Chapter 8 - Statistical Modeling and Inference

STAT 251

Lecture 30

Hypothesis Testing About Difference of Two Population Means

Dr. Lasantha Premarathna

Chapter 8 - Learning Outcomes

- Point Estimation for μ ans σ
- Bias of an estimator
- Confidence Interval for μ
- Testing of Hypotheses about μ
- One sample problems
- Two sample problems

Hypothesis Testing About Difference of Two Population Means, $\mu_1 - \mu_2$

Two Sample Problems

Compare the means of two independent populations, assuming equal population standard deviations

- Suppose we draw a random sample from each of the two independent populations

  Population means μ_1 and μ_2

  Population standard deviations σ_1 and σ_2

Hypothesis Testing About Difference of Two Population Means, $\mu_1 - \mu_2$

Hypotheses: takes one of the following 3 forms

- $H_0 : \mu_1 - \mu_2 \geq \Delta_0$ vs $H_a : \mu_1 - \mu_2 < \Delta_0$ $\Leftarrow$ left-tailed test
- $H_0 : \mu_1 - \mu_2 \leq \Delta_0$ vs $H_a : \mu_1 - \mu_2 > \Delta_0$ $\Leftarrow$ right-tailed test
- $H_0 : \mu_1 - \mu_2 = \Delta_0$ vs $H_a : \mu_1 - \mu_2 \neq \Delta_0$ $\Leftarrow$ two-tailed test

where Δ_0 is the hypothesized value of the population mean.

Example: if the hypotheses are

$$H_0 : \mu_1 \geq \mu_2 \text{ vs } H_a : \mu_1 < \mu_2$$

Then we can give hypotheses as

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ vs } H_a : \mu_1 - \mu_2 < 0$$

in this case $\Delta_0 = 0$

Hypothesis Testing About Difference of Two Population Means, $\mu_1 - \mu_2$

Assumptions

- random samples from each of the population is drawn
- sample individuals are independent of each other
- both populations are normal or we need reasonably large samples to validate using the CLT
- both population distributions have equal variances ($\sigma_1^2 = \sigma_2^2$)

Hypothesis Testing About Difference of Two Population Means, $\mu_1 - \mu_2$

Test Statistic

- We select a simple random sample of size n_1 from population 1 and a simple random sample of size n_2 from population 2
- Let $\bar{x}_1$ is the mean of the sample 1 and $\bar{x}_2$ is the mean of the sample 2

$$\text{test statistic}: \quad t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$$df = n_1 + n_2 - 2$$

s_p is the pooled standard deviation

The Pooled Standard Deviation (s_p)

- This method requires the assumption that population variances are equal

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- The pooled standard deviation (s_p) estimates the common value σ

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$(1-\alpha)100\%$ Confidence Interval (CI) for the Difference Between Two Population Means, $(\mu_1 - \mu_2)$

Point estimator of the $\mu_1 - \mu_2$ is $(\bar{x}_1 - \bar{x}_2)$

$$\text{CI} \quad \Rightarrow \quad \text{point estimate} \pm \text{margin of error}$$

$(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2},\, n_1+n_2-2} \left(s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Example: 5

Average densities of two types of brick are compared (say type A and type B)

Using the following sample data, test the claim that the true mean densities are equal. Use a 0.05 significance level and assume normality of the two density distributions. Also assume that population variances are equal.

Also find the 95% confidence interval for the population mean difference of densities of type A and the type B bricks.

Type A	Type B
$n_A = 8$	$n_B = 10$
$\bar{x}_A = 22.7$	$\bar{x}_B = 21.5$
$s_A = 0.8$	$s_B = 0.6$

Example 5 - Solution

Let μ_A be the true average density for type A bricks

Let μ_B be the true average density for type B bricks

<u>Hypotheses</u>

$$H_0 : \mu_A = \mu_B$$

$$H_a : \mu_A \neq \mu_B \qquad \Rightarrow \text{this is a two-tail test}$$

Equivalently we can write hypotheses as follows

$$\Rightarrow H_0 : \mu_A - \mu_B = 0 \qquad \Rightarrow \Delta_0 = 0$$

$$H_a : \mu_A - \mu_B \neq 0$$

here, $\alpha = 0.05$

Example 5 - Solution

Pooled variance

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$
$$= \frac{(8-1)0.8^2 + (10-1)0.6^2}{8 + 10 - 2}$$
$$s_p^2 = 0.4825$$

Pooled standard deviation $= s_p = \sqrt{0.4825} = 0.695$

Test statistic

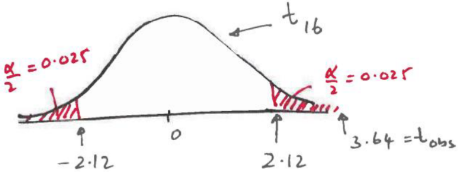
$$t_{obs} = \frac{(\bar{x}_A - \bar{x}_B) - 0}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t_{n_A+n_B-2}$$
$$= \frac{22.7 - 21.5}{0.695\sqrt{\frac{1}{8} + \frac{1}{10}}} \sim t_{16}$$
$$t_{obs} = 3.64$$

Example: 5 - Solution

Critical value approach

$$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$$

need to consider $\frac{\alpha}{2}$ because this is a two-tailed test



critical values are $t_{0.025,16} = 2.12 \leftarrow$ get this value using the t-table and the corresponding negative value

Calculated test statistic is in the rejection region.

$$i.e.\ |t_{obs}| = 3.64 > t_{0.025,16} = 2.12$$

$\Rightarrow$ Reject H_0 at $\alpha = 0.05$

Conclusion:

We conclude that the true mean densities of two types of brick are not equal at the significance level 0.05.

Example: 5 - Solution

95% Confidence Interval for $\mu_A - \mu_B$

$$\Rightarrow \quad (\bar{x}_1 - \bar{x}_2) \pm \; t_{\frac{\alpha}{2}, n_1+n_2-2} \; \left(s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\Rightarrow \quad (22.7 - 21.5) \pm \; 2.12 \; \left((0.695) \sqrt{\frac{1}{8} + \frac{1}{10}}\right)$$

$$\Rightarrow \quad 1.2 \pm 0.33$$

$$\Rightarrow \quad [0.87 \; , \; 1.53]$$

Conclusion:

95% confidence interval for the population mean difference of densities of type A and the type B bricks is between 0.87 and 1.53.

CI approach also can be used to test hypotheses about difference of two population means: ($H_0 : \mu_A - \mu_B = 0$ vs $H_a : \mu_A - \mu_B \neq 0$)

Above calculated confidence interval does not contain the hypothesized value 0 ($\Delta_0 = 0$). Therefore we reject the null hypothesis and conclude that the true mean densities of two types of brick are not equal at the significance level 0.05.

Before the next class ...

Visit the course website at canvas.ubc.ca

- Review Lecture 30 and related sections in the text book
- Topic of next class: **Chapter 10: Analysis of Variance (ANOVA)**