Large-Sample Tests/CI's for $p_1 - p_2$

Engineering Statistics
Section 9.4

Josh Engwer

TTU

29 April 2016

Unbiased Point Estimation of $p_1 - p_2$

Proposition

Given any two populations with proportions p_1, p_2 *of some "success".*
Let $\mathbf{X} := (X_1, X_2, \ldots, X_{n_1})$ *be a random sample from the* 1^{st} *population.*
Let $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_{n_2})$ *be a random sample from the* 2^{nd} *population.*
Moreover, let random samples $\mathbf{X}$ & $\mathbf{Y}$ *be independent of one another.*

Then:

- $\widehat{p}_1 - \widehat{p}_2$ *is a unbiased estimator of* $p_1 - p_2$.
- *The standard error of* $\widehat{p}_1 - \widehat{p}_2$ *is* $\quad \sigma_{\widehat{p}_1 - \widehat{p}_2} = \sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$
- *The estimated standard error of* $\overline{X} - \overline{Y}$ *is* $\quad \widehat{\sigma}_{\overline{X} - \overline{Y}} = \sqrt{\dfrac{\widehat{p}_1 \widehat{q}_1}{n_1} + \dfrac{\widehat{p}_2 \widehat{q}_2}{n_2}}$

where $\begin{array}{l} q_1 := 1 - p_1 \\ q_2 := 1 - p_2 \end{array}, \quad \begin{array}{l} \widehat{p}_1 := X/n_1 \\ \widehat{p}_2 := Y/n_2 \end{array}, \quad \begin{array}{l} \widehat{q}_1 := 1 - \widehat{p}_1 \\ \widehat{q}_2 := 1 - \widehat{p}_2 \end{array}$

PROOF: This was shown in the 6.1 Homework Problem #3

Large-Sample z-Test for $p_1 - p_2$ (Test Statistic)

Proposition

Given any two populations with proportions p_1 and p_2 of some "success".
Let $\mathbf{X} := (X_1, X_2, \ldots, X_{n_1})$ be a random sample from the 1^{st} population.
Let $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_{n_2})$ be a random sample from the 2^{nd} population.
Moreover, let random samples $\mathbf{X}$ & $\mathbf{Y}$ be independent of one another.

Suppose an α-level hypothesis test for $p_1 - p_2$ is desired with one of the forms:

$$\begin{array}{ll} H_0: & p_1 - p_2 = 0 \\ H_A: & p_1 - p_2 > 0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & p_1 - p_2 = 0 \\ H_A: & p_1 - p_2 < 0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & p_1 - p_2 = 0 \\ H_A: & p_1 - p_2 \neq 0 \end{array}$$

Then, the corresponding test statistic $W(\mathbf{X}, \mathbf{Y})$ is

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \Longrightarrow \quad Z \overset{approx}{\sim} \text{Standard Normal}$$

where $\begin{array}{l} \widehat{p}_1 := X/n_1 \\ \widehat{p}_2 := Y/n_2 \end{array}$, $\begin{array}{l} \widehat{q}_1 := 1 - \widehat{p}_1 \\ \widehat{q}_2 := 1 - \widehat{p}_2 \end{array}$, $\min\{n_1\widehat{p}_1, n_1\widehat{q}_1, n_2\widehat{p}_2, n_2\widehat{q}_2\} \geq 10$

Large-Sample z-Test for $p_1 - p_2$

Proposition

Populations:	*Two Pop's with proportions* p_1, p_2 *of some "success"*
Realized Samples:	$\mathbf{x} := (x_1, x_2, \ldots, x_{n_1})$ $\quad \mathbf{y} := (y_1, y_2, \ldots, y_{n_2})$ *Samples* $\mathbf{x}$ & $\mathbf{y}$ *are independent of one another*
Test Statistic Value $W(\mathbf{x}, \mathbf{y})$	$z = \dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}\widehat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ *where* $\widehat{p}_1 := X/n_1,\ \widehat{p}_2 := Y/n_2$, $\widehat{p} := (X+Y)/(n_1+n_2)$, $\widehat{q} := 1 - \widehat{p}$
	$\min\{n_1\widehat{p}_1, n_1\widehat{q}_1, n_2\widehat{p}_2, n_2\widehat{q}_2\} \geq 10$

HYPOTHESIS TEST:	***P-VALUE DETERMINATION:***
$H_0 : p_1 - p_2 = 0$ *vs.* $H_A : p_1 - p_2 > 0$	*P-value* $\approx 1 - \Phi(z)$
$H_0 : p_1 - p_2 = 0$ *vs.* $H_A : p_1 - p_2 < 0$	*P-value* $\approx \Phi(z)$
$H_0 : p_1 - p_2 = 0$ *vs.* $H_A : p_1 - p_2 \neq 0$	*P-value* $\approx 2 \cdot [1 - \Phi(|z|)]$

Decision Rule:
If P-value $\leq \alpha$ *then reject* H_0 *in favor of* H_A
If P-value $> \alpha$ *then accept* H_0 *(i.e. fail to reject* H_0*)*

Large-Sample z-CI for $p_1 - p_2$

Proposition

Given any two populations with proportions p_1 and p_2 of some "success".
Let $x_1, x_2, \ldots, x_{n_1}$ be a sample taken from the 1^{st} population.
Let $y_1, y_2, \ldots, y_{n_2}$ be a sample taken from the 2^{nd} population.

*Then the $100(1-\alpha)\%$ **large-sample CI for** $p_1 - p_2$ is approximately*

$$\left((\widehat{p}_1 - \widehat{p}_2) - z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}}, \quad (\widehat{p}_1 - \widehat{p}_2) + z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}} \right)$$

—— OR WRITTEN MORE COMPACTLY ——

$$(\widehat{p}_1 - \widehat{p}_2) \pm z^*_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}}$$

where $\begin{matrix} \widehat{p}_1 := X/n_1 \\ \widehat{p}_2 := Y/n_2 \end{matrix}$, $\begin{matrix} \widehat{q}_1 := 1 - \widehat{p}_1 \\ \widehat{q}_2 := 1 - \widehat{p}_2 \end{matrix}$, $\min\{n_1\widehat{p}_1, n_1\widehat{q}_1, n_2\widehat{p}_2, n_2\widehat{q}_2\} \geq 10$

Textbook Logistics for Section 9.3

- Difference(s) in Notation:

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Probability of Event	$P(A)$	$\mathbb{P}(A)$
Alternative Hypothesis	H_a	H_A
Sample Sizes	m, n	n_1, n_2

- Ignore "Type II Error Probabilities and Sample Sizes" (pg 394-395)

- Ignore "Small-Sample Inferences" section (pg 397)
  - Turns out to be not very straightforward.
  - One can employ **Fisher's Exact Test** or **Barnard's Exact Test**, but they bring complications (and sometimes controversies!)

- Ignore any mention of **one-sided CI's**.

Fin.