Chapter 20 — Comparing two Proportions

The following method works well when n_1 and n_2 are quite large and there are 10+ successes and failures in each sample.

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \text{ is approx } N(0,1)$$

Thus an approximate confidence interval is

$$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm Z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Ex: 986 of 2253 **men** and 923 of 2629 females (aged 19-25) who were sampled lived with their parents. With 95% confidence,

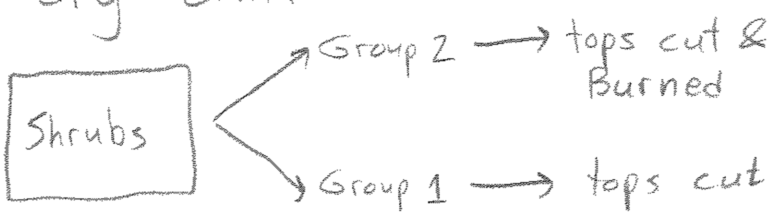
$$p_1 - p_2 = .0866 \pm 1.96(.014)$$

$$= .0866 \pm .0277$$

Again here in the two-population case, the "plus 4" method gives better results and also works for small sample sizes.

If n_1 & $n_2 \geq 5$ this method is appropriate.

Ex: Fire is a serious threat to shrubs in dry climates.



By how much does burning the shrub reduce its chance of growing back?
(ie what is $p_1 - p_2$)

Let p_1 be the proportion of shrubs in the population that would grow back if cut.

Let p_2 be the proportion of shrubs in the population that would grow back if BURNED.

Suppose 12 of 12 that were cut grew back and 8 of 12 that were burned grew back.

Using the "plus four" method we will all 1 success and 1 failure to each sample.

So, $\tilde{p}_1 = \frac{13}{14}$ $\tilde{p}_2 = \frac{9}{14}$

$$p_1 - p_2 = (\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$$

A 90% confidence interval is

$$p_1 - p_2 = \left(\frac{13}{14} - \frac{9}{14}\right) \pm 1.645 \sqrt{\frac{\frac{13}{14}\left(1-\frac{13}{14}\right)}{14} + \frac{\frac{9}{14}\left(1-\frac{9}{14}\right)}{14}}$$

$$= .2857 \pm .2392$$

$$= .047 \text{ to } .525$$

Thus we are 90% confident that burning reduces the percent of shrubs that resprout by 4.7 to 52.5 percentage points.

Hypothesis Testing with 2 proportions.

Ex: Would you marry a person from a lower social class?

This question was asked to 385 black students at historically black schools.

91 of 149 men said yes. $\left(\hat{p}_1 = \frac{91}{149} = .611\right)$
117 of 236 women said yes. $\left(\hat{p}_2 = \frac{117}{236} = .496\right)$

Is there reason to think that the true population proportions p_1 and p_2 are different?

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 \neq 0$

$$z = \frac{-(p_1 - p_2) + (\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$ is approx $N(0,1)$

$\boxed{\text{The problem is that we don't know } p_1 \text{ or } p_2}$

We are assuming (see H_0) that $p_1 = p_2$. Thus an estimate for p_1 & p_2 would be the "pooled sample proportion"

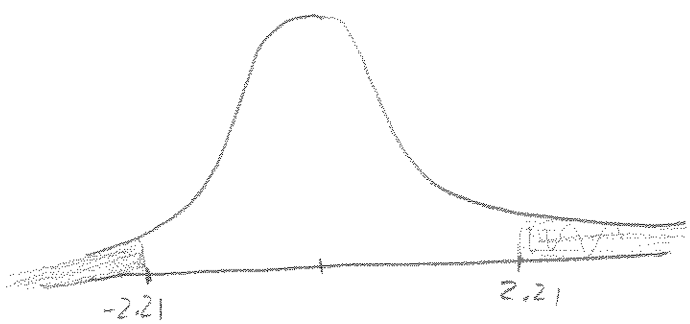
$$\hat{p} = \frac{\text{total successes in both samples combined}}{\text{total individuals in both samples combined}} = \frac{91+117}{149+236}$$

$= .5403$

We compute our test statistic by plugging $\hat{p}$ in for p_1 & p_2.

$$z = \frac{+(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{+(.611 - .496)}{\sqrt{(.5403)(1-.5403)\left(\frac{1}{149} + \frac{1}{236}\right)}}$$

$$= +2.21$$



The P-value is $2(.0136) = .0272$

Thus we can reject at $\alpha = .05$ but not at $\alpha = .01$.

Note: We use the standard Normal table. We do <u>NOT</u> use the t-table. The t-table is only used when testing the **mean**. It requires the assumption that the population is normally distributed.