Hypothesis testing for the difference of two population means

A study is conducted to assess the differences in performance during the first years of services between employees that stayed in a certain company during 15 years and those who left the company. The performance is measured by the company’s annual performance appraisal which produce ratings on a 5 point scale; 1 for low performance and 5 for high performance.

The data are summarized in the table:

<table>
<thead>
<tr>
<th></th>
<th>Stayers</th>
<th>Leavers</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>174</td>
<td>355</td>
</tr>
<tr>
<td>(\bar{y}_1)</td>
<td>3.51</td>
<td>3.24</td>
</tr>
<tr>
<td>(s_1)</td>
<td>0.51</td>
<td>0.52</td>
</tr>
</tbody>
</table>

To consider the previous problem in a hypothesis testing framework we assume that the group of stayers has a population mean equal to \(\mu_1\) and the group of leavers has a population mean of \(\mu_2\). Then

- \(H_0 : \mu_1 = \mu_2\) vs \(H_1 : \mu_1 > \mu_2\) one-tailed
- \(H_0 : \mu_1 = \mu_2\) vs \(H_1 : \mu_1 \neq \mu_2\) two-tailed

The test statistics is given by

\[
t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}
\]

where \(s_p^2\) is the pooled sample variance

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

calculated assuming that \(\sigma_1^2 = \sigma_2^2\).

The rejection region is given by

- \(t < -t_\alpha\) one-tailed
- \(|t| > t_{\alpha/2}\) two-tailed

where \(t_\alpha\) is the \(\alpha\)-quantile of a student \(t\) distribution with \((n_1 + n_2 - 2)\) degrees of freedom.

A confidence interval is given by

\[
(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}
\]

In our previous example we have that

\[
s_p^2 = \frac{173 \times 0.51 + 354 \times 0.52}{527} = 0.5167 \quad \text{and} \quad t = 4.0588
\]

For 527 DF we have that \(t_{0.005} = 2.58\) so a one-tailed test would have produced the same conclusion. The p-value of this test is in the order of \(10^{-5}\).

A confidence 99% confidence interval is given by

\[
(3.51 - 3.24) \pm 2.58 \times 0.0665 = 0.27 \pm 0.1715 = (0.0985, 0.4415)
\]

Notice that, given the large number of samples involved in the study, a test based on the normal distribution would have given very similar results.

**Warning:** you have to specify if your test is a two-tailed or a one-tailed test beforehand.
Comparing two population variances

When comparing two normal populations we can be interested in differences between the variances of them and not just the means. Consider the following example. The data in the table correspond to the amount of ethylene oxide (ETO) (in mmgrms) measured in the bloodstream of 30 subjects manipulating sterilizing hospital supplies. 11 subjects where randomly assigned to one task and 19 to another. Is there enough evidence in the data to conclude that there are differences in the variability of the ETO levels between the two groups?

<table>
<thead>
<tr>
<th>task 1</th>
<th>task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>11</td>
</tr>
<tr>
<td>Mean</td>
<td>5.60</td>
</tr>
<tr>
<td>SD</td>
<td>4.10</td>
</tr>
</tbody>
</table>

For our example, assume a significant level of 5%, then \( F_\alpha = 2.79 \). The observed ratio is \( F = 4.51 > 2.79 \) so we reject the null hypothesis and conclude that there is a significant difference in the variances of the two groups.

Notice that we have only stated the one-tailed test. Given that the \( F \) distribution is not a symmetric distribution, splitting the probability in \( \alpha \) in two equal halves is not as obvious as in the mean case.

The textbook proposes to calculate the ratio by always taking as the numerator the highest population variance and dividing the level by two. We shall follow that approach.

The hypothesis testing problem can be stated as

\[
H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad \sigma_1^2 > \sigma_2^2
\]

The test statistics in this case is

\[
F = \frac{s_1^2}{s_2^2}
\]

\( s_i \) represents the sample variance. The rejection region is given by

\[ F > F_\alpha \]

where \( F_\alpha \) is such that

\[
Pr(X > F_\alpha) = \alpha
\]

where \( X \) is a random variable following a Fisher’s \( F \) distribution with \( n_1 - 1 \) degrees of freedom in the numerator and \( n_2 - 1 \) in the denominator.

Comparing the means when the variances are not equal

Consider the comparison of two population means

\[
H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2 \quad \text{one-tailed}
\]

\[
H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2 \quad \text{two-tailed}
\]

When we can not assume that the variances of the two populations are equal the test statistics is given by

\[
t = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}
\]

And the degrees of freedom of the student are calculated as

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

To investigate the effect of amphetamines on water consumption, 15 lab rats were injected with amphetamine and 10 with saline solution. The water consumed by each rat in ml/kg of body weight were recorded and the results are summarized in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Amphetamine</th>
<th>Saline</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>115</td>
<td>135</td>
</tr>
<tr>
<td>$s$</td>
<td>40</td>
<td>15</td>
</tr>
</tbody>
</table>

To compare the two variances, we use an $F$ test.

$F = \frac{40^2}{15^2} = 7.11$. From the table we obtain $F_{14.9}^{0.975} \approx 3.87$. Since $7.11 > 3.87$ we reject the null hypothesis, consisting of the two variances being equal.

To perform the $t$ test we calculate the degrees of freedom

$$\nu = \frac{\left(\frac{1600}{15}\right)^2 + \frac{225}{10}}{\left(\frac{1600}{15-1}\right)^2 + \left(\frac{225}{10-1}\right)^2} = 19.2 < 23 = n_A + n_S - 2$$

We test the hypothesis

$$H_0 : \mu_A \geq \mu_S \text{ vs } H_1 : \mu_A < \mu_S$$

using the test statistics

$$t = \frac{115 - 135}{\sqrt{\frac{1600}{15} + \frac{225}{10}}} = -1.759$$

Letting $\alpha = 0.05$ we find that the critical value is $t_{1.05}^{0.9} = -1.729$. Since $-1.759 < -1.729$ we reject the null hypothesis.