Stat 3411  Section 6.5  Proportions

(1) \[ \hat{p} = \frac{9}{235} \]  For proportions \( \sigma^2 = p(1-p) \approx \hat{p}(1-\hat{p}) \)

\[ SE_{\hat{p}} = \sqrt{\frac{\sigma^2}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

\[ \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{235}} \]

(b) 

Z-Test

Ho: \( p = 0.07 \)

SE if \( H_0 \) true 0.017

Calculated \( z \) -1.905

\( p \)-values

\( Ha: \mu \neq \) 0.057

\( Ha: \mu < \) 0.028

\( Ha: \mu > \) 0.972

The only statistically significant test at the \( \alpha=0.05 \) level is for \( Ha: \mu < 0.07 \). For this test we reject \( H_0: p = 0.07 \) in favor of \( H_a: p < 0.07 \).

(2) Going through the details again behind the usual formulas:

\[ Var(\hat{p}_1) = \frac{\sigma^2}{n_1} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \]

\[ Var(\hat{p}_2) = \frac{\sigma^2}{n_2} = \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \]

For independent \( \hat{p}_1 \) and \( \hat{p}_2 \), \( Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2) \approx \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \)

\[ SE(\hat{p}_1 - \hat{p}_2) = \sqrt{Var(\hat{p}_1 - \hat{p}_2)} \approx \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.56(1-0.56)}{100} + \frac{0.33(1-0.33)}{100}} \]

0.23 ± 1.96 * \[ \sqrt{\frac{0.56(1-0.56)}{100} + \frac{0.33(1-0.33)}{100}} \]

- We use \( Z_{0.025} \) rather than \( t_{0.025} \) because the 1's and 0's that go into the means \( \hat{p}_1 \) and \( \hat{p}_2 \) are not normally distributed.

- In order for the Central Limit Theorem's approximate normality to work OK we need
o # of "successes": 56 and 33
  ▪ Both > 5. OK

o # of "failures": 100-56 = 44 and 100-33 = 67
  ▪ Both > 5. OK

○ The approximation is OK.

(b, c)
  • If \( p = 0.50 \), then the variance = \( 0.5 \times (1-0.5) = 0.25 \).
  • If \( p = 0.80 \), then the variance = \( 0.8 \times (1-0.8) = 0.16 \).
  • We have large variance if \( p = 0.50 \), so this require the larger sample size.
    o We used 0.50 when finding \( n \) when we didn't have any restrictions on what \( p \) might be.
      ▪ The "worst case" scenario.
      ▪ Since \( p = 0.5 \) is the worst case scenario, the required sample size is less when \( p = 0.8 \).

(d)
Z-Test

<table>
<thead>
<tr>
<th>Combined</th>
<th>( n )</th>
<th>( X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ho: ( p_1 - p_2 = 0 )</td>
<td>200</td>
<td>67</td>
</tr>
</tbody>
</table>

SE if \( H_0 \) true
\[ 0.067 \]

Calculated\[ z \]
\[ 1.348 \]

p-values

<table>
<thead>
<tr>
<th>Ha: ( \mu \neq )</th>
<th>Ha: ( \mu &lt; )</th>
<th>Ha: ( \mu &gt; )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.178</td>
<td>0.911</td>
<td>0.089</td>
</tr>
</tbody>
</table>

None of these test have significant differences since all p-values > 0.05.

(3)
(a) The worst case scenario with largest variance \( p(1-p) \) requiring the largest sample is for \( p=0.5 \)
\[
1.96 \times SE_{\hat{p}} \leq 0.01
\]
\[
1.96 \times \sqrt{\frac{0.5(1-0.5)}{n}} \leq 0.01
\]
\[ n \geq 9604 \]

(b) If \( p \leq 0.25 \), then the worst case \( p \) closest to 0.5 is \( p = 0.25 \).
\[ 1.96 \times SE_{\hat{p}} \leq 0.01 \]
\[ 1.96 \times \sqrt{\frac{0.25(1 - 0.25)}{n}} \leq 0.01 \]
\[ n \geq 7203 \]

(4)

\[ \hat{p} \geq 5 \text{ and } n(1 - \hat{p}) \geq 5 \quad \hat{p} \Rightarrow \geq 5 \text{ and } n(1 - \hat{p}) \geq 5 \]
\[ \text{count} \Rightarrow \geq 5 \text{ and } n - \text{count} \geq 5 \]

\[ \bar{X} = 1.802 \quad S = 0.232 \]
\[ P(Lifetime \leq 100) \approx P(X \leq \log(100)) = P\left(Z \leq \frac{2.000 - 1.802}{0.232}\right) \]
\[ \approx P(Z \leq 0.853) = 0.803 \]
\[ Z_{\text{Low}} = 0.853 - \frac{1.96}{\sqrt{23}} \sqrt{1 + \frac{(23/2) \times 0.853^2}{23 - 1}} = 0.373 \]
\[ P_{\text{Low}} = P(Z \leq Z_{\text{Low}}) = P(Z \leq 0.373) = 0.645 \]
\[ Z_{\text{Low}} = 0.853 + \frac{1.96}{\sqrt{23}} \sqrt{1 + \frac{(23/2) \times 0.853^2}{23 - 1}} = 1.333 \]
\[ P_{\text{Low}} = P(Z \leq Z_{\text{Low}}) = P(Z \leq 1.333) = 0.909 \]