Large-Sample Tests/CI's for $\mu_1 - \mu_2$

Engineering Statistics
Section 9.1

Josh Engwer

TTU

25 April 2016

Unbiased Point Estimation of $\mu_1 - \mu_2$

Proposition

Given any two <u>normal</u> populations with means μ_1,μ_2 and std dev's σ_1,σ_2.
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:

- $\overline{X} - \overline{Y}$ *is a unbiased estimator of* $\mu_1 - \mu_2$.
- *The standard error of* $\overline{X} - \overline{Y}$ *is* $$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
- *The estimated standard error of* $\overline{X} - \overline{Y}$ *is* $$\widehat{\sigma}_{\overline{X}-\overline{Y}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

  where s_1^2, s_2^2 *are the corresponding sample variances.*

<u>PROOF:</u> This was shown in the 6.1 Outline problem EX 6.1.2

Large-Sample z-Test for $\mu_1 - \mu_2$ (Test Statistic)

Proposition

Given any two populations with means μ_1 and μ_2.
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.
Moreover, let the sample sizes be "large" meaning $n_1, n_2 > 40$.

Suppose an α-level hypothesis test for $\mu_1 - \mu_2$ is desired with one of the forms:

$$\begin{array}{ll} H_0: & \mu_1 - \mu_2 = \delta_0 \\ H_A: & \mu_1 - \mu_2 > \delta_0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & \mu_1 - \mu_2 = \delta_0 \\ H_A: & \mu_1 - \mu_2 < \delta_0 \end{array} \quad \text{OR} \quad \begin{array}{ll} H_0: & \mu_1 - \mu_2 = \delta_0 \\ H_A: & \mu_1 - \mu_2 \neq \delta_0 \end{array}$$

Then, the corresponding test statistic $W(\mathbf{X}, \mathbf{Y}; \delta_0)$ is

$$Z = \frac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \Longrightarrow \qquad Z \overset{approx}{\sim} \text{Standard Normal}$$

PROOF: (Beyond the scope of this course.)

Large-Sample z-Test for $\mu_1 - \mu_2$ $(\sigma_1, \sigma_2$ unknown$)$

Proposition

Populations:	*Any Two Populations with σ_1, σ_2 unknown*
Realized Samples:	$\mathbf{x} := (x_1, x_2, \ldots, x_{n_1}) \quad (n_1 > 40)$ $\mathbf{y} := (y_1, y_2, \ldots, y_{n_2}) \quad (n_2 > 40)$ *Samples $\mathbf{x}$ & $\mathbf{y}$ are independent of one another*
Test Statistic Value $W(\mathbf{x}, \mathbf{y}; \delta_0)$	$z = \dfrac{(\bar{x} - \bar{y}) - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

HYPOTHESIS TEST:	***P-VALUE DETERMINATION:***
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 > \delta_0$	*P-value* $\approx 1 - \Phi(z)$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 < \delta_0$	*P-value* $\approx \Phi(z)$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 \neq \delta_0$	*P-value* $\approx 2 \cdot [1 - \Phi(\lvert z \rvert)]$

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 $\mu_1 - \mu_2$ $(\sigma_1, \sigma_2$ unknown)

Proposition

Given any two populations with means μ_1 and μ_2.
Let $x_1, x_2, \ldots, x_{n_1}$ be a large sample $(n_1 > 40)$ taken from the 1^{st} population.
Let $y_1, y_2, \ldots, y_{n_2}$ be a large sample $(n_2 > 40)$ taken from the 2^{nd} population.

*Then the $100(1-\alpha)\%$ **large-sample CI for** $\mu_1 - \mu_2$ is approximately*

$$\left((\bar{x} - \bar{y}) - z^*_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad (\bar{x} - \bar{y}) + z^*_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

—— OR WRITTEN MORE COMPACTLY ——

$$(\bar{x} - \bar{y}) \pm z^*_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

If a half-width of w is desired for the $100(1-\alpha)\%$ CI yielding $(\bar{x} - \bar{y}) \pm w$, then the minimum sample size n required to achieve this is:

$$n = \left\lceil \frac{4(z^*_{\alpha/2})^2(s_1^2 + s_2^2)}{w^2} \right\rceil,$$

where s_1, s_2 are "best guesses" for the population std dev's σ_1, σ_2

Textbook Logistics for Section 9.1

- 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
Hypothesized Mean Difference	Δ_0	δ_0

- Skip "Test Procedures for Pop's with Known Variances" (pg 363-365)
  - In practice, population std dev's σ_1, σ_2 will not be known a priori.
- Ignore the "β and the Choice of Sample Size" section (pg 366-367)
- Ignore any mention of **one-sided CI's**.

Fin.