10.3 Hypothesis Test for Two Means When $\sigma_1 \neq \sigma_2$

Suppose we want to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \geq \mu_2$ or $H_1 : \mu_1 \leq \mu_2$ or $H_1 : \mu_1 \neq \mu_2$.

We studied the test when we assumed that $\sigma_1 = \sigma_2$. Now we study the case when $\sigma_1 \neq \sigma_2$.

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

The test statistic is given by the formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

which, under H_0, has a *t*-distribution with the number of degrees of freedom *df* defined as the largest integer such that

$$df \leq \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}.$$

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 random sample of 5 male customers at a clothing store showed that they spent on average \$89 with a standard deviation of \$27.50. Another random sample of 6 female customers revealed the sample mean of \$104 with the standard deviation of \$26.60. Can we conclude that the average amount spent by males doesn't exceed that spent by females? Assume that $\sigma_1 \neq \sigma_2$.

Solution. We want to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \le \mu_2$. The test statistic is

$$t = \frac{89 - 104}{\sqrt{\frac{(27.5)^2}{5} + \frac{(26.6)^2}{6}}} = -0.914.$$

The number of degrees of freedom is the largest integer such that

$$df \le \frac{\left(\frac{27.5^2}{5}+\frac{26.6^2}{6}\right)^2}{\frac{\left(\frac{27.5^2}{5}\right)^2}{5-1}+\frac{\left(\frac{26.6^2}{6}\right)^2}{6-1}} = 8.52, \text{ so } \; df = 8.$$

The P-value =

$$P(t < -0.914) = P(t > 0.914) > 0.15 > 0.05 = \alpha.$$

Therefore, we accept the null, and conclude that there is no supporting evidence that the average amount spent by males doesn't exceed that spent by females.

Example. A sample of 50 seniors and a sample of 40 freshmen were surveyed, and it was found that the mean time seniors spend studying for a final exam is 17 hours with the standard deviation of 6 hours, while freshmen spend on average 14 hours with the standard deviation of 9 hours. Test whether seniors spend on average more hours studying than freshmen. Assume $\sigma_1 \neq \sigma_2$.

Solution. We want to test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$.

The test statistic is

$$t = \frac{17-14}{\sqrt{\frac{(6)^2}{50}+\frac{(9)^2}{40}}} = 1.811.$$

The number of degrees of freedom is the largest integer such that

$$df \le \frac{\left(\frac{6^2}{50}+\frac{9^2}{40}\right)^2}{\frac{\left(\frac{6^2}{50}\right)^2}{50-1}+\frac{\left(\frac{9^2}{40}\right)^2}{40-1}} = 65.1,\ \text{so}\quad df = 65.$$

In the *t*-table we take the closest value *df*=60.

The P-value = $P(t > 1.811) < 0.05 = \alpha$.

Hence, we reject H_0, and conclude that the average number of hours seniors spent studying for final is larger than that for freshmen.

Example. A study found that the mean number of children under 18 per household in community A was 1.6 with a standard deviation of 0.7. In community B the mean was 2.1 with a standard deviation 1.3. The data were based on two independent samples of sizes 155 and 160, respectively. A statistician is trying to determine whether the population means are different. There is no reason to believe that the population standard deviations are the same. Conduct the test and draw conclusion. Assume that the probability of type I error is 0.025.

Solution. We want to test $H_0 : \mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$.

The test statistic is

$$t = \frac{1.6 - 2.1}{\sqrt{\frac{(0.7)^2}{155} + \frac{(1.3)^2}{160}}} = -4.268.$$

The number of degrees of freedom is the largest integer such that

$$\mathrm{df} \le \frac{\left(\frac{0.7^2}{155} + \frac{1.3^2}{160}\right)^2}{\frac{\left(\frac{0.7^2}{155}\right)^2}{155-1} + \frac{\left(\frac{1.3^2}{160}\right)^2}{160-1}} = 245.695, \quad so \quad \mathrm{df} = 245.$$

In the *t*-table we take the closest value *df*=100.

The P-value = $2P(t \leq -4.268) = 2P(t \geq 4.268)$
$< 0.001 < 0.025 = \alpha$.

Hence, we reject H_0, and conclude that the population mean number of children is different for these two communities.