Chapter 9-10

Confidence Intervals and Hypothesis Testing

CI and HT for difference of Two Means

Statistics

Mehmet Güray Güler, PhD

Last updated 17.07.2020

HT and CI for $\mu_1 - \mu_2$

$\boldsymbol{\sigma}$ Known

σ UNKnown

Normal distribution
Linear combinations

Theorem: linear combinations

If $X_1, X_2, \ldots, X_n$ are independent random variables, where

$$X_i \sim N(\mu_i, \sigma_i^2), \text{ for } i = 1, 2, \ldots, n,$$

and $a_1, a_2, \ldots, a_n$ are constant, then the linear combination

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n \sim N(\mu_Y, \sigma_Y{}^2),$$

where

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

HT For $\mu_1 - \mu_2$ when σ *known*

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

is approximately a standard normal variable.

HT For $\mu_1 - \mu_2$ when σ *known*

- **Step1:** The hypothesis are:
  - $H_0: \mu_1 - \mu_2 = d_0$
  - $H_1: \mu_1 - \mu_2 \neq d_0$
- **Step2:** The test statistic for step 2:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

- **Step 3**: R- [$-z_{\alpha\backslash 2}, z_{\alpha\backslash 2}$],
- **Step 4**: Calculate z_{obs} using the formula in step2 using $\bar{x}_{\mathbf{1},\boldsymbol{obs}}$ and $\bar{x}_{\mathbf{2},\boldsymbol{obs}}$
- **Step5:** if z_{obs} is in the critical region, we reject the null hypothesis.

CI For $\mu_1 - \mu_2$ when σ *known*

- Very similar to CI for a single mean.
  - Find the related statistic,
  - Write the probability statement
  - Substitute the statistic.
  - Establish the confidence interval.
- Now, consider the following
  - $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

CI For $\mu_1 - \mu_2$ when σ *known*

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(l - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

HT and CI for $\mu_1 - \mu_2$

σ Known

$\boldsymbol{\sigma}$ UNKnown ➔ Two cases: (i) equal variances and (ii) unequal variances.

HT For $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$

- *Let* ${S_1}^2$ and ${S_2}^2$ be the sample variances to estimate ${\sigma_1}^2$ and ${\sigma_2}^2$.
- Then following is a T random variable

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- with v d.o.f

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}$$

HT For $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$

- **Step1:** The hypothesis are:
  - $H_0: \mu_1 - \mu_2 = d_0$
  - $H_1: \mu_1 - \mu_2 \neq d_0$

- **Step2:** Test statistic HT For $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$

- **Step 3**: R- [-$t_{\alpha\backslash 2}$, $t_{\alpha\backslash 2}$],

- **Step 4**: Calculate t_{obs} using the formula in step2 using $\bar{x}_{\mathbf{1},\boldsymbol{obs}}$ and $\bar{x}_{\mathbf{2},\boldsymbol{obs}}$

- **Step5:** if t_{obs} is in the critical region, we reject the null hypothesis.

CI For $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$

- An approximate $(1-\alpha)$ 100% confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Compare with the previous case. Very Similar!!!:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

HT For $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$

- **Step1:** The hypothesis are:
  - $H_0: \mu_1 - \mu_2 = d_0$
  - $H_1: \mu_1 - \mu_2 \neq d_0$
- **Step2:** The test statistic $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$ with $v = n_1 + n_2 - 2$ and $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
- **Step 3**: R- [$-t_{\alpha\backslash 2}$, $t_{\alpha\backslash 2}$],
- **Step 4**: Calculate t_{obs} using the formula in step2 using $\bar{x}_{\boldsymbol{1,obs}}$ and $\bar{x}_{\boldsymbol{2,obs}}$
- **Step5:** if t_{obs} is in the critical region, we reject the null hypothesis.

CI For $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where s_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t-value with $v = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.