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

Engineering Statistics II
Section 9.1

Josh Engwer

TTU

2018

PART I:

Standard Normal Distribution

Carl Friedrich Gauss (1777-1855)

Sometimes a normal distribution is called a **Gaussian** distribution.

Standard Normal Distribution

Definition

Notation	$Z \sim \text{Normal}(0,1)$ or $Z \sim \text{StdNormal}$
Parameters	(None)
Support	$\text{Supp}(Z) = (-\infty, \infty)$
pdf	$f_Z(z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
cdf	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\xi^2/2}\, d\xi$
Mean	$\mathbb{E}[Z] = 0$
Variance	$\mathbb{V}[Z] = 1$
Model(s)	(Used mainly for Statistical Inference)

Standard Normal Distribution (Properties)

Proposition

Properties of Standard Normal distribution:

- *The standard normal pdf curve is unimodal.*
- *The standard normal pdf curve is bell-shaped.*
- *The standard normal pdf curve is symmetric.*

PROOF: Beyond scope of course. Take **Mathematical Statistics**.

z-Cutoffs (AKA z Critical Values) (Definition)

A key component to some CI's & hypothesis tests is the **z-cutoff**:

Definition

z^*_α is called a **z-cutoff** of the std normal distribution such that
its upper-tail probability is exactly its subscript value α: ($Z \sim$ StdNormal)

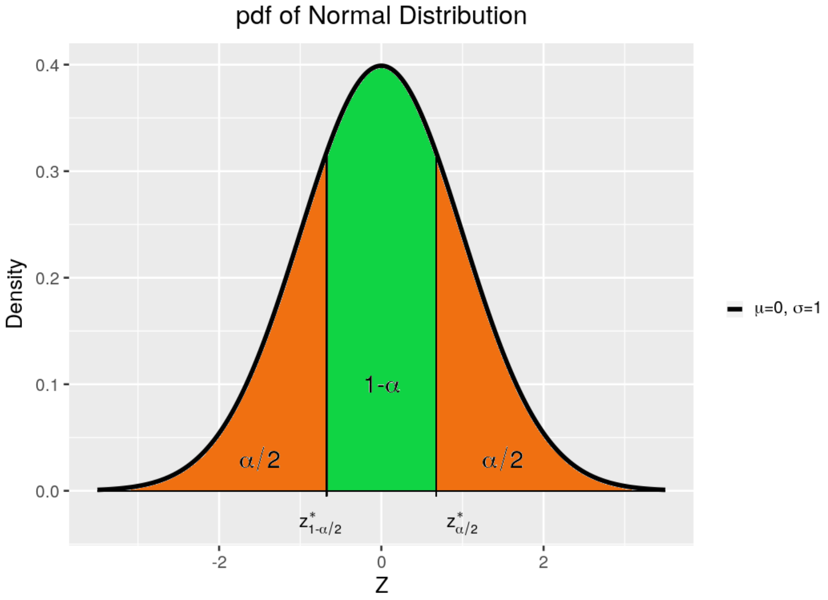
$$\mathbb{P}(Z > z^*_\alpha) = \alpha$$

NOTE: Do not confuse z-cutoff z^*_α with z percentile z_α:

$$\mathbb{P}(Z \leq z_\alpha) = \alpha$$

Another name for z-cutoff is z **critical value**.

z-Cutoffs (Example Plot)



z-Cutoffs (Key Property)

Proposition

Lower-tail z-cutoffs can be determined from appropriate upper-tail t-cutoffs:

$$z^*_{1-\alpha} = -z^*_\alpha$$

PROOF: Follows from the fact that the std normal distribution is symmetric.

z-Cutoffs Table

STD NORMAL z-CUTOFFS, z^*_α $\quad \mathbb{P}(Z > z^*_\alpha) = \alpha, \quad z^*_{1-\alpha} = -z^*_\alpha$

α	**0.2**	**0.1**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**	**0.0005**
z^*_α	0.842	1.282	1.645	1.960	2.054	2.326	2.576	3.090	3.291

PART II:

Large-Sample z-Tests & z-CI's
(Unknown Population Variances σ_1^2, σ_2^2)

A Statistic Related to the Standard Normal Distribution

Theorem

Given any two populations with means μ_1, μ_2 and unknown σ_1, σ_2.
Let $\mathbf{X} := (X_1, \ldots, X_{n_1})$ be a random sample from the 1^{st} population.
Let $\mathbf{Y} := (Y_1, \ldots, Y_{n_2})$ be a random sample from the 1^{st} population.
Moreover, suppose random samples $\mathbf{X}$ & $\mathbf{Y}$ are independent of each other.
Moreover, let the sample sizes be "large", meaning $n_1, n_2 > 40$. Then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{approx}{\sim} \textit{StdNormal}$$

PROOF: Beyond scope of course. Take **Mathematical Statistics**.

Proposition

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

HYPOTHESIS TEST:	***REJECTION REGION AT LVL*** α :
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 > \delta_0$	$z \geq z^*_\alpha$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 < \delta_0$	$z \leq z^*_{1-\alpha}$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 \neq \delta_0$	$z \leq z^*_{1-\alpha/2}$ *or* $z \geq z^*_{\alpha/2}$

Large-Sample z-Test for $\mu_1 - \mu_2$ (Unknown σ_1, σ_2)

Proposition

Population:	*Any Two Populations with unknown* σ_1, σ_2
Realized Samples:	$\mathbf{x} := (x_1, x_2, \cdots, x_{n_1})$, *mean* $\overline{x}$, *std dev* s_1 $(n_1 > 40)$ $\mathbf{y} := (y_1, y_2, \cdots, y_{n_2})$, *mean* $\overline{y}$, *std dev* s_2 $(n_2 > 40)$ *Samples* $\mathbf{x}$ & $\mathbf{y}$ *are independent of each other*
Test Statistic Value: $W(\mathbf{x}, \mathbf{y}; \delta_0)$	$z = \dfrac{(\overline{x} - \overline{y}) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

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

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 $\mu_1 - \mu_2$ (Motivation)

Given any two populations with means μ_1, μ_2 and unknown σ_1, σ_2.
Let $\mathbf{X} := (X_1, \ldots, X_{n_1})$ be random sample from the 1^{st} population with $n_1 > 40$.
Let $\mathbf{Y} := (Y_1, \ldots, Y_{n_2})$ be random sample from the 2^{nd} population with $n_2 > 40$.
Moreover, suppose random samples $\mathbf{X}$ & $\mathbf{Y}$ are independent of each other.
Then, construct the $100(1-\alpha)\%$ CI for parameter difference $\mu_1 - \mu_2$:

1. Produce suitable **pivot**: $Q(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2) := [(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)] / \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
2. Then pivot is approx. standard normal: $Q(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2) \overset{approx}{\sim}$ StdNormal
3. Find constants $a < b$ such that $\mathbb{P}(a < Q(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2) < b) = 1 - \alpha$

   Since std normal pdf is symmetric, $\begin{cases} a = z^*_{1-\alpha/2} = -z^*_{\alpha/2} \\ b = z^*_{\alpha/2} \end{cases}$
4. Manipulate the inequalities to isolate parameter difference $\mu_1 - \mu_2$:
$$(\overline{X} - \overline{Y}) - z^*_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X} - \overline{Y}) + z^*_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$
5. Take independent samples $\mathbf{x} := (x_1, \cdots, x_{n_1})$ & $\mathbf{y} := (y_1, \cdots, y_{n_2})$.
6. Replace point estimators $\overline{X}, \overline{Y}, S_1, S_2$ with $\bar{x}, \bar{y}, s_1, s_2$ from the samples:
$$(\bar{x} - \bar{y}) - z^*_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x} - \bar{y}) + z^*_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Proposition

Given any two populations with means μ_1, μ_2 and unknown σ_1, σ_2.
Let $x_1, x_2, \cdots, x_{n_1}$ be a sample taken from the 1^{st} population with $n_1 > 40$.
Let $y_1, y_2, \cdots, y_{n_2}$ be a sample taken from the 2^{nd} population with $n_2 > 40$.
Moreover, suppose samples $\mathbf{x}$ & $\mathbf{y}$ are independent of each other.
*Then the $100(1-\alpha)\%$ **large-sample z-CI for** $\mu_1 - \mu_2$ is*

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

—- OR WRITTEN MORE COMPACTLY —-

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

Textbook Logistics for Section 9.1

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Probability of Event $A \subseteq \Omega$	$P(A)$	$\mathbb{P}(A)$
Expected Value of rv X	$E(X)$	$\mathbb{E}[X]$
Variance of rv X	$V(X)$	$\mathbb{V}[X]$
Alternative Hypothesis	H_a	H_A
Sample Sizes	m, n	n_1, n_2
z-Cutoffs	$z_\alpha, \quad z_{\alpha/2}$	$z^*_\alpha, \quad z^*_{\alpha/2}$
Hypothesized Mean Difference	Δ_0	δ_0

- Skip "Test Procedures for Pop's with Known Variances" (pg 363-365)
  - In practice, population variances σ_1^2, σ_2^2 will not be known a priori.
- Ignore "β and the Choice of Sample Size" section (pg 366-367)
- Ignore any mention of **one-sided CI's**.

Fin

Fin.