10.2 Hypothesis Test for Two Means When $\sigma_1 = \sigma_2$

Suppose we have two populations. We draw a random sample from each population. The sample sizes are, respectively, n_1 and n_2. Let μ_1 and μ_2 denote the true population means.

Suppose we want to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \underset{>}{\geq} \mu_2$ or $H_1 : \mu_1 \underset{<}{\leq} \mu_2$ or $H_1 : \mu_1 \neq \mu_2$.

We assume that the true population standard deviations σ_1 and σ_2 are estimated from the data by the sample standard deviations s_1 and s_2, respectively.

Two cases are distinguished:

- When it is assumed that the population standard deviations are equal, that is, $\sigma_1 = \sigma_2$.

- When it is assumed that the population standard deviations are <u>not</u> equal, that is, $\sigma_1 \neq \sigma_2$.

Now we will study the test when it is assumed that the population standard deviations are equal, that is, $\sigma_1 = \sigma_2$.

We are given $n_1, \bar{x}_1, s_1, n_2, \bar{x}_2, s_2$, and α.

The first step is to estimate the common standard deviation by pooling two samples together. The **pooled estimate** is computed as

$$s_{\mathrm{p}} = \sqrt{\frac{\sum (x-\bar{x}_1)^2 + \sum (x-\bar{x}_2)^2}{n_1 - 1 + n_2 - 1}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}.$$

Next, we compute the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\mathrm{p}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

which, under H_0, has a *t*-distribution with $df = n_1 + n_2 - 2$.

Denote by t_0 the observed value of the test statistic. The *P*-value for this test is defined as

- $P(t \geq t_0)$ if $H_1 : \mu_1 \geq \mu_2$
- $P(t \leq t_0)$ if $H_1 : \mu_1 \leq \mu_2$
- $2P(t \leq -|t_0|)$ if $H_1 : \mu_1 \neq \mu_2$.

Finally, we use the *t*-table to compare the *P*-value to α, and then we make a decision and draw conclusion.

Example. A hotel manager wants to check whether advertisement increases average hotel stay. He picks a random sample of 10 visitors before the advertisement takes place, and finds that the mean stay is 2.2 nights with the standard deviation of 0.9 nights. He also selects a random sample of 10 visitors after the advertisement takes place and finds that the mean stay for this sample is 2.9 nights and the standard deviation is 1.1 nights. Test the claim at the 5% significance level. Assume $\sigma_1 = \sigma_2$.

Solution. We are given that $n_1 = n_2 = 10$, $\bar{x}_1 = 2.2, \bar{x}_2 = 2.9, s_1 = 0.9, s_2 = 1.1,$ and $\alpha = 0.05.$ We want to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$. We compute the pooled standard deviation

$$s_\mathrm{p} = \sqrt{\frac{(10-1)(0.9)^2 + (10-1)(1.1)^2}{10+10-2}} = 1.005.$$

The test statistic is $t = \dfrac{2.2 - 2.9}{(1.005)\sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = -1.557.$

The P-value = $P(t < -1.557) = P(t > 1.557)$.

Now we use the t-table to find bounds for the P-value. We see that for $df = 10 + 10 - 2 = 18$, $1.330 < 1.557 < 1.734$, and, hence, $0.1 > P(t > 1.557) > 0.05 = \alpha$.

Therefore, we accept H_0, and conclude that the advertisement didn't increase an average stay in the hotel.

Example. Cholesterol levels are measured for 28 heart attack patients (2 days after their attacks) and 30 other hospital patients who did not have a heart attack (control group). The sample quantities are:

	Mean	StDev
Attack	233.7	56.3
Control	184.2	49.8

Test at $\alpha = 0.001$ that mean cholesterol level is higher for the heart attack patients, assuming equal standard deviations.

Solution. We have $n_1 = 28, n_2 = 30, \bar{x}_1 = 233.7,$ $\bar{x}_2 = 184.2, s_1 = 56.3, s_2 = 49.8,$ and $\alpha = 0.05.$ We want to test $H_0: \mu_1 = \mu_2$ against H_1: $\mu_1 > \mu_2$. We compute the pooled standard deviation

$$s_p = \sqrt{\frac{(28-1)(56.3)^2 + (30-1)(49.8)^2}{28+30-2}} = 53.03.$$

The test statistic is $t = \dfrac{233.7 - 184.2}{(53.03)\sqrt{\dfrac{1}{28} + \dfrac{1}{30}}} = 3.55.$

The degrees of freedom *df=28+30-2=56*.
We will use *df=60*.

P-value = $P(t > 3.55) < 0.0005 < 0.001 = \alpha,$

therefore, we reject $H_0,$ and conclude that mean cholesterol level is higher for the heart attack patients.

Example. A company is interested in finding out whether mean customer satisfaction scores are the same for two stores owned by this company. The data for two random samples of sizes 5 and 7 are available, with respective sample means of 30 and 24, and sample standard deviations of 3.4 and 5.1. Carry out the test at the 5% level of significance. Assume $\sigma_1 = \sigma_2$.

<u>Solution</u>. Given $n_1 = 5, \bar{x}_1 = 30, s_1 = 3.4, n_2 = 7, \bar{x}_2 = 24,$ $s_2 = 5.1,$ and $\alpha = 0.05.$ We want to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2.$

The pooled standard deviation is

$$s_{\mathrm{p}} = \sqrt{\frac{(5-1)(3.4)^2 + (7-1)(5.1)^2}{5+7-2}} = 4.498.$$

The test statistic is

$$t = \frac{30-24}{(4.498)\sqrt{\frac{1}{5} + \frac{1}{7}}} = 2.278.$$

The P-value = $2P(t < -2.278) = 2P(t > 2.278)$.

From the t-table, for $df = 5 + 7 - 2 = 10$,

2.228< 2.278<2.764, and, thus

$$\alpha = 0.05 > 2P(t > 2.278) > 0.02.$$

Hence, we reject the null, and conclude that the mean satisfaction scores for the two stores are different.