Section 8-4 Inferences about Two Proportions

1. We have proportions from two **independent** simple random samples.
2. For both samples, the conditions $np \geq 5$ and $nq \geq 5$ are satisfied.

For population 1, we let:

p_1 = population proportion

n_1 = size of the sample

x_1 = number of successes in the sample

$\hat{p}_1 = x_1/n_1$ (the *sample* proportion)

$\hat{q}_1 = 1 - \hat{p}_1$

The corresponding meanings are attached to p_2, n_2, x_2, $\hat{p}_2$, and $\hat{q}_2$, which come from population 2.

Mar 18-7:14 PM

Pooled Estimate of p_1 and p_2

- The **pooled estimate of p_1 and p_2** is denoted by $\bar{p}$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

- $\bar{q} = 1 - \bar{p}$

Test Statistic for Two Proportions

For H_0: $p_1 - p_2 = 0$ H_0: $p_1 - p_2 \geq 0$ H_0 $p_1 - p_2 \leq 0$

H_1: $p_1 - p_2 \neq 0$ H_1: $p_1 - p_2 < 0$ H_1: $p_1 - p_2 > 0$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \overset{0}{(p_1 - p_2)}}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

Mar 18-7:27 PM

where $p_1 - p_2 = 0$ (assumed in the null hypothesis)

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{and} \quad \bar{q} = 1 - \bar{p}$$

Confidence Interval Estimate of $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Mar 18-7:29 PM

Ex1a) Treatment Group: $n_1 = 20$, $x_1 = 10$ Placebo Group: $n_2 = 25$, $x_2 = 15$

Perform a hypothesis test to see if the following groups are different. Use .05 for significance level.

1. H_0: $p_1 - p_2 = 0$ H_1: $p_1 - p_2 \neq 0$

2. find $\bar{p}$ and $\bar{q}$: $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ $\frac{10 + 15}{20 + 25} = .556$

3. Test statistic: $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$

$\hat{p}_1 = 10/20 = .2$

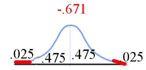
$\hat{q}_1 = 15/25 = .6$

$$\frac{.5 - .6}{\sqrt{\frac{.556(.444)}{20} + \frac{.556(.444)}{25}}}$$

$$\frac{-.1}{.1491}$$

$-.671$

4. Critical Value:



± 1.96

5. Fail to reject. There is not sufficient evidence to support the claim that they are different or $p_1 - p_2 \neq 0$.

Mar 21-8:06 PM

Ex1b) Use the data in the preceding example to construct a 95% confidence interval.

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

$$E = 1.96\sqrt{\frac{.5(.6)}{20} + \frac{.5(.6)}{25}} = .322$$

$(\hat{p}_1 - \hat{p}_2) = .5 - .6 = -.1$

$-.1 - .322 < (\hat{p}_1 - \hat{p}_2) < -.1 + .322$

$-.422 < (\hat{p}_1 - \hat{p}_2) < .222$

Not a significant difference between the proportions

Mar 21-8:25 PM

Ex2a) In a preliminary study, it was found that 16% of the men in the treatment group had headaches. There were 734 men getting treatment. In the placebo group, 725 men were tested and 4% of them experienced headaches. Is there sufficient evidence to support the claim that men taking the medication had an increase in headaches compared to the placebo group? Use .01 as the significance level.

1. H_0: $p_1 - p_2 \leq 0$

   H_1: $p_1 - p_2 > 0$

2. Find $\bar{p}$ and $\bar{q}$: $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$x_1 = .16(734) = 117.44 \approx 117$

$x_2 = .04(725) = 29$

$\frac{117 + 29}{734 + 725} = 0.100069 = \bar{p}$

$1 - .100069 = .899931 = \bar{q}$

Mar 19-3:04 PM

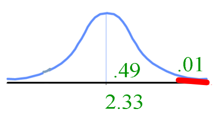
3. Test statistic:

$$z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1}+\frac{\bar{p}\bar{q}}{n_2}}}$$

$$=\frac{\frac{117}{734}-\frac{29}{725}}{\sqrt{\frac{.100069(.899931)}{734}+\frac{.100069(.899931)}{725}}}$$

7.60

4. Critical Value:



5. Conclusion: reject H_0

There is sufficient evidence to support $p_1>p_2$.

Mar 20-2:00 PM

Ex2b) Find the confidence interval between the treatment and control group with 99% confidence.

$$(\hat{p}_1-\hat{p}_2)-E<(p_1-p_2)<(\hat{p}_1-\hat{p}_2)+E$$

where $E=z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$

$$E=2.575\sqrt{\frac{\left(\frac{117}{734}\right)\left(\frac{617}{734}\right)}{734}+\frac{\left(\frac{29}{725}\right)\left(\frac{696}{725}\right)}{725}}$$

$E=.0395$

$$(\hat{p}_1-\hat{p}_2)=\frac{117}{734}-\frac{29}{725}=.1194$$

$$.1194-.0395<p_1-p_2<.1194+.0395$$

$$.0799<p_1-p_2<.1589$$

Conclusion: The interval does NOT contain 0 so there is a significance difference between the two proportions.

Mar 20-2:01 PM