different values of the sample statistic
  - estimates from samples will be closer to the true values in the population if the samples are larger
Recall from previous chapters:

**Sampling Distribution**
tells what values a statistic takes and how often it takes those values in repeated sampling.

- Sample proportions (\( \hat{p} \)'s) from repeated sampling would have a normal distribution with a mean equal to the population mean and a certain standard deviation.
- Sample means (\( \bar{x} \)'s) from repeated sampling have a normal distribution with a mean equal to the population mean and a certain standard deviation.

95% of all sample means will be within roughly 2 standard deviations (\( 2 \sigma / \sqrt{n} \)) of the population parameter \( \mu \).

Because distances are symmetrical, this implies that the population parameter \( \mu \) must be within roughly 2 standard deviations from the sample average \( \bar{x} \), in 95% of all samples.

This reasoning is the essence of statistical inference.

Reworded

With 95% confidence, we can say that \( \mu \) should be within roughly 2 standard deviations (\( 2 \sigma / \sqrt{n} \)) from our sample mean \( \bar{x} \).

- In 95% of all possible samples of this size \( n \), \( \mu \) will indeed fall in our confidence interval.
- In only 5% of samples would \( \bar{x} \) be farther from \( \mu \).
The Rule for Sample Proportions

If numerous simple random samples of size \( n \) are taken from the same population, the sample proportions \( \hat{p} \) from the various samples will have an approximately normal distribution. The \textit{mean} of the sample proportions will be \( p \) (the true population proportion). The \textit{standard deviation} will be:

\[
\sqrt{\frac{p(1-p)}{n}}
\]

Rule Conditions

• For a rule to be valid, it must have
  • Random sample
  • “Large” sample size

Derivation of the Margin of Error

Two formulas for a 95% confidence interval:

• sample proportion ± 1/\( \sqrt{n} \) (from Chapter 4)
• sample proportion ± 2(SD)

Two formulas are equivalent when the proportion used in the formula for SD is \( p = 0.50 \).

Then 2(SD) is simply \( 1/\sqrt{n} \)

\[
\sqrt{\frac{0.5(1-0.5)}{n}} = \frac{1}{\sqrt{n}}
\]
Confidence Intervals from the Media

Formula for a 95% confidence interval:

\[ \text{sample proportion} \pm \text{margin of error} \]

Fingerprints A Question

Assume that the proportion of all men who have left asymmetry is 15%.
Is it unusual to observe a sample of 66 men with a sample proportion (\( \hat{p} \)) of 30% if the true population proportion (\( p \)) is 15%?

Case Study: Fingerprints

Sampling Distribution

\[
\hat{p} = \frac{p(1-p)}{n} = \frac{0.15(1-0.15)}{66} = 0.044 \text{ (s.d.)}
\]
Case Study: Fingerprints

Answer to Question

• Where should about 95% of the sample proportions lie?
  • mean plus or minus two standard deviations
    \[0.15 - 2(0.044) = 0.062\]
    \[0.15 + 2(0.044) = 0.238\]
  • 95% should fall between [0.062, 0.238]

Formula for a 95% Confidence Interval for the Population Proportion (Empirical Rule)

\[\hat{p} \pm 2\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}\]

Inference for Population Means

Sampling Distribution, Confidence Intervals

• making conclusions about population means rather than population proportions
  – includes the rule for the sampling distribution of sample means (\(\bar{X}\))
  – includes confidence intervals for one mean or a difference in two means
The Rule for Sample Means
If numerous simple random samples of size \( n \) are taken from the same population, the sample means (\( \bar{x} \)) from the various samples will have an approximately normal distribution. The mean of the sample means will be \( \mu \) (the population mean). The standard deviation will be:

\[
\sigma / \sqrt{n}
\]

We do not know the value of \( \sigma \)!

Standard Error of the (Sample) Mean

\[
SEM = \text{standard error of the mean} = \frac{s}{\sqrt{n}}
\]

Results
Exercise and Pulse Rates

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>mean</th>
<th>std. dev.</th>
<th>std. err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercisers</td>
<td>29</td>
<td>66</td>
<td>8.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Nonexer.</td>
<td>31</td>
<td>75</td>
<td>9.0</td>
<td>1.6</td>
</tr>
</tbody>
</table>

* Typical deviation of an individual pulse rate (for Exercisers) is \( s = 8.6 \)
* Typical deviation of a mean pulse rate (for Exercisers) is \( \frac{s}{\sqrt{n}} = \frac{8.6}{\sqrt{29}} = 1.6 \)
Confidence Intervals
Exercise and Pulse Rates

- 95% C.I. for the population mean:
  - sample mean ± 2 × (standard error)
    \[ \bar{x} \pm \frac{2}{\sqrt{n}} \]
- Exercisers: \( 66 \pm 2(1.6) = 66 \pm 3.2 = (62.8, 69.2) \)
- Nonexercisers: \( 75 \pm 2(1.6) = 75 \pm 3.2 = (71.8, 78.2) \)
- Do you think the population means are different?
  Yes, because the intervals do not overlap

Confidence Intervals:
difference between means
Exercise and Pulse Rates

- 95% C.I. for the difference in population means (nonexercisers minus exercisers):
  \[ \bar{x}_n - \bar{x}_e \pm 2 \times (\text{SE of the difference}) \]
- Difference in sample means: \( 75 - 66 = 9 \)
- SE of the difference = 2.26 (given)
- 95% confidence interval: (4.4, 13.6)
  - interval does not include zero (⇒ means are different)

Careful Interpretation of a Confidence Interval

- “We are 95% confident that the mean resting pulse rate for the population of all exercisers is between 62.8 and 69.2 bpm.” (We feel that plausible values for the population of exercisers’ mean resting pulse rate are between 62.8 and 69.2.)
- ** This does not mean that 95% of all people who exercise regularly will have resting pulse rates between 62.8 and 69.2 bpm.
- Statistically: 95% of all samples of size 29 from the population of exercisers should yield a sample mean within two standard errors of the population mean; i.e., in repeated samples, 95% of the C.I.s should contain the true population mean.
Understanding Confidence Level

For a confidence level of 95%, we expect that about 95% of all such intervals will actually cover the true population value.

The remaining 5% will not. Confidence is in the procedure over the long run.

Levels of Confidence

- 95% confidence interval:
  sample proportion ± 2(SD)
  In 95% of all samples, the true proportion will fall within plus/minus 2 standard deviations of the sample proportion. (exact multiplier= 1.96)

- 90% confidence interval:
  sample proportion ± 1.645(SD)

- 99% confidence interval:
  sample proportion ± 2.576(SD)

- More confidence => Wider Interval

Factors affecting the width of the confidence interval:

- The sample size, n
  – the larger n is, the smaller the CI

- The standard deviation, s
  – the smaller s is, the smaller the CI

- The confidence level
  – the smaller the level of confidence, the smaller the CI
Which of the following would cause a decrease in the width of a confidence interval?

a) An increase in n  
b) A decrease in the SE  
c) A decrease in the confidence level  
d) All of the above

Summary of the Variety of Information Given in Journals

Can determine CIs for individual means or difference in two means if you have:

- Direct confidence intervals
- Means and standard errors of the means
- Means, standard deviations, and sample sizes

Perceived racial/ethnic harassment and tobacco use among African American young adults.

We examined the association between perceived racial/ethnic harassment and tobacco use in 2129 African American college students in North Carolina. Age-adjusted and multivariate analyses evaluated the effect of harassment on daily and less-than-daily tobacco use. Harassed participants were twice as likely to use tobacco daily (odds ratio = 2.01; 95% confidence interval=1.94, 2.08) compared with those with no reported harassment experiences. Experiences of racial/ethnic harassment may contribute to tobacco use behaviors among some African American young adults.