Section 3.5 Calculate and Interpret Confidence Levels from Two Independent Samples

1. mod3.5

1.1 Calculate and Interpret Confidence Levels from Two Independent Samples

Notes:

Now we are moving on to calculate and interpret confidence intervals from 2 independent samples and a range of different parameters.

Two Samples - Independent

a) Continuous - mean difference
b) Dichotomous - risk difference
c) Dichotomous - risk ratio
d) Dichotomous - odds ratio
1.2 Learning Outcomes

Learning Outcomes

Calculate and interpret confidence intervals for parameters derived from an independent 2-group study design: mean difference, risk difference, risk ratio, and odds ratio

Notes:

1. Here are the module learning objectives.
2. Remember, the course evaluation you get to fill out will ask whether the learning objectives were appropriately covered and presented.
1.3 CI for Two Samples – Independent Continuous Outcome

Common parameter of interest is difference in means between the two groups, $\bar{X}_1$ and $\bar{X}_2$, and denoted for the population as: $\mu_1 - \mu_2$

Since there are 2 independent groups, we also have: $n_1$ and $n_2$ and $s_1$ and $s_2$

If the sample variances are approximately equal, then we can “pool” the standard deviations, $s_1$ and $s_2$. A typical rule of thumb to pool is: $s^2_1 / s^2_2 > 0.5$ and $s^2_1 / s^2_2 < 2.0$

The pooled (common) standard deviation is a weighted average:

$$s_p = \sqrt{\frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2}}$$

Notes:

1. Here are the guidelines for calculating a confidence interval for 2 independent samples and a continuous outcome variable, such as blood pressure.
2. The parameter of interest is the mean difference in the outcome variable between the 2 groups.
3. We will use a pooled (common) standard deviation for the 2 groups.
1.4 CI for Two Samples - Independent Continuous Outcome

Notes:

1. In this example, we want to calculate a 95% confidence interval for the difference in mean systolic blood pressure between men and women.
2. Note the use of a pooled standard deviation denoted $S_p$. 

Parameter: Mean difference in systolic blood pressure between sample of men and a sample of women

$\bar{X}_{men} = 128.2$; $n_1 = 1623$; $s_1 = 17.5$

$\bar{X}_{women} = 126.5$; $n_2 = 1911$; $s_2 = 20.1$

Note: $s_1^2 / s_2^2 = 0.76$, so can use pooled SD ($S_p$)

Confidence Level: 95%  
Z value: 1.96

$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{359.12} = 19.0$

Formula: $(\bar{X}_1 - \bar{X}_2) \pm zS_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
**1.5 CI for Two Samples – Independent Continuous Outcome**

### CI for Two Samples - Independent Continuous Outcome

**Parameter:** Mean difference in systolic blood pressure between a sample of men and women

\[
\bar{x}_{\text{men}} = 128.2; \quad n_1 = 1623; \quad s_1 = 17.5 \\
\bar{x}_{\text{women}} = 126.5; \quad n_2 = 1911; \quad s_2 = 20.1
\]

**Formula:**

\[
(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

\[
(128.2 - 126.5) \pm 1.96(19.0) \sqrt{\frac{1}{1623} + \frac{1}{1911}}
\]

\[
1.7 \pm 1.26 = (0.44, 2.96)
\]

**Notes:**

1. In this example, the mean difference in systolic blood pressure between men and women is 1.7 mmHg, and the 95% confidence interval for the difference is from 0.44 to 2.96.
2. The confidence interval is narrow due to the large samples of men and women, and is symmetric around the mean difference.
1.6 Practice Exercise

Parameter: Mean difference in depression scores between a sample of men and women

\[ \bar{X}_{\text{men}} = 5.77; \quad n_1 = 163; s_1 = 7.674 \]
\[ \bar{X}_{\text{women}} = 6.86; \quad n_2 = 333; \quad s_2 = 8.714 \]

Note: \[ \frac{s_1^2}{s_2^2} = \]

\[ s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} \]

\[ (\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]

\[ (\bar{X}_1 - \bar{X}_2) = \]

Notes:

1. Now it is time to practice.
2. From your handout and the information provided, calculate a 95% confidence interval for a two-sample (independent) continuous outcome.
3. Provide an interpretation for the confidence interval.
1.7 Practice Exercise Answers

Notes:

1. Here are the answers.
2. Note that this example used a $z$-value of 1.96 for a 95% confidence interval.
3. As seen, the 95% confidence interval for the mean difference in depression scores between men and women ranges from -2.66 to 0.49.
1.8 Calculate the 95% CI Using SPSS

Calculate the 95% CI Using SPSS

Parameter: Mean difference in depression scores between a sample of men and women

\[
\bar{X}_{\text{men}} = 5.77; \quad n_1 = 163; \quad s_1 = 7.674 \\
\bar{X}_{\text{women}} = 6.88; \quad n_2 = 333; \quad s_2 = 8.714
\]

From the sample, we estimate a mean difference in depression scores between men and women of 1.09, and we are 95% confident that the true mean difference lies between the interval of -2.66 to 0.49.

Notes:

1. Let’s use the same data and calculate the 95% confidence interval using SPSS.
2. The SPSS syntax to be used is listed on this slide.

1.9 Calculate the 95% CI using SPSS

Calculate the 95% CI Using SPSS
1.10 CI for Two Samples – Independent: Risk Difference

Notes:

1. In this example, we want to calculate a 95% confidence interval for the risk difference, meaning the difference in 2 incidence proportions.

2. As the outcome, we will used the occurrence (incidence) of cardiovascular disease (CVD), and calculate the confidence interval for the risk difference comparing current smokers to non-smokers.
1.11 CI for Two Samples – Independent: Risk Difference

**Example:** Compare the incidence proportion of CHD among smokers (exposed) and non-smokers (not exposed) \(^\wedge\)

- Smokers: \(n_1 = 744\) w/CHD \(x_1 = 81\) \(\hat{p}_1 = 0.1089\)
- Non-smokers: \(n_2 = 3055\) w/CHD \(x_2 = 298\) \(\hat{p}_2 = 0.0975\)

**Confidence Level:** 95%

\[
\hat{p}_1 - \hat{p}_2 \pm z \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}}
\]

\[
(0.1089 - 0.0975) \pm 1.96 \sqrt{\frac{0.1089(1 - 0.1089)}{744} + \frac{0.0975(1 - 0.0975)}{3055}}
\]

\[
= 0.0114 \pm 0.0247 = (-0.0133, 0.0361)
\]

**Notes:**

1. In this example, the risk difference (the difference in incidence proportions) is \(0.1089 - 0.0975 = 0.0114\).
2. The \(z\)-value of 1.96 corresponds to the 95% confidence interval.
3. The 95% confidence interval for the risk difference ranges from -0.0133 to 0.0361.
1.12 Practice Exercise

Example: Compare the incidence proportion of sleep disorder among person on statins (exposed) and not on statins (not exposed)

Confidence Level: 95% Z value: ______

<table>
<thead>
<tr>
<th></th>
<th>Sleep OK</th>
<th>Sleep Dx</th>
<th>Total</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statin user</td>
<td>91</td>
<td>14 (k1)</td>
<td>105</td>
<td>p1 = 14 / 105 = 0.1333</td>
</tr>
<tr>
<td>Non-statin user</td>
<td>369</td>
<td>28 (k2)</td>
<td>397</td>
<td>p2 = 28 / 397 = 0.0705</td>
</tr>
<tr>
<td>Total</td>
<td>460</td>
<td>42</td>
<td>502</td>
<td></td>
</tr>
</tbody>
</table>

\[
\hat{p}_1 - \hat{p}_2 \pm z \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

Notes:

1. Now it is time to practice.
2. From your handout and the information provided, calculate a 95% confidence interval for a two-sample (independent) risk difference in the incidence of sleep disorder by use versus non-use of statins.
3. Provide an interpretation for the confidence interval.
1.13 Practice Exercise Answers

**Example: **Compare the incidence proportion of sleep disorder among person on statins (exposed) and not on statins (not exposed)

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>statin user</th>
<th>non-statin user</th>
<th>Total</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>91</td>
<td>369</td>
<td>460</td>
<td>( \hat{p}_1 = 14 / 105 = 0.1333 )</td>
</tr>
<tr>
<td></td>
<td>14 (k1)</td>
<td>28 (k1)</td>
<td>42</td>
<td>( \hat{p}_2 = 28 / 397 = 0.0705 )</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>397</td>
<td>502</td>
<td></td>
</tr>
</tbody>
</table>

\[
\hat{p}_1 - \hat{p}_2 \pm z \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.1333 \pm 0.0697 = [-0.007, 0.133]
\]

**Notes:**

1. Here are the answers.
2. Note that this example used a \( z \)-value of 1.96 for a 95% confidence interval.
3. As seen, the 95% confidence interval for the risk difference in the incidence of sleep disorder between statin uses and non-users ranges from -0.007 to 0.133.
1.14 Practice Exercise Interpretation

Notes:

1. Here is the interpretation for the 95% confidence interval.
2. Because the range of the 95% confidence interval for the risk difference includes zero, it is plausible that the incidence of sleep disorder could be higher or lower in statin users compared to non-statin users.
1.15 CI for Two Samples - Independent: Risk Ratio

Parameter of interest is the ratio of the incidence proportions for the population, denoted as \( RR = p_1 / p_2 \). 

For a sample, the point estimate for the risk ratio (RR) is denoted as:

\[ \hat{RR} = \frac{\hat{p}_1}{\hat{p}_2} \]

Note that the RR does not follow a normal distribution, but the natural log (ln) of the RR is approximately normally distributed and is used to calculate the confidence interval - this entails 2 steps:

1. Calculate CI for ln(RR)
2. Calculate CI for RR (i.e. transform)

\[
\text{CI for ln(RR): } \ln \hat{RR} \pm z \left( \frac{n_1 - x_1}{x_1} + \frac{n_2 - x_2}{x_2} \right) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
\text{CI for (RR): } \exp(\text{Lower limit}), \exp(\text{Upper limit})
\]

Notes:

1. Here are the guidelines for calculating a confidence interval for 2 independent samples and the risk ratio as the parameter of interest.
2. As seen, the parameter of interest, the risk ratio, is the ratio of the incidence proportion of the 2 groups being compared.
3. Note that calculation of the confidence interval requires an additional step of performing a calculation in natural logarithm scale, and then doing a transformation.
1.16 CI for Two Samples – Independent: Risk Ratio

Notes:

1. In this example, we want to calculate a 95% confidence interval for the risk ratio, meaning the ratio of the 2 incidence proportions.

2. The outcome of interest is the interest of coronary heart disease (CHD), and the groups being compared are smokers versus non-smokers.

3. After calculating the confidence interval in the natural logarithm scale, the values are transformed by exponentiation (e.g. the function “exp” on your calculator or in Excel).

4. In this example, smokers appear to be at higher risk of developing CHD (risk ratio of 1.12), but the 95% confidence interval of 0.89 to 1.41 includes the possibility of either lower or higher risk associated with smoking.
1.17 Practice Exercise

Practice Exercise

\[ RR = \frac{\hat{p}_1}{\hat{p}_2} \]

CI for \( \ln(\text{RR}) \):
\[ \ln\text{RR} \pm z \left( \frac{(n_1-x_1)/x_1 + (n_2-x_2)/x_2}{n_1} \right) \]

CI for (RR):
\[ \exp(\text{Lower limit}), \exp(\text{Upper limit}) \]

Example: Compare the future risk of sleep disorder among statin users (exposed) versus non-statin users (not exposed)

Confidence Level: 95%

<table>
<thead>
<tr>
<th></th>
<th>Sleep OK</th>
<th>Sleep Dx</th>
<th>Total</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statin user</td>
<td>91</td>
<td>14</td>
<td>105</td>
<td>0.3333</td>
</tr>
<tr>
<td>Non-statin user</td>
<td>309</td>
<td>28</td>
<td>337</td>
<td>0.0705</td>
</tr>
<tr>
<td>Total</td>
<td>460</td>
<td>42</td>
<td>502</td>
<td></td>
</tr>
</tbody>
</table>

\[ RR = \frac{\hat{p}_1}{\hat{p}_2} = \]

CI for \( \ln(\text{RR}) \):

Notes:

1. Now it is time to practice.
2. From your handout and the information provided, calculate a 95% confidence interval for a two-sample (independent) risk ratio in the incidence of sleep disorder by use versus non-use of statins.
3. Make sure to first calculate the confidence interval in the natural logarithm scale.
4. Provide an interpretation for the confidence interval.
1.18 Practice Exercise Answers

Notes:

1. Here are the answers.
2. Note that this example used a z-value of 1.96 for a 95% confidence interval.
3. As seen, the 95% confidence interval for the risk ratio in the incidence of sleep disorder between statin users and non-users ranges from 1.03 to 3.46. Thus, in this example statin users appear to be at a higher risk of developing sleep disorder.
1.19 Practice Exercise Interpretation

**Practice Exercise Interpretation**

**Example:** Compare the future risk of sleep disorder among statin users (exposed) versus non-statin users (not exposed)  
**Confidence Level:** 95%  
**Z value:** 1.96

<table>
<thead>
<tr>
<th></th>
<th>Sleep OK</th>
<th>Sleep Dx</th>
<th>Total</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statin user</td>
<td>91</td>
<td>14 (c1)</td>
<td>105</td>
<td>p₁ = 14 / 105 = 0.1333</td>
</tr>
<tr>
<td>Non-statin user</td>
<td>369</td>
<td>28 (c2)</td>
<td>397</td>
<td>p₂ = 28 / 397 = 0.0705</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>460</td>
<td>42</td>
<td>502</td>
<td><strong>RR</strong> = p₁ / p₂ = 0.1333 / 0.0705 = 1.89</td>
</tr>
</tbody>
</table>

\[
\ln(1.89) \pm 1.96 \sqrt{\frac{91}{105} + \frac{369}{397}}
\]

CI for \(\ln(RR)\):  
\[0.6366 \pm 0.6044 = (0.0322, 0.6044)\]  
\[\exp(0.0322), \exp(1.24)\]  
\[= (1.03, 3.46)\]

From the sample, we estimate that risk of sleep disorder is 1.89 times higher in statin users compared to non-users, and we are 95% confident that the true risk lies between the interval of 1.03 to 3.46.

**Notes:**

1. Here is the interpretation for the 95% confidence interval.
2. The 95% confidence interval for the risk ratio does not include the null value of 1.0, so it appears that statin users are at higher risk of developing sleep disorder compared to non-statin users.
1.20 CI for Two Samples – Independent: Odds Ratio

### CI for Two Samples - Independent: Odds Ratio

Conceptually similar to risk ratio, yet the parameter of interest is the odds ratio (OR), defined as:

\[
\text{Odds of exposure among cases} / \text{Odds of exposure among controls}
\]

#### Example: Prevalence of CVD in Smokers and Non-Smokers (95% C.I.)

<table>
<thead>
<tr>
<th></th>
<th>CVD (D)</th>
<th>No-CVD (D')</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current smoker (E')</td>
<td>81</td>
<td>663</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smoker (E')</td>
<td>298</td>
<td>2757</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{OR} = \frac{81 / 298}{663 / 2757} = 1.13
\]

\[Z = 1.96\]

\[
\ln(\text{OR}) \pm 1.96 = \frac{1}{\sqrt{\frac{81}{663} + \frac{298}{2757}}}
\]

\[
\ln(1.13) \pm 1.96 = 0.122 \pm 0.260 = (0.138, 0.382)
\]

\[
(\exp(0.138), \exp(0.382)) = (0.87, 1.37)
\]

### Notes:

1. Similar to the risk ratio, here are the guidelines for calculating a confidence interval for 2 independent samples and the odds ratio as the parameter of interest.
2. Note that calculation of the confidence interval requires the additional step of performing a calculation in natural logarithm scale, and then doing a transformation.
3. In this example, the 95% confidence interval for the odds ratio ranges from 0.87 to 1.47.
4. Since the 95% confidence interval includes the null value of 1.0, it is plausible (from these data) that smokers are at either lower or higher risk of cardiovascular disease (CVD) compared to non-smokers.
### 1.21 Practice Exercise

#### Practice Exercise

| OR = Odds of exposure among cases / Odds of exposure among controls |
|---|---|---|---|---|---|
| | Sleep Dx | Sleep OK | Cases | Controls |
| Statin user (E-1) | 14 | 91 | a | b |
| Non-statin user (E-) | 28 | 369 | c | d |

#### Prevalence of Sleep Disorder Among Statin and Non-Statin Users (95% CI.)

| OR = \( \frac{a}{c} / \frac{b}{d} \) = | Z = |

\[
\text{Cl for ln(OR)}: \quad \ln\hat{OR} \pm z \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}
\]

\[
\text{Cl for OR:} \quad \exp(\text{Lower limit}), \exp(\text{Upper limit})
\]

#### Notes:

1. Now it is time to practice.
2. From your handout and the information provided, calculate a 95% confidence interval for a two-sample (independent) odds ratio for the odds of sleep disorder by use versus non-use of statins.
3. Make sure to first calculate the confidence interval in the natural logarithm scale.
4. Provide an interpretation for the confidence interval.
1.22 Practice Exercise Answers

Notes:

1. Here are the answers.
2. Note that this example used a $z$-value of 1.96 for a 95% confidence interval.
3. As seen, the 95% confidence interval for the odds ratio for the odds of sleep disorder between statin uses and non-users ranges from 1.03 to 4.01. Thus, in this example statin users appear to be at a higher risk of developing sleep disorder.
### 1.23 Practice Exercise Interpretation

#### Practice Exercise Interpretation

**OR = Odds of exposure among cases / Odds of exposure among controls**

**Example: Prevalence of Sleep Disorder Among Statin and Non-Statin Users**

<table>
<thead>
<tr>
<th>Statin user (E+)</th>
<th>Sleep Dx</th>
<th>Sleep CK</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14</td>
<td>91</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Non-statin user (E-)</td>
<td>28</td>
<td>369</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

\[
OR = \frac{14 / 28}{91 / 369} = 2.027
\]

\[
\text{or } 0.7066 \pm 0.6813 = (0.253, 1.3879)
\]

\[
\exp\{0.0253\}, \exp\{1.3879\} = (1.03, 4.01)
\]

From the sample, we estimate that the odds of statin use among persons with sleep disorder are 2.03 times higher that the odds of statin-use among persons without sleep disorder, and we are 95% confident that the value lies between the interval of 1.03 to 4.01.

#### Notes:

1. Here is the formal interpretation for the results, including the point estimate (odds ratio) and the 95% confidence interval.
2. As seen, being a statin-user appears to increase the risk of sleep disorder.
End of Section 3.5