

Review 3

$(1 - \alpha) \cdot 100\%$ confidence interval:
 (point estimate)

$$\pm \left[\left(z_{\alpha/2}, t_{n-1, \alpha/2}, t_{n_1+n_2-2, \alpha/2} \right) \cdot (\text{standard error of point estimate}) \right]$$

Hypothesis testing:

$$\begin{aligned} & \text{test statistic} \\ & = \frac{\text{point estimate} - \text{mean of point estimate under } H_0}{\text{standard deviation (error) of point estimate under } H_0} \end{aligned}$$

Standard deviation (error) of point estimate $\bar{X}_1 - \bar{X}_2$:

(a) Large sample $n_1 \geq 30, n_2 \geq 30$:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(b) Small sample $n_1 < 30, n_2 < 30$, normal populations:

$$s_{\bar{X}_1 - \bar{X}_2}^* = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)},$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{\sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2,i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Standard deviation (error) of point estimate $\bar{P}_1 - \bar{P}_2$:

$$s_{\bar{P}_1 - \bar{P}_2} = \sqrt{\frac{\bar{p}_1(1 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2(1 - \bar{p}_2)}{n_2}}$$

$$s_{\bar{P}_1 - \bar{P}_2}^* = \sqrt{\bar{p}(1 - \bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2}$$

	One sample	Two samples
Point estimate	\bar{x}, \bar{p}	$\bar{x}_1 - \bar{x}_2, \bar{d}, \bar{p}_1 - \bar{p}_2$
Test Statistic	$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \text{ or } z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$ $t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$ $z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}}$	$z = \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{\sigma_{\bar{X}_1 - \bar{X}_2}} \text{ or } z = \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{s_{\bar{X}_1 - \bar{X}_2}}$ $t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{s_{\bar{X}_1 - \bar{X}_2}^*}$ $z \text{ or } t = \frac{\bar{d} - \mu_0}{s_{\bar{D}}}$ $z = \frac{\bar{p}_1 - \bar{p}_2}{s_{\bar{p}_1 - \bar{p}_2}^*}$
Classical approach (critical values)	$-z_\alpha, z_\alpha, z_{\alpha/2}$ $-t_{n-1,\alpha}, t_{n-1,\alpha}, t_{n-1,\alpha/2}$	$-z_\alpha, z_\alpha, z_{\alpha/2}$ $-t_{n_1+n_2-2,\alpha}, t_{n_1+n_2-2,\alpha}, t_{n_1+n_2-2,\alpha/2}$ $-t_{n-1,\alpha}, t_{n-1,\alpha}, t_{n-1,\alpha/2}$
Null distribution (p -value)	$Z, T(n-1)$	$Z, T(n_1 + n_2 - 2), T(n-1)$
C. I.	$\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}} \text{ or } \bar{x} \pm z_{\alpha/2} s_{\bar{x}}$ $\bar{x} \pm t_{n-1,\alpha/2} s_{\bar{x}}$ $\bar{p} \pm z_{\alpha/2} s_{\bar{p}}$	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2}$ $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} s_{\bar{X}_1 - \bar{X}_2}$ $(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,\alpha/2} s_{\bar{X}_1 - \bar{X}_2}^*$ $\bar{d} \pm z_{\alpha/2} s_{\bar{D}}$ $\bar{d} \pm t_{n-1,\alpha/2} s_{\bar{D}}$ $(\bar{p}_1 - \bar{p}_2) \pm z_{\alpha/2} s_{\bar{p}_1 - \bar{p}_2}$

Example 1:

Consider the following hypothesis test.

$$H_0: \mu_1 - \mu_2 = 0 \text{ vs } H_a: \mu_1 - \mu_2 \neq 0.$$

The following data are for two **independent samples** taken from the two normal populations with equal variances.

Sample 1	Sample 2
6, 10, 9, 8, 7	9, 12, 10, 11, 9

(a) With $\alpha = 0.05$, test the hypothesis based on the **classical hypothesis test procedure**.

(b) With $\alpha = 0.01$, test the hypothesis based on the **p-value**.

(c) With $\alpha = 0.05$, using the **confidence interval** method to test the hypothesis.

[Solution:]

$$n_1 = 5, \bar{x}_1 = 8, s_1^2 = 2.5, n_2 = 5, \bar{x}_2 = 10.2, s_2^2 = 1.7, \mu_0 = 0.$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{4 \cdot 2.5 + 4 \cdot 1.7}{5 + 5 - 2} = 2.1.$$

(a)

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \right| = \left| \frac{8 - 10.2 - 0}{\sqrt{2.1 \left(\frac{1}{5} + \frac{1}{5} \right)}} \right| = 2.4 > 2.306 = t_{8,0.025},$$

we reject H_0 .

(b)

$$\begin{aligned} p\text{-value} &= P(|T(n_1 + n_2 - 2)| > |t|) \\ &= P(|T(8)| > 2.4) > P(|T(8)| > 3.355) = 0.01 = \alpha \end{aligned}$$

we do **not** reject H_0 .

(c) A 95% confidence interval for $\mu_1 - \mu_2$ is

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) &\pm t_{n_1+n_2-2, \alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \\ &= (8 - 10.2) \pm t_{8,0.025} \sqrt{2.1 \left(\frac{1}{5} + \frac{1}{5} \right)} = -2.2 \pm (2.306 \cdot 0.9165) \\ &= [-4.313, -0.086]. \end{aligned}$$

Since

$$\mu_0 = 0 \notin [-4.313, -0.086],$$

we reject H_0 .

Example 2:

To determine the effectiveness of a new weight control diet, 8 randomly selected students observed the diet for 4 weeks with the results shown below.

Dieter	Weight (before)	Weight (after)
A	138	135
B	151	147
C	129	132
D	125	127

E	168	155
F	139	131
G	152	144
H	140	142

We like to test the hypothesis $H_0: \mu_1 = \mu_2 + 2$, where μ_1 and μ_2 are the mean weights of the students before and after taking the weight control diet, respectively.

The above data can be considered as the **matched-sample** data.

- (a) For $\alpha = 0.1$, test the above hypothesis using the **classical hypothesis test**.
- (b) For $\alpha = 0.05$, please use **the confidence interval method** to test the above hypothesis.

[Solution:]

(a)

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
3	4	-3	-2	13	8	8	-2

Therefore,

$$\bar{d} = 3.625, s_d = 5.7802.$$

Thus, we do **not** reject H_0 since

$$|t| = \left| \frac{\bar{d} - \mu_0}{(s_d / \sqrt{n})} \right| = \left| \frac{3.625 - 2}{(5.7802 / \sqrt{8})} \right| = 0.795 < 1.895 = t_{7,0.05} = t_{n-1,\alpha/2}.$$

(b) A 95% confidence interval for $\mu_1 - \mu_2$ is

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}} = 3.625 \pm t_{7,0.025} \frac{5.7802}{\sqrt{8}} = [-1.21, 8.46].$$

Since

$$\mu_0 = 2 \in [-1.21, 8.46],$$

we do **not** reject H_0 .

Example 3:

The results of a recent poll on the preference of shoppers regarding two products are shown below.

<u>Product</u>	<u>Shoppers Favoring</u>	
	<u>Shoppers Surveyed</u>	<u>This Product</u>
A	800	560
B	900	612

Let p_1 be the proportion favoring product A and p_2 be the proportion favoring product B.

- (a) Develop a 90% confidence interval estimate for the difference $p_1 - p_2$

between the proportions favoring each product.

(b) Test $H_0: p_1 = p_2$ at $\alpha = 0.05$ based on the classical approach.

(c) Test $H_0: p_1 = p_2$ at $\alpha = 0.05$ based on the p-value method.

[Solution:]

$$n_1 = 800, n_2 = 900, \bar{p}_1 = \frac{560}{800} = 0.7, \bar{p}_2 = \frac{612}{900} = 0.68.$$

(a)

$$\begin{aligned} s_{\bar{p}_1 - \bar{p}_2} &= \sqrt{\frac{\bar{p}_1(1 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2(1 - \bar{p}_2)}{n_2}} = \sqrt{\frac{0.7(1 - 0.7)}{800} + \frac{0.68(1 - 0.68)}{900}} \\ &= 0.0225 \end{aligned}$$

Thus, a 90% confidence interval is

$$(\bar{p}_1 - \bar{p}_2) \pm z_{\alpha/2} s_{\bar{p}_1 - \bar{p}_2} = (0.7 - 0.68) \pm z_{0.05} \cdot 0.0225 = [-0.017, 0.057]$$

(b)

$$\bar{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{560 + 612}{800 + 900} = \frac{1172}{1700} = 0.689.$$

$$s_{\bar{p}_1 - \bar{p}_2}^* = \sqrt{\bar{p}(1 - \bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0.689 \cdot 0.311 \cdot \left(\frac{1}{800} + \frac{1}{900} \right)} = 0.0225.$$

Thus,

$$|z| = \left| \frac{\bar{p}_1 - \bar{p}_2}{s_{\bar{p}_1 - \bar{p}_2}^*} \right| = \left| \frac{0.7 - 0.68}{0.0225} \right| = 0.89 < 1.96 = z_{0.025} = z_{\alpha/2},$$

we do not reject H_0 .

(c)

$$p\text{-value} = P(|Z| > |z|) = P(|Z| > 0.89) = 0.3734 > 0.05 = \alpha,$$

we do not reject H_0 .