Confidence Interval and Hypothesis Testing for Two Samples

Independent Sample - Samples selected from two or more populations in such a way that the occurrence of values in one sample has no influence on the probability of the occurrence of values in the other sample(s).

- The population standard deviations are known

- The population standard deviations are unknown

Confidence interval for $\mu_1 - \mu_2$ for small samples is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm t_c * s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{1}$$

where

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \qquad \text{Pooled standard deviation}$$

t - critical t-value from t-distribution with $df = n_1 + n_2 - 2 = n - 2$

Confidence interval for $\mu_1 - \mu_2$ for small samples

Example 1.: An insurance company wishes to estimate the difference in mean damage to cars that crash into a barricade at 20 mph with a new bumper system versus the older bumper system. A random sample on 8 cars with the new bumper system provided a mean damage equal to \$ 3,950 and a sample standard deviation equal to \$600. A random sample of 10 cars with the older bumper system provided a mean damage of \$3,475 and a sample standard deviation equal to \$650. Develop a 95% confidence interval estimate.

1. Calculate the point estimate:

$$\bar{x}_1 - \bar{x}_2 = \$3,950 - \$3,475 = \$475$$

2. Calculate the pooled standard deviation:

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

$$s_p = \sqrt{\frac{(8 - 1)600^2 + (10 - 1)650^2}{16}}$$

$$s_p = \$628.61$$

3. Determine the critical value, t_c, d_f, from the t-distribution table for a 95% confidence level.

$$d_f = n - 2 = 18 - 2 = 16$$

and with 95% confidence is $t_c = 2.1199$ (from t-chart)

4. The 95% confidence interval is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm t_c * s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$475 \pm (2.120)(628.61)\sqrt{\frac{1}{8} + \frac{1}{10}}$$

$$475 \pm (2.120)(628.61)(0.47)$$

$$\mathbf{475 \pm 626.35}$$

or

$$[475 - 626.35, 475 + 626.35]$$

$$[-151.35, 1101.35]$$

5. ***Conclusion:*** Based on the sample data with 95% confidence we conclude that the difference between mean damage for the two bumpers is between $-151.35 and $1101.35. We can say that there is no significance difference btw two groups.

2. Confidence interval for $\mu_1 - \mu_2$ for big samples
Confidence interval for $\mu_1 - \mu_2$ for big samples is given by:

$$(\text{Point Estimate}) \pm (\text{Critical Value})(\text{Standard Error})$$

or

$$(\bar{x}_1 - \bar{x}_2) \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (2)$$

where the most used z-values are given in the following table: **Table 1.**

Confidence Level	Critical z-value
80%	$z = 1.28$
90%	$z = 1.645$
95%	$z = 1.96$
99%	$z = 2.575$

2. Confidence interval for $\mu_1 - \mu_2$ for big samples

Example 2. The broker for a real estate agency wishes to estimate the difference in the mean number of days a home tales to sell for one-story versus two-story. A 95% confidence level is to be used with sample sizes for one and two-story homes equal to $n_1 = 80$ and $n_2 = 100$ respectively. The standard deviation for one-story homes is 10 days and the standard deviation for two-story homes in 14 days. The sample mean for one-story is 96 days and for two-story is 75 days.

1. Calculate the point estimate:

$$\bar{x}_1 - \bar{x}_2 = (96 - 75) = 21$$

2. Determine the z critical vale

For c=0.95 (level of confidence) the z critical value is

$$z_c = 1.96$$

3. Calculate the Standard Error

$$\text{Standard Error} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{80} + \frac{14^2}{100}} = 1.86$$

4. Calculate he confidence interval:

$$21 \pm (1.96)(1.86)$$
$$21 \pm 3.65 \qquad (\text{Margin of Error} = 3.65)$$

$$[21 - 3.65, 21 + 3.65]$$

[17.35, 24.65]   **is 95% CI**

Conclusion: We are 95% confident that the difference in the mean number of days a home tales to sell for one-story versus two-story is between 17.35 and 24.65

3. Hypothesis Tests for Two Population Means Using Independent Samples

Some situations require to test whether two populations have **equal** means or whether one population mean is **larger (or smaller)** then another.

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_0 : \mu_1 - \mu_2 \geq 0 \quad H_0 : \mu_1 - \mu_2 \leq 0$$
$$H_a : \mu_1 - \mu_2 \neq 0 \quad H_a : \mu_1 - \mu_2 < 0 \quad H_a : \mu_1 - \mu_2 > 0$$

We will use t-Test Statistic which is given as:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example 3. (Two-Tailed Test) One of the big costs associated with using computer printers is the cost of ink cartridges. An independent testing laboratory is interested in testing to see whether the mean number of pages generated per cartridge is the same for the popular name-brand as it is for the leading generic brand. To conduct the test at an alpha = 0.05 level, the lab has randomly selected 15 cartridges of each type and counted the number of pages generated by each. The following sample data were observed:

$$n_1 = 15, \quad n_2 = 15$$
$$\bar{x}_1 = 2,119.7, \quad \bar{x}_2 = 1,777.7$$
$$s_1 = 709.7, \quad s_2 = 593.9$$

1. State the null and alternative hypotheses.

$$H_0 : \quad \mu_1 - \mu_2 = 0$$
$$H_A : \quad \mu_1 - \mu_2 \neq 0$$

2. Determine the critical value t_c.

$df = 15 + 15 - 2 = 28$ and for two tailed test we are looking at $\dfrac{\alpha}{2} = 0.025$ implies

From table:
$$t_c = t_{0.025,28} = 2.0484$$

3. Calculate the pooled standard deviation s_p:

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

$$s_p = \sqrt{\frac{(15 - 1)709.7^2 + (15 - 1)593.9^2}{28}}$$

$$s_p = \sqrt{428195.3}$$

$$s_p = \$654.37$$

4. Compute the test statistic (t) or (p-value):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

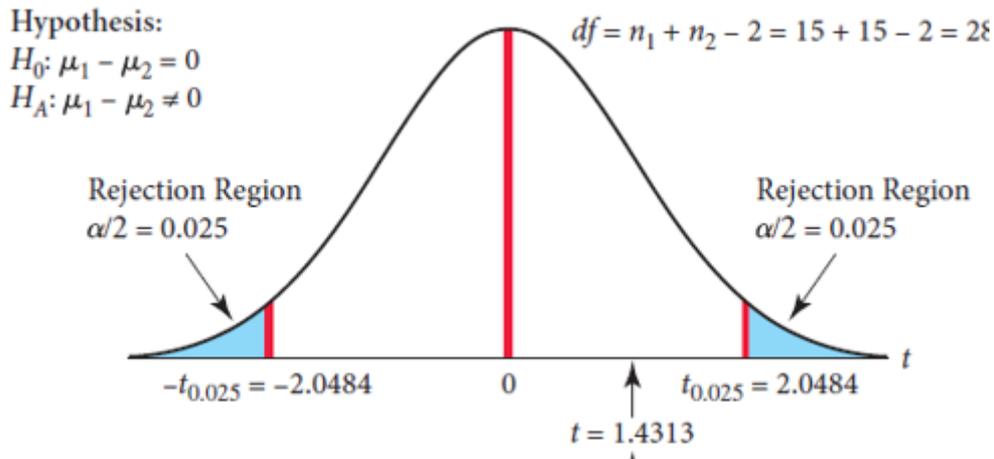$$t = \frac{342 - 0}{654.37\sqrt{\frac{1}{15} + \frac{1}{15}}}$$

$$t = 1.4313$$

5. Decision rule (make graph to see rejection area):

If $t > 2.0484$,
or $t < -2.0484$, reject H_0

Otherwise, do not reject H_0 (fail to reject)

6. Decision:

$1.43 < 2.0484$ so we fail to reject H_0 (do not reject H_0)

Hypothesis:
$H_0: \mu_1 - \mu_2 = 0$
$H_A: \mu_1 - \mu_2 \neq 0$

$df = n_1 + n_2 - 2 = 15 + 15 - 2 = 28$

Rejection Region
$\alpha/2 = 0.025$

Rejection Region
$\alpha/2 = 0.025$

$-t_{0.025} = -2.0484$

0

$t_{0.025} = 2.0484$

$t = 1.4313$

7. **Conclusion:** There is no significant difference in the mean number of pages generated per popular name-brand cartridges and is for generic brand cartridges.

Hypotheses Testing - Big Samples Example: The company is interested in determining whether the average thickness of brick facing products made by Plant 2 has larger m the two plants. You can use the significance level $\alpha = 0.05$

Plant 1	Plant 2
$\bar{x}_1 = 0.501$	$\bar{x}_2 = 0.509$
$\sigma_1 = 0.025$	$\sigma_2 = 0.034$

1. State the null and alternative Hypothesis:

$$H_0: \quad \mu_1 - \mu_2 \geq 0 \quad (\mu_1 \geq \mu_2)$$
$$H_A: \quad \mu_1 - \mu_2 < 0 \quad (\mu_1 < \mu_2)$$

2. Determine the z critical value for ($\alpha = 0.05$). You can use from t- table, and the $z_c = z_{0.05} = -1.645$

3.Calculate z-test statistic:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{3}$$

6

$$z = \frac{(0.501 - 0.509) - 0}{\sqrt{\frac{0.025^2}{n_1} + \frac{0.034^2}{n_2}}}$$

$$z = -1.90$$

4. Decision rule:

If $z < -1.645$ reject H_0

                   Otherwise, do not reject H_0 (fail to reject)

5. Decision:

Since $-1.90 < -1.645$,   we reject H_0