Chapter 10 - Lecture 4
Inference About Two Population Proportions

Yuan Huang

April 1, 2013

Review

- Until now we have seen how we make inference on two sample means
- In this lecture we will learn how to make inference on two sample proportions.
- Assume $X \sim Bin(m, p_1)$ and $Y \sim Bin(n, p_2)$.
- We will see only the test that uses the normal approximation.
    - One special case: z test for comparison of proportions.

Construction

- What is a natural estimator for $p_1 - p_2$?
- What is the distribution of the estimator using the normal approximation to binomial?
- Under which conditions, this approximation is valid?

- $X \sim Bin(m, p_1) \rightarrow E(\bar{X}) = p_1, V(\bar{X}) = \frac{1}{m} p_1(1 - p_1)$.
- $Y \sim Bin(n, p_2) \rightarrow E(\bar{Y}) = p_2, V(\bar{Y}) = \frac{1}{n} p_2(1 - p_2)$.

Denote $\bar{X} = \hat{p}_1, \bar{Y} = \hat{p}_2$,

$$\begin{aligned} E(\bar{X} - \bar{Y}) &= p_1 - p_2 \\ V(\bar{X} - \bar{Y}) &= \frac{p_1(1 - p_1)}{m} + \frac{p_2(1 - p_2)}{n} \end{aligned}$$

With with large sample size (this is different for CI and testing,)

Confidence Interval

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{1}{m}p_1(1-p_1) + \frac{1}{n}p_2(1-p_2)}}$$

Conditions:

- $m\hat{p}_1 > 10, m(1-\hat{p}_1) > 10$
- $n\hat{p}_2 > 10, n(1-\hat{p}_2) > 10$

Two-sided: $1-\alpha$ CI is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{1}{m}\hat{p}_1(1-\hat{p}_1) + \frac{1}{n}\hat{p}_2(1-\hat{p}_2)}$$

Hypothesis test

- Null Hypothesis: $H_0 : p_1 = p_2$
- Test statistic: $z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m} + \frac{1}{n}\right)}}$
- Rejection Regions:
  - $z \geq z_\alpha$ if $H_A : p_1 > p_2$
  - $z \leq -z_\alpha$ if $H_A : p_1 < p_2$
  - $z \leq -z_{\alpha/2}$ and $z \geq z_{\alpha/2}$ if $H_A : p_1 \neq p_2$
- Conditions: This test is used only if $m\hat{p} \geq 10, m(1-\hat{p}) \geq 10, n\hat{p} \geq 10, n(1-\hat{p}) \geq 10$
- Note that:

$$\hat{p} = \frac{X+Y}{m+n} = \frac{m}{m+n}\hat{p}_1 + \frac{n}{m+n}\hat{p}_2$$

Example: Let p_1 denote the true proportion of family who owns second car in Beijing at 2007 and p_2 denote true proportion of family who owns second car in Beijing at 2008. To access whether the odd-and-even license plate rule in Beijing (started at 2008) increases the proportion of family who own the second car, we collect the data from surveys which reports

$$n_1 = 1000, \hat{p}_1 = 0.14, n_2 = 800, \hat{p}_2 = 0.15$$

Based on these two samples, what conclusion can you get?