Comparing Two Means

Independent Samples
Paired Samples

Samples

- Most comparisons of groups use **independent samples** from the groups. The observations in one sample are *independent* of those in the other sample.
- Significance test for comparing two means with **paired samples**, that is, sample observations are *naturally paired*. This is a result of a group being tested twice. Usually, this is used in determining the difference of means of before and after observations.

Independent Samples

Most comparisons of groups use **independent samples** from the groups. The observations in one sample are *independent* of those in the other sample.

Step 1: Assumptions

- Check assumptions
  - A quantitative response variable for the two groups
  - Independent random samples
  - Approximately normal population distribution for each group
  - For large sample sizes, use the z-table
  - For small sample sizes, $n_1, n_2 < 30$, use the t-table

Example:

Records of 40 used passenger cars and 40 used pickup trucks (none used commercially) were randomly selected to investigate whether there was any difference in the mean time in years that they were kept by the original owner before being sold. For cars the mean was 5.3 years with standard deviation 2.2 years. For pickup trucks the mean was 7.1 years with standard deviation 3.0 years.

Test whether there is a difference between the means. Use a 10% level of significance.

Step 1: Assumptions

- Both sample sizes are large enough (use the z-table)

Step 2: State the Hypotheses

Null Hypothesis

$$H_0: \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$$

where μ_1 is the mean for the first group and μ_2 is the mean for the second group

Alternative Hypothesis

$$H_a: \mu_1 \neq \mu_2$$

Step 3: Compute the Test Statistic

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{se}$$

where $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$\overline{x_1}$: mean of the first group

$\overline{x_2}$: mean of the second group

s_1: standard deviation of the first group

s_2: standard deviation of the second group

n_1: sample size of the first group

n_2: sample size of the second group

Step 3: Compute the Test Statistic

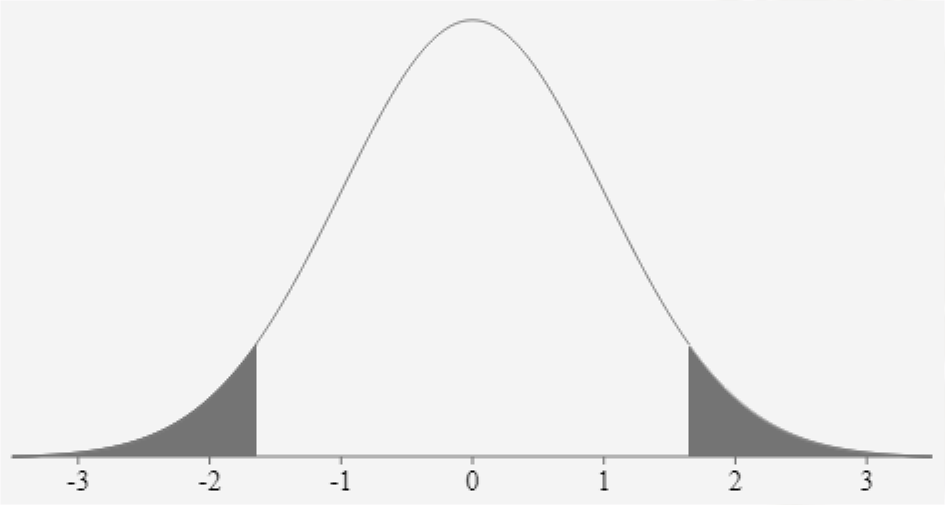
$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{2.2^2}{40} + \frac{3.0^2}{40}} = 0.5882$$

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{se} = \frac{(5.3 - 7.1) - 0}{0.5882} = -3.06$$

$$\overline{x_1} = 5.3$$
$$\overline{x_2} = 7.1$$
$$s_1 = 2.2$$
$$s_2 = 3.0$$
$$n_1 = 40$$
$$n_2 = 40$$

Step 4: Interpret the Test Statistic (Using Rejection Region)

- $\alpha = 0.1$
- $z_c = \pm 1.645$



Step 5: Make a Conclusion

- Since the test statistic lies in the RR
- then we reject the null hypothesis.
- Therefore, there is a difference between the means of the two groups.

Example:

A gardener sets up a flower stand in a busy business district and sells bouquets of assorted fresh flowers on weekdays. To find a more profitable pricing, she sells bouquets for Php15 each for ten days, then for Php10 each for five days. Her average daily profit for the two different prices are given below.

	n	$\bar{x}$	s
Php 15	10	171	26
Php 10	5	198	29

Test whether there is a difference between the means. Use a 10% level of significance.

Step 1: Assumptions

- Both sample sizes are small, $n_1, n_2 < 30$, use the t-table

- Equal variance: $df = n_1 + n_2 - 2$

- Unequal variance: $df = \dfrac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$

Step 2: State the Hypotheses

Null Hypothesis

$$H_0: \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$$

where μ_1 is the mean for the first group and μ_2 is the mean for the second group

Alternative Hypothesis

$$H_a: \mu_1 \neq \mu_2$$

Step 3: Compute the Test Statistic

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{26^2}{10} + \frac{29^2}{5}} = 15.3558$$

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{se} = \frac{(171 - 198) - 0}{15.3558}$$

$$= -1.76$$

Step 4: Interpret the Test Statistic (Using p-values)

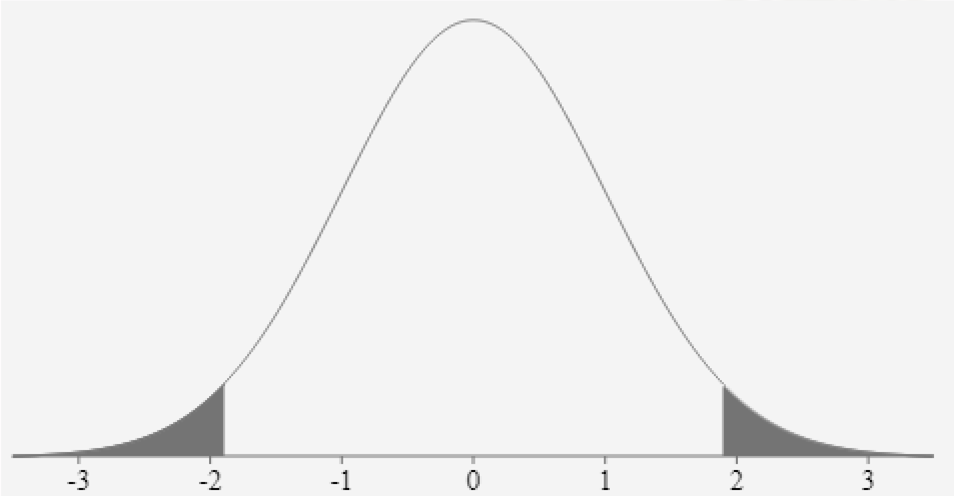
- $\alpha = 0.1$
- Use unequal variance,

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

$$= \frac{\left[\frac{26^2}{10} + \frac{29^2}{5}\right]^2}{\frac{1}{10 - 1}\left(\frac{26^2}{10}\right)^2 + \frac{1}{5 - 1}\left(\frac{29^2}{5}\right)^2} = 7.33 \approx 7$$

Step 4: Interpret the Test Statistic (Using rejection region)

- $\alpha = 0.1$ (two-tailed), $\frac{\alpha}{2} = 0.05$
- df=7
- $t_c = \pm 1.894579$



Step 5: Make a Conclusion

- Since the test statistic lies in the FTR
- then we do not reject the null hypothesis.
- Therefore, there is no difference between the means of the two groups.

Paired Samples

a significance test for comparing two means with paired samples, that is, sample observations are naturally paired. This is a result of a group being tested twice. Usually, this is used in determining the difference of means of before and after observations.

Paired Samples

1. Compute the difference between the observations for each sample. Denote it by x_d
2. Compute for the sample mean difference, $\bar{x}_d$ of the difference scores, x_d
3. Hypotheses
- To test the hypothesis $H_0: \mu_1 = \mu_2$ of equal means, conduct a one-sample test of $H_0: \mu_d = 0$ with the difference scores.
- $H_0: \mu_d = 0 (or\ \mu_d \geq 0, or\ \mu_d \leq 0)$, where $\mu_d = \mu_2 - \mu_1$
- $H_a: \mu_d \neq 0 (or\ \mu_d < 0, or\ \mu_d > 0)$
4. Test Statistic
- $t = \frac{\bar{x}_d - 0}{se}$ where $se = \frac{s_d}{\sqrt{n}}$ where s_d is the standard deviation of the samples x_d

5. Compute for the p – value. The degree of freedom (*df*)= n – 1. Or use the rejection method.

6. Based on the p-value or according to the rejection region, make a decision about H_0. Relate the conclusion to the context of the study.

Example:

- Eight golfers were asked to submit their latest scores on their favorite golf courses. These golfers were each given a set of newly designed clubs. After playing with the new clubs for a few months, the golfers were again asked to submit their latest scores on the same golf courses. The results are summarized below.

Golfer	1	2	3	4	5	6	7	8
Own Clubs	77	80	69	73	73	72	75	77
New Clubs	72	81	68	73	75	70	73	75

- Test, at the 1% level of significance, the hypothesis that on average golf scores are the same with the new clubs.

Example:

Golfer	1	2	3	4	5	6	7	8
Own Clubs	77	80	69	73	73	72	75	77
New Clubs	72	81	68	73	75	70	73	75
x_d	5	-1	1	0	-2	2	2	2

$$\overline{x_d} = 1.125$$
$$sd = 2.167124$$
$$se = 0.766194$$
$$t = 1.468296 \approx 1.47$$

Example:

- $df = 7;\ \alpha = .01;\ \frac{\alpha}{2} = .005$
- $t_c = 3.49948$
- Hence, the test statistic lies in the FTR. So, we fail to reject the null hypothesis, $\mu_d = 0, \mu_1 = \mu_2$.
- Therefore, there is no significant difference between the before and after or the old club and the new club.