# Hypotheses Testing about the Difference in Two Means : Independent Samples and Population Variances Known

Session 19

01/05/2021

# Hypotheses Testing about the Difference in Two Means : Independent Samples and Population Variances Known

- We consider both samples are independent.
- We will consider several different techniques for analyzing data that come from two samples.
- For eg:- the effectiveness of two brands of toothpaste or whether two brands of tires differ significantly.
- Research might be conducted to study the difference in the productivity of men and women on an assembly line under certain conditions.
- An engineer might want to determine differences in the strength of aluminum produced under two different temperatures

- Let $\overline{x_1}$ be the mean of a sample of size $n_1$ from population with mean $\mu_1$ and variance ${\sigma_1}^2$ and $\overline{x_2}$ be the mean of a sample of size $n_2$ from another population with mean $\mu_2$ and variance ${\sigma_2}^2$ .

- The central limit theorem states that the difference in two sample means ($\overline{x_1}$-$\overline{x_2}$) is normally distributed for large sample sizes (both $n_1$ and $n_2 \geq 30$ ) regardless of the shape of the population.

- The value of Z corresponding to ($\overline{x_1}$-$\overline{x_2}$) is given by

$$Z = \frac{(\overline{x_1} - \overline{x_2}) - E(\overline{x_1} - \overline{x_2})}{S.E.(\overline{x_1} - \overline{x_2})} \sim N(0,1)$$

- Under null hypotheses, $H_0: \mu_1 = \mu_2$ i.e., there is no significant difference in means, we get

$$E(\overline{x_1} - \overline{x_2}) = E(\overline{x_1}) - E(\overline{x_2}) = \mu_1 - \mu_2 = 0$$

$$V(\overline{x_1} - \overline{x_2}) = V(\overline{x_1}) + V(\overline{x_2}) = \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}$$

The covariance term vanishes since the populations are independent.

Thus under $H_0$, the test statistic becomes

$$\boldsymbol{Z} = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} \sim N(0,1)$$

# Remark

- If $\sigma_1^{\,2} = \sigma_2^{\,2} = \sigma^2$ i.e., the samples have been drawn from the populations with common S.D σ, then under $H_0 : \mu_1 = \mu_2$

$$Z = \frac{(\overline{x_1} - \overline{x_2})}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1).$$

# Confidence Interval

- Sometimes being able to estimate the difference in the means of two populations is valuable.
- By how much do two populations differ in size or weight or age? By how much do two products differ in effectiveness? Do two different manufacturing or training methods produce different mean results? The answers to these questions are often difficult to obtain through census techniques. The alternative is to take a random sample from each of the two populations and study the difference in the sample means.

# Confidence Interval

**CONFIDENCE INTERVAL TO ESTIMATE $\mu_1 - \mu_2$ (10.2)**

$$(\bar{x}_1 - \bar{x}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

# Example

- Suppose a study is conducted to estimate the difference between middle-income shoppers and low-income shoppers in terms of the average amount saved on grocery bills per week by using coupons. Random samples of 60 middle-income shoppers and 80 low income shoppers are taken, and their purchases are monitored for one week. The average amounts saved with coupons, as well as sample sizes and population standard deviations are in the table below.

| Middle-Income Shoppers | Low-Income Shoppers |
|---|---|
| $n_1 = 60$ | $n_1 = 80$ |
| $\bar{x}_1 = \$5.84$ | $\bar{x}_2 = \$2.67$ |
| $\sigma_1 = \$1.41$ | $\sigma_2 = \$0.54$ |

- Construct a 95% confidence interval to estimate the difference between the mean amount saved with coupons by middle-income shoppers and the mean amount saved with coupons by low-income shoppers

# Solution

- The $z_{\alpha/2}$ value associated with a 95% level of confidence is 1.96.
- Using C.I. formula,
- $(5.84-2.87)-1.96\left(\sqrt{\frac{1.41^2}{60}+\frac{0.54^2}{80}}\right) \leq \mu_1 - \mu_2 \leq$

  $(5.84-2.87)+1.96\left(\sqrt{\frac{1.41^2}{60}+\frac{0.54^2}{80}}\right)$
- $\Rightarrow$

# Test-5:Test about difference in population means (Population variances are known and Independent samples )

| | Left tail test | Right tail test | Two tail test |
|---|---|---|---|
| Hypotheses | $H_0:\mu_1-\mu_2= 0$<br>$H_1:\mu_1-\mu_2 < 0$ | $H_0:\mu_1-\mu_2= 0$<br>$H_1:\mu_1-\mu_2 > 0$ | $H_0:\mu_1-\mu_2= 0$<br>$H_1:\mu_1-\mu_2 \neq 0$ |
| Test statistic | $Z=\dfrac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1{}^2}{n_1}+\dfrac{\sigma_2{}^2}{n_2}}}$ | | |
| Rejection Rule | Reject $H_0$ if<br>$Z \leq -z_{\alpha}$ | Reject $H_0$ if<br>$Z \geq z_{\alpha}$ | Reject $H_0$ if<br>$\lvert Z\rvert \geq z_{\alpha/2}$ |

Where $z_{\alpha/2}$ is the critical value obtained from the standard normal distribution.

# Example

- Suppose you own a plumbing repair business and employ 15 plumbers. You are interested in estimating the difference in the average number of calls completed per day between two of the plumbers. A random sample of 40 days of plumber A's work results in a sample average of 5.3 calls, with a population variance of 1.99. A random sample of 37 days of plumber B's work results in a sample mean of 6.5 calls, with a population variance of 2.36. Use this information and a 95% level of confidence to estimate the difference in population mean daily efforts between plumber A and plumber B. Interpret the results. Is it possible that, for these populations of days, the average number of calls completed between plumber A and plumber B do not differ?

# Solution

- $n_1$=40, $n_2$=37, $\bar{x}_1$=5.3, $\bar{x}_2$=6.5, ${\sigma_1}^2$=1.99, ${\sigma_2}^2$=2.36
- The $z_{\alpha/2}$ value associated with a 95% level of confidence is 1.96.
- Using C.I. formula,
- (5.3-6.5)-1.96$\left(\sqrt{\frac{1.99}{40}+\frac{2.36}{37}}\right) \le \mu_1 - \mu_2 \le$

  (5.3-6.5)+1.96$\left(\sqrt{\frac{1.99}{40}+\frac{2.36}{37}}\right)$

$\Rightarrow (-1.8604 \le \mu_1 - \mu_2 \le -0.5396)$

- If the confidence interval contains the hypothesized value 0 (here), do not reject $H_0$. Otherwise, reject $H_0$.

# Hypothesis testing

- $H_0:\mu_1-\mu_2 = 0$ against $H_1:\mu_1-\mu_2$ ‡ $0$
- $\alpha = 0.05$ and Test statistic is $\boldsymbol{Z} = \dfrac{(\overline{x_1}-\overline{x_2})-(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}}$

$$= \frac{(5.3-6.5)-0}{\left(\sqrt{\frac{1.99}{40}+\frac{2.36}{37}}\right)} = -3.5619$$

- Rejection rule is Reject $H_0$ if $|Z| \geq z_{\alpha/2}$
- $|-3.5619| > 1.96$, Reject $H_0$ and conclude that there is enough statistical evidence to conclude that the average number of calls completed between plumber A and plumber B differ.