Section 8.2 - Comparing Two Proportions

Statistics 104

Autumn 2004

Comparing Two Proportions

Two-sample problems

- Want to compare the responses in two groups or treatments
- Each sample is considered to be a sample from a distinct population
- The responses in each group are independent of those in the other group

Setup:

Population	Population Proportion	Sample Size	Count of Successes	Sample Proportion
1	p_1	n_1	X_1	$\hat{p}_1 = \frac{X_1}{n_1}$
2	p_2	n_2	X_2	$\hat{p}_2 = \frac{X_2}{n_2}$

X_1 and X_2 both have binomial distributions

Two problems of interest:

1. Confidence interval for $p_1 - p_2$
2. Tests on the hypothesis $H_0 : p_1 = p_2$ (or $H_0 : p_1 - p_2 = 0$)

Will base inference on $D = \hat{p}_1 - \hat{p}_2$.

When both sample sizes are sufficiently large, D will be approximately normally distributed with

$$\mu_D = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2$$

and

$$\sigma_D^2 = \sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Confidence interval for $p_1 - p_2$

Like the one sample case, we want to adjust the estimates used for p_i.

However instead of adding 2 successes and 2 failures to each group, it will be one and one for each.

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} \quad \text{and} \quad \tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2}$$

The estimated difference between the populations for the confidence interval is

$$\tilde{D} = \tilde{p}_1 - \tilde{p}_2$$

The standard error of this difference is

$$SE_{\tilde{D}} = \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2}}$$

An approximate C level confidence interval for $p_1 - p_2$ is

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

As in the one sample case we want to us D as our point estimate of $p_1 - p_2$.

Example: Use of aspirin to prevent strokes

Aspirin group: 63 of 78 with no strokes

Placebo group: 43 of 77 with no strokes

Group	$\hat{p}$	$\tilde{p}$
Aspirin	$\frac{63}{78} = 0.808$	$\frac{63+1}{78+2} = 0.800$
Placebo	$\frac{43}{74} = 0.558$	$\frac{43+1}{77+2} = 0.557$

$$D = 0.808 - 0.558 = 0.250$$
$$\tilde{D} = 0.800 - 0.557 = 0.243$$

$$SE_{\tilde{D}} = \sqrt{\frac{0.800(1-0.800)}{78+2} + \frac{0.557(1-0.557)}{77+2}} = 0.0716$$

Then a 95% CI for $p_1 - p_2$ is

$$\begin{aligned} CI &= 0.243 \pm 1.96 \times 0.0716 \\ &= 0.243 \pm 0.140 = (0.103, 0.383) \end{aligned}$$

The interval suggested here is valid as long as both sample sizes are at least 10, and C ranges from 90% to 99%.

To calculate this interval in Stata we can use the `prtesti` function

```
. prtesti 80 64 79 44, count

Two-sample test of proportion                    x: Number of obs = 80
                                                 y: Number of obs = 79

------------------------------------------------------------------------------
Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |         .8   .0447214                      .7123477    .8876523
       y |    .556962   .0558881                      .4474233    .6665008
---------+--------------------------------------------------------------------
    diff |    .243038   .0715785                      .1027466    .3833293
         |  under Ho:   .0740355     3.28   0.001
------------------------------------------------------------------------------
```

Significance Tests

Since D is approximately normally distributed (assuming the sample sizes are large enough) we will use a z test similar to the one sample case.

Will only look at $H_0 : p_1 - p_2 = 0$. We will skip the case of testing a non-zero difference.

Under the null hypothesis, call the common proportion p. Then the standard deviation of D is

$$
\begin{aligned}
\sigma_D &= \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} \\
&= \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}
\end{aligned}
$$

If there is no difference in the proportion of successes between the two populations, we can combine the samples, giving the pooled estimate

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1}{n_1 + n_2}\hat{p}_1 + \frac{n_2}{n_1 + n_2}\hat{p}_2$$

So this is a weighted average like s_p^2. Since its a weighted average, $\hat{p}$ must lie between $\hat{p}_1$ and $\hat{p}_2$.

Then the standard error of D is

$$SE_D = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Then the test statistic for examining the hypothesis $H_0 : p_1 = p_2$ is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_D} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

For the aspirin example

$$\begin{aligned}
D &= 0.249 \\
\hat{p} &= \frac{63+43}{78+77} = \frac{106}{155} = 0.684 \\
SE_D &= \sqrt{0.684(1-0.684)\left(\frac{1}{78} + \frac{1}{77}\right)} = 0.0747 \\
z &= \frac{0.249}{0.0747} = 3.337
\end{aligned}$$

For $H_A : p_1 \neq p_2$,

$$p - \text{value} = 2 \times P[Z \geq 3.337] = 0.00085$$

Not surprisingly, we are getting strong evidence that the two probabilities are different.

Stata output:

```
. prtesti 78 63 77 43, count

Two-sample test of proportion                     x: Number of obs =  78
                                                  y: Number of obs =  77

------------------------------------------------------------------------------
Variable |       Mean   Std. Err.       z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |   .8076923   .0446246                         .7202298    .8951548
       y |   .5584416   .0565897                         .4475277    .6693554
---------+--------------------------------------------------------------------
    diff |   .2492507   .0720677                         .1080007    .3905008
         |  under Ho:   .0746952     3.34   0.001
------------------------------------------------------------------------------

               Ho: proportion(x) - proportion(y) = diff = 0

      Ha: diff < 0              Ha: diff != 0               Ha: diff > 0
       z =  3.337                 z =  3.337                 z =  3.337
   P < z =  0.9996           P > |z| =  0.0008           P > z =  0.0004
```

As we are using a normal approximation, we need to worry about the sample size. A common rule of thumb is to have at least 5 successes and 5 failures in each group.

Relative Risks

Another summary to compare two proportions

$$\widehat{RR} = \frac{\hat{p}_1}{\hat{p}_2}$$

Example: Infant mortality in New York City in 1974

Looked at one year death rates

Birthweight	Dead	Alive	Total
$\leq$ 2500 grams	530	4340	4870
$>$ 2500 grams	333	32637	32970

What is the relative risk of death?

$$\leq 2500\text{grams} \quad : \quad \hat{p}_1 = \frac{530}{4870} = 0.1088$$
$$> 2500\text{grams} \quad : \quad \hat{p}_2 = \frac{333}{32970} = 0.0101$$

$$\widehat{RR} = \frac{0.1088}{0.0101} = 10.78$$

So the risk of death for low birthweight babies is almost 11 times higher than the risk for normal birthweight babies.

What if we look at the chance of being alive

$$\begin{aligned} \leq 2500\text{grams} &: \quad \hat{q}_1 = \frac{4340}{4870} = 0.8912 = 1 - \hat{p}_1 \\ > 2500\text{grams} &: \quad \hat{q}_2 = \frac{32337}{32970} = 0.9899 = 1 - \hat{p}_2 \end{aligned}$$

$$\widehat{RR}_{alive} = \frac{0.8912}{0.9899} = 0.9086$$

Note that the RR based on the q's in not a simple function of the RR based on the p's.

The difference in proportions works much nicer. Its easy to show that

$$\hat{q}_1 - \hat{q}_1 = -(\hat{p}_1 - \hat{p}_2)$$

Even with this lack of symmetry with the relative risk, it is a useful measure, particularly with small proportions

Example: Incidence of Rhabdomyolysis and Lipid-Lowering Drugs

(JAMA, December 1, 2004 – Vol 292, No. 21, pages 2585-2590)

Drug	n	Rhabdomyolysis Cases	$\hat{p}$
Cerivastatin (Baycol)	12695	10	0.000788
Atorvastatin (Lipitor)	130865	8	0.000061

$$\widehat{RR} = \frac{0.000788}{0.000061} = 12.89$$

$$\hat{p}_1 - \hat{p}_2 = 0.000788 - 0.000061 = 0.000727$$

While the absolute difference between the two probabilities is small, it's not really the correct scale to be describing the difference.

The ratio of almost 13 times is a better description of the increased risk of problems with Cerivastatin.

Note that Cerivastatin was voluntarily removed from the market by Bayer in August 2001 for due to reports of fatal cases of Rhabdomyolysis (a severe muscle reaction to the drug).