MATH 124 Spring 2005

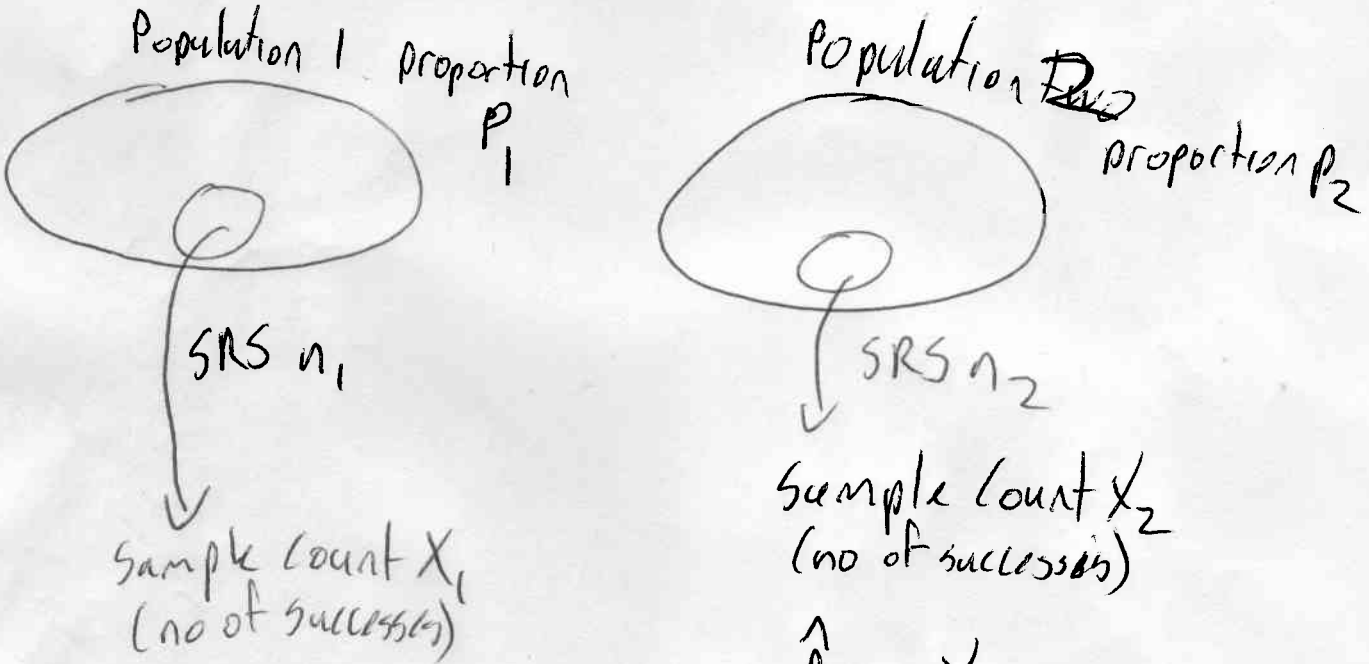
Lecture: 24
Date: May 19, 2005

Skills you should acquire from this lecture:
- Two sample proportion situation
- Confidence intervals for difference between two proportions
- Ability to use confidence interval to determine the results of a hypothesis test
- Hypothesis tests about difference between two proportions (online notes only)

Related readings in the textbook:
- Section 8.2

Todays topic is inference for comparing two proportions. Note that, as with inference for a single proportion, these formulaes differ from those in the book.

Two proportion situation



Sample proportion $\hat{p}_1 = \frac{X_1}{n_1}$

Sample proportion. $\hat{p}_2 = \frac{X_2}{n_2}$

To compare p_1 and p_2 we use the difference $D = p_1 - p_2$

Our sample estimate of this quantity is

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

and the Standard Error is

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

It can be shown that provided n_1 and n_2 are both reasonably large that

$\hat{D} = \hat{p}_1 - \hat{p}_2$ is approximately normal

$$N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

Confidence interval for two proportions

Suppose that a SRS of size n_1 is chosen from a large population with proportion p_1 successes and that an independent (separate) SRS of size n_2 is chosen from a large population with proportion p_2 successes.

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

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

where $P(-z^* < z < z^*) = C$

Example

A researcher is interested in whether there is gender bias in text books. She gathers a sample of sentences from 10 different texts. She classifies each reference as either juvenille (boy/girl) or adult (man/woman). From one particular text she finds the following

Gender	n	X(Juvenille)
Female	60	48
Male	132	52

Using a 90% CI determine whether there is a difference between juvenille references bygender.

Let p_1 = proportion of references that are juvenille for females

p_2 = proportion of references that are juvenille for males

$$\hat{p}_1 = \frac{48}{60} = .8$$

$$\hat{p}_2 = \frac{52}{132} = .394$$

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.8(1-.8)}{60} + \frac{.39(1-.39)}{132}}$$

$$= .0668$$

So a 90% CI is given by

$$.8 - .39 \pm 1.645\,(.0668)$$

$$.406 \pm .101$$

So 90% CI for $p_1 - p_2$ is

$(.296 \quad .516)$

Since 0 is not in the interval there is a difference between p_1 and p_2. Furthermore since the interval is above 0 it means that p_1 is higher than p_2. In other words, it seems like there is a gender bias in this text.

Hypothesis testing

For carrying out hypothesis tests we change the standard error estimate a little. Because under the null hypothesis we assume $p_1 = p_2 = p$ this means that

$$\sigma_{\hat{D}} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

reduces to

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

to estimate p we "pool" together data from both samples

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

And so our standard error for the difference (under the null hypothesis assumption) is

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

↑ this p is to signify pooled.

$H_0: p_1 = p_2$

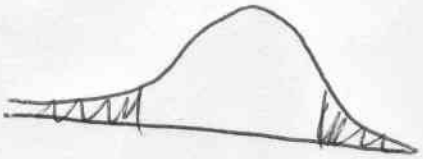
$H_A: p_1 \neq p_2$

(or equivalently $H_0: p_1 - p_2 = 0$, $H_A: p_1 - p_2 \neq 0$)

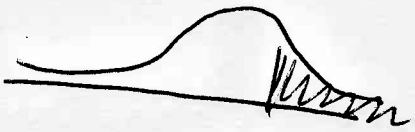
$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE(\hat{D}_p)}$$ ← test statistic distributed $N(0,1)$.

So Pvalues come from Normal distribution table.
As always the alternative gives you the area to look at

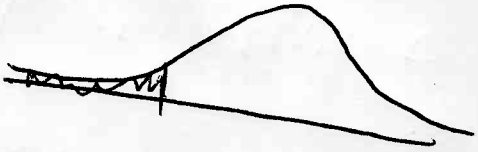
$H_A: p_1 - p_2 \neq 0$ $\quad 2P(Z > |z|)$



$H_A: p_1 - p_2 > 0$ $\quad P(Z > z)$



$H_A: p_1 - p_2 < 0$ $\quad P(Z < z)$



Example

Using earlier data. Lets test whether or not there are more female juvenille references

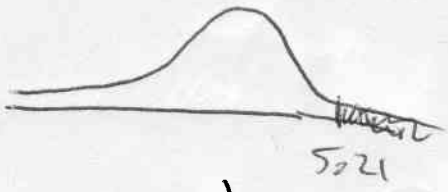
$H_0: p_1 - p_2 \leq 0$

$H_A: p_1 - p_2 > 0$

$$\hat{p} = \frac{48+52}{60+132} = \frac{100}{192} = .521$$

$$SE_{\hat{p}} = \sqrt{.521(1-.521)\left(\frac{1}{60} + \frac{1}{132}\right)} = .078$$

$$z = \frac{.8 - .394}{.078} = 5.21$$



$$\text{Pvalue} = P(z > 5.21) \leq P(z > 3.49)$$

$$= 1 - P(z < 3.49)$$

$$= 1 - .9998$$

$$= .0002$$

So reject H_0 and accept H_A. There are more female juvenille references.