Math 1070-2: Spring 2008

Lecture 12

Davar Khoshnevisan

Department of Mathematics
University of Utah
http://www.math.utah.edu/~davar

April 16, 2008

Reading from Chapter 9

- §9.1 [CIs & testing for two proportions]
- §9.2 [CIs for the difference of two means; skip testing for the difference of two means]
- §9.4 [CIs & testing for paired data]
- **Skip:** §9.3 and §9.5

CI for the difference of two proportions (Recap)

- 95% CI for $p_1 - p_2$
- Sample propotions: $\hat{p}_1$ and $\hat{p}_2$
- Standard error

$$\mathrm{SE} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- Estimate the standard error

$$\widehat{\mathrm{se}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- 95% CI is $\hat{p}_1 - \hat{p}_2 \pm 1.96 \times \widehat{\mathrm{se}}$
- What about other confidence intervals? Say, 99%?

An example (aspirin vs heart disease)

Example 2 & 3, pp. 427–429

- 5-year double-blind study; 22071 subjects who took aspirin/placebo every other day
- Results:

Group	Heart attack	No heart attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037

- p_1 = proportion of heart attacks in the "placebo group"
- p_2 = proportion of heart attacks in the "aspirin group"
- $\hat{p}_1 = 189/11034 \approx 0.0171$, $\hat{p}_2 = 104/11037 \approx 0.0094$

An example (aspirin vs heart disease)

Example 2 & 3, pp. 427–429, continued

- $p_1 =$ proportion of heart attacks in the "placebo group"
- $p_2 =$ proportion of heart attacks in the "aspirin group"
- $\hat{p}_1 \approx 0.0171$, $\hat{p}_2 \approx 0.0094$
- Standard error estimate

$$\mathrm{SE} = \sqrt{\frac{0.0171(1-0.0171)}{11034} + \frac{0.0094(1-0.0094)}{11037}} \approx 0.0015$$

- 95% CI for $p_1 - p_2$:

$$\underbrace{(0.0171 - 0.0094)}_{0.0077} \pm \underbrace{(1.96 \times 0.0015)}_{0.00294} \approx (0.00476\,,0.01064)$$

- Meaning?

Testing for two proportions

- $H_0: \; p_1 = p_2$ vs $H_a: \; p_1 \neq p_2$
- Similar, but use the *pooled SE*:

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

where $\hat{p}$ denotes the *pooled proportion*.

  - In the aspiring-versus-heart disease problem, $\hat{p}$ is the total proportion of heart attacks; i.e.,

$$\hat{p} = \frac{\overbrace{189 + 104}^{\text{total no of heart attacks}}}{\underbrace{11034 + 11037}_{\text{total size of study}}} = \frac{293}{22071} \approx 0.0133.$$

A testing example

Aspirin vs heart disease, the calculations

- $H_0: \ p_1 = p_2$ vs $H_a: \ p_1 \neq p_2$
- *Pooled SE*:

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

- $\hat{p} \approx 0.0133$, $n_1 = 11034$, $n_2 = 11037$
- pooled SE

$$\approx \sqrt{0.0133(1-0.0133) \times \left(\frac{1}{11034} + \frac{1}{11037}\right)} \approx 0.00154$$

- Therefore,

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\text{pooled SE}} \approx \frac{0.0171 - 0.0094}{0.00154} = 5 \quad \rightarrow \quad P\text{-value} \approx 0$$

A testing example

Aspirin vs heart disease, the report

1. Assumptions:
   - Variables? Quant?
   - Random sampling?
   - Normality?
2. Hypotheses: $H_0: \; p_1 = p_2$ vs $H_a: \; p_1 \neq p_2$
3. Test statistic: $z \approx 5$
4. *P*-value: ≈ 0
5. Conclusion[s]?

CI for the difference of two means

- Two populations with respective means μ_1 and μ_2
  - E.g., men's annual salary vs women's annual salary
- Take 2 independent random sample, one from each population
- 95% CI for $\mu_1 - \mu_2 =$

$$\bar{x}_1 - \bar{x}_2 \pm (1.96 \times \mathrm{SE}),$$

where

$$\mathrm{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

An example (cell-phone use in accidents)

Example 9, pp. 446–447

- University of Utah student data ☺
- Two groups: cell vs no-cell (control); 32 in each
- Variable: mean response time

	no	Mean	SD
Cell phone	32	585.2	89.6
No cell	32	533.7	65.3

- $\mu_1 =$ mean response time for cell-phone users; $\mu_2 =$ mean response time for non cell-phone users
- $\bar{x}_1 = 585.2$, $\bar{x}_2 = 533.7$ $\quad s_1 = 89.6$, $s_2 = 65.3$
  $n_1 = n_2 = 32$

An example (cell-phone use in accidents)

Cell-phone example, continued

- $\mu_1 =$ mean response time for cell-phone users; $\mu_2 =$ mean response time for non cell-phone users
- $\bar{x}_1 = 585.2$, $\bar{x}_2 = 533.7$ $\quad s_1 = 89.6$, $s_2 = 65.3$
  $n_1 = n_2 = 32$
- Standard error =

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \approx 19.599$$

- 95% CI for $\mu_1 - \mu_2 =$

$$\bar{x}_1 - \bar{x}_2 \pm (1.96 \times \text{SE}) \approx \underbrace{585.2 - 533.7}_{51.5} \pm \underbrace{(1.96 \times 19.599)}_{38.41}$$
$$= (13.1\,, 90)$$

- Use the CI for two-sided testing [requires large n_1 and n_2]