13. Paired and Two-Sample Tests

Two-sample t-test

The two-sample test is used when one needs to compare means of two populations. To perform the test we take two *independent* samples from each population:

- $X_1, \dots, X_{n_1}$ from the first population with unknown mean μ_1
- $Y_1, \dots, Y_{n_2}$ from the second population with unknown mean μ_2

The questions that we can ask about μ_1 and μ_2 are:

- Is μ_1 is greater than μ_2?
- Is μ_1 is less than μ_2?
- Are μ_1 and μ_2 different?

Two-sample test: setup

The null hypothesis

$$H_0: \mu_1 = \mu_2$$

Possible alternatives:

$$H_a: \mu_1 > \mu_2$$
$$H_a: \mu_1 < \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

Which type of alternative will be used is determined by the problem context.

Two-sample test: test statistic

To test the research hypothesis we use the following test statistic

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

- $\bar{X}$ is the sample mean of the first sample, s_1^2 is the corresponding sample variance, and n_1 is the corresponding sample size
- $\bar{Y}$ is the sample mean of the second sample, s_2^2 is the corresponding sample variance, and n_2 is the corresponding sample size

Two-sample test: rejection region

Three possible expressions for the rejection region:

- **If we test $H_a\colon \mu_1 > \mu_2$, then $RR = [t_{\alpha,\nu}, +\infty]$**
- **If we test $H_a\colon \mu_1 < \mu_2$, then $RR = [-\infty, -t_{\alpha,\nu}]$**
- **If we test $H_a\colon \mu_1 \neq \mu_2$, then $RR = \left[-\infty, -t_{\alpha/2,\nu}\right] \cup [t_{\alpha/2,\nu}, +\infty]$**

Here

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \times \frac{1}{n_1 - 1} + \left(\frac{s_2^2}{n_2}\right)^2 \times \frac{1}{n_2 - 1}}$$

Two-sample test: rejection region

- If $n_1 < 30$, or $n_2 < 30$, or both, we *must* make sure that both sampled populations are approximately normal.
- If $n_1 \geq 30$ and $n_2 \geq 30$, the test statistic has approximately normal distribution and in the formulas for the rejection region we can use z-critical values instead of critical values of t-distribution.

Two-sample test: conclusion

As before, the decision depends on the relationship between the rejection region and the test statistic. The same universal rule is used.

- **If the test statistic falls *inside* the rejection region we *accept* H_a, the research claim.**
- **If the test statistic falls *outside* the rejection region we *reject* H_a.**

Two-sample test: a remark

Sometimes we might want to ask something like this: is it true that μ_1 is greater than μ_2 *by* D_0 units? This leads to the following natural modifications of the test.

- *The null hypothesis*

Instead of $H_0: \mu_1 = \mu_2$ we use $H_0: \mu_1 - \mu_2 = D_0$

- *The possible alternatives*

Instead of $H_a: \mu_1 > \mu_2$ we use $H_a: \mu_1 - \mu_2 > D_0$

Instead of $H_a: \mu_1 < \mu_2$ we use $H_a: \mu_1 - \mu_2 < D_0$

Instead of $H_a: \mu_1 \neq \mu_2$ we use $H_a: \mu_1 - \mu_2 \neq D_0$

- *Test statistic*

$$t = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Two-sample test: example

Example 1. A processor of recycled aluminum cans is concerned about the level of impurities (principally other metals) contained in lots from two sources. Laboratory analysis of sample lots yields the following data (kilograms of impurities per hundred kilogram of product):

$$n_1 = 12, \bar{X} = 3.267, s_1 = .676$$
$$n_2 = 12, \bar{Y} = 3.617, s_2 = 1.365$$

Can the processor conclude, using the confidence level of 5%, that there is a nonzero difference in means?

1. Setup

Let μ_1 and μ_2 be the true means of all the lots from the 1st and 2nd source, respectively.

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

Note that the sample sizes are small, therefore, we need to check normality assumption for two populations.

2. Test statistic

The test statistic is given by

$$\boldsymbol{t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.267 - 3.617}{\sqrt{\frac{.676^2}{12} + \frac{1.365^2}{12}}} \approx -.796}$$

3. Rejection Region

First, note that

$$\nu = \frac{\left(\frac{.676^2}{12} + \frac{1.365^2}{12}\right)^2}{\left(\frac{.676^2}{12}\right)^2 \times \frac{1}{11} + \left(\frac{1.365^2}{12}\right)^2 \times \frac{1}{11}} \approx 16$$

Then, since here we test a two-sided alternative, the RR is given by

$$\begin{aligned} RR &= \left[-\infty, -t_{\alpha/2,\nu}\right] \cup \left[t_{\alpha/2,\nu}, +\infty\right] \\ &= \left[-\infty, -t_{.025,16}\right] \cup \left[t_{.025,16}, +\infty\right] \\ &= [\infty, -2.12] \cup [2.12, +\infty] \end{aligned}$$

4. Conclusion

The test statistic does not fall into the rejection region. Therefore, we failed to reject the null at significance level of 5%, and we reject the alternative. That is, there is *no* difference between two sources.

Paired t-test

Consider a population that consist of *pairs of measurements* (x, y). Take a sample of n_D pairs:

$$(X_1, Y_1), \dots, (X_{n_D}, Y_{n_D})$$

The goal of the *paired test* is to compare the average of x-measurements with the average of y-measurements.

Examples:

- Population of husband-wife couples, height measurements for husband and wife
- Morning and bed time systolic blood pressure readings for male patients

Paired test vs two-sample test: warning

It is easy to confuse these two tests if in the two-sample situation the sample sizes, n_1 and n_2, are equal to each other.

We employ the paired test if we can show that each observation in the first sample is *coupled* (*matched, paired*) in a meaningful way with a corresponding observation in the second sample.

Paired t-test: how it is done

Instead of working with the sample of n_D pairs

$$(X_1, Y_1), \dots, (X_{n_D}, Y_{n_D})$$

we form a new sample of differences:

$$X_1 - Y_1, \dots, X_{n_D} - Y_{n_D}$$

and test a claim about the population *mean difference* μ_D. It is clear that

- $\mu_D > 0$ means that the x-average is greater than the y-average
- $\mu_D < 0$ means that the x-average is less than the y-average
- $\mu_D \neq 0$ means that the x-average and the y-average are different

Thus a paired t-test is just a regular one-sample t-test on the sample of differences.

Paired test: setup

Null hypothesis

$$H_0: \mu_D = 0$$

Possible alternatives:

$$H_a: \mu_D > 0$$

$$H_a: \mu_D < 0$$

$$H_a: \mu_D \neq 0$$

Which type of alternative will be used is determined by the problem context.

Paired test: test statistic

To test the research hypothesis we use the following test statistic

$$t = \frac{\bar{X}_D}{s_D/\sqrt{n_D}}$$

Here $\bar{X}_D$ is the sample mean of differences, s_D is the sample standard deviation of differences, and n_D is the numbers of pairs.

- If H_0 is true then one can show that random variable t has t-distribution with $n_D - 1$ degrees of freedom (when the population of differences is normal).
- If $n_D \geq 30$, the CLT tells us that the test statistic has normal distribution under H_0. In this case the normality assumption check is not needed.

Paired test about: rejection region

Three possible expressions for the rejection region:

- **If we test $H_a: \mu_D > 0$, then $RR = [t_{\alpha, n_D - 1}, \infty]$**
- **If we test $H_a: \mu_D < 0$, then $RR = [-\infty, -t_{\alpha, n_D - 1}]$**
- **If we test $H_a: \mu_D \neq 0$, then $RR = \left[-\infty, -t_{\alpha/2, n_D - 1}\right] \cup [t_{\alpha/2, n_D - 1}, \infty]$**

However, if $n_D \geq 30$, under H_0 the test statistics has approximately normal distribution and in the formulas for the rejection region we can use z-critical values instead of critical values of t-distribution.

Paired test: conclusion

As before, the decision depends on the relationship between the rejection region and the test statistic.

- If the test statistic falls *inside* the rejection region we *accept* H_a, the research claim.
- If the test statistic falls *outside* the rejection region we *reject* H_a.

Paired test: example

Example 2. A tasting panel of 15 people is asked to rate two new kinds of tea on a scale ranging from 0 to 100; 25 means “I would try to finish it only to be polite,” 50 means “I would drink it but not buy it,” 75 means “It’s about as good as any tea I know,” and 100 means “It’s superb; I would drink nothing else.” The difference in rating is recorded for each person. The mean of differences is 7, and the standard deviation is 16.08. Does these data indicate the difference in ratings (at 5 % significance level)?

1. Setup

Let μ_D will be true mean difference in ratings of two kinds of tea. Then

$$H_0: \mu_D = 0$$

$$H_a: \mu_D \neq 0$$

2. Test Statistic

The test statistic is

$$t = \frac{\bar{X}}{s_D/\sqrt{n_D}} = \frac{7}{16.08/\sqrt{15}} \approx 1.69$$

3. Rejection region

The rejection region is

$$
\begin{aligned}
\boldsymbol{RR} &= \left[-\infty, -\boldsymbol{t}_{\boldsymbol{\alpha/2, n_D - 1}}\right] \cup \left[\boldsymbol{t}_{\boldsymbol{\alpha/2, n_D - 1}}, +\infty\right] \\
&= \left[-\infty, -\boldsymbol{t}_{\boldsymbol{.025, 14}}\right] \cup \left[\boldsymbol{t}_{\boldsymbol{.025, 14}}, +\infty\right] \\
&= [-\infty, -\mathbf{2.145}] \cup [\mathbf{2.145}, +\infty]
\end{aligned}
$$

4. Conclusion

The test statistic does not fall into the rejection region. Therefore, we reject the alternative at significance level of 5%. That is, there is no difference in ratings of these two kinds of tea.

Exercises

Exercise 1. Business schools A and B reported the following summary of GMAT (Graduate Management Aptitude Test) verbal scores:

School	Sample size	Sample mean	Sample variance
A	21	34.75	48.6
B	15	30.25	30.7

At a 5% level of significance, is there sufficient evidence to believe there is a difference in the population means?

Exercises

Exercise 2. Twenty-four males age 25-29 were selected from the Framingham Heart Study. Twelve were smokers and 12 were nonsmokers. The subjects were *paired*, with one being a smoker (A) and the other a nonsmoker (B). Otherwise, each pair was similar with regard to age and physical characteristics. Systolic blood pressure readings were as follows:

A	122	146	120	114	124	126	118	128	130	134	116	130
B	114	134	114	116	138	110	112	116	132	126	108	116

Exercises

List the differences A-B and verify that $\bar{X}_D = 6$ and $s_D = 8.4$. Use a 5% level of significance to determine whether the data indicate a difference in mean systolic blood pressure levels for the populations from which the two groups were selected. You may assume that the population of differences is approximately normal.

Exercises

Exercise 3. A salesman for a shoe company claimed that runners would record quicker times, on the average, with the company's brand of sneaker. A track coach decided to test the claim. The coach selected eight runners. Each runner ran two 100-yard dashes on different days. In one 100-yard dash, the runners wore the sneakers supplied by the school (A); in the other, they wore the sneakers supplied by the salesman (B). Each runner was randomly assigned the sneakers to wear for the first run. Their times, measured in seconds, were as follows:

A	11.4	12.5	10.8	11.7	10.9	11.8	12.2	11.7
B	10.8	12.3	10.7	12.0	10.6	11.5	12.1	11.2

Note. For the differences, $\overline{X}_D = .225$ and $s_D = .276$. Assume the population of differences is approximately normal.