The T-statistic is used to test differences in the means of two groups. The grouping variable is categorical and data for the dependent variable is interval scaled. The following table shows alternative statistical techniques that can be used to analyze this type of data when different levels of measurement are available.

The t-distribution was developed by W. S. Gosset (1908)\(^1\). As an employee of the Guinness brewery in Dublin, Gosset was not permitted to publish research findings in his own name, and hence penned the pseudonym "Student". The t-distribution, as it was first designated, has been known under a variety of names, including the Student's distribution and Student’s t-distribution.

The t-distribution revolutionized statistics and the ability to work with small samples. Prior to this time, statistical work was based largely on the value of \( z \), which was used to designate a point on the normal distribution where population parameters were known. The \( z \) value is the deviation of the sample mean from the mean of the population and is expressed in terms of variance within a normally distributed population.\(^2\)

<table>
<thead>
<tr>
<th>Sample Cases</th>
<th>Level of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td><strong>Ordinal</strong></td>
</tr>
<tr>
<td>Two-sample related samples</td>
<td>McNemar test for the</td>
</tr>
<tr>
<td>Independent samples</td>
<td>significance of changes</td>
</tr>
<tr>
<td></td>
<td>Fischer exact</td>
</tr>
<tr>
<td></td>
<td>probability test</td>
</tr>
<tr>
<td></td>
<td>( \chi^2 ) test for</td>
</tr>
<tr>
<td></td>
<td>two independent samples</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The purpose of the \( z \) value is to express the amount of deviation between the sample mean and the population mean and to permit the making of inferences as to whether the sample mean belongs to the population in question. The mean and variance characteristics of the population (\( \mu \) and \( \sigma \)), to which we desire to make inferences, are rarely known. This state of perfect knowledge is the assumption made by the \( z \) test and in actual use is difficult to justify. The \( t \) statistic does not require the population variance information needed for the \( z \) test, but instead uses the sample variance (and sample standard deviation).

\[
t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}
\]

The \( t \)-distribution is symmetrical about the mean and is approximately normal. It is centered at the population mean of 0 and for large samples has a variance \( \sigma = 1 \).

\(^1\) November 1, 2011 Version: This tutorial is edited from the BIOMED statistical package t-test program, as developed under a National Science Foundation grant.

\(^2\) Qualtrics.com, Selected Nonparametric and Parametric Statistical Tests for Two-Sample Cases.
The central limit theorem tells us that the sampling distribution of all possible sample means \( \bar{x} \), approaches normality as the size of the samples increase. This is true, even when the population is not normally distributed. The t-distribution rapidly approaches the shape of a normal distribution as sample size increases. The t-distribution is considered normal when \( n=30 \).

The t-distribution is used to make inferences concerning the difference between the two populations \( \mu_1 \) and \( \mu_2 \). The specific statistical theory used relates to the distribution of differences between the two sets of independent sample means, the sampling distribution of \( \bar{x}_1 - \bar{x}_2 \).

**MATHEMATICAL COMPUTATIONS FOR THE T-TEST**

Statistical analysis programs compute t-statistics and associated probability levels for the equality of the means of two groups based on pooled and separate variance estimates. An F-statistic and associated probability level for the equality of group variances is also computed. Groups may be defined by specifying codes to be included. Several dependent variables may often be analyzed concurrently. Paired comparison t-ratios may be obtained through the use of IF and RECODE commands (SPSS).

Typical t-test computations include:
1. F-ratio of variance
2. t-value (based on pooled variance estimate)
3. t-value (based on separate variance estimate)
4. Two-tailed probability levels for each t and for the F
5. Means
6. Standard deviations
7. Standard error of the means
8. Number of observations included in computing 5-7 above

In computing the variance, it is necessary to pool variance estimates when for the two groups, sample sizes are unequal and variances are unequal (Levine’s test or O’Brien’s test). In case of pooling, we pool the point estimate by simple averages. The pooled standard error is given by:

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

**COMPUTATIONAL PROCEDURE**

Each problem is divided into two groups: an X and a Y category. For each analysis, the number of non-missing observations, the mean, standard deviation, and standard error are computed for each variable of each category. The t-values, F-values, and corresponding probability level for between-category comparison are computed for each variable.
STEP 1

\[ X_{ijk} \], where \( i=1,2,\ldots,n \)
\[ j=1,2,\ldots,(p+q) \]
\[ k=1,2;1=x \text{ category};2=y \text{ category} \]

STEP 2: Mean Computation

\[ \bar{X}_{jk} = \frac{\sum x_{ijk}}{n_{jk}} \]

Variance

\[ S_{jk}^2 = \left( \frac{1}{n_{jk} - 1} \right) \times \left( \frac{\sum x_{ijk}^2}{n_{jk}} - \left( \frac{\sum x_{ijk}}{n_{jk}} \right)^2 \right) \]

Standard Deviation

\[ S_{jk} = \sqrt{S_{jk}^2} \]

Standard error (of Mean)

\[ SE_{jk} = \frac{S_{jk}}{\sqrt{n_{jk}}} \]

\[ F_j = \frac{S_{j1}^2}{S_{j2}^2} \quad \text{if} \quad S_{j1}^2 \geq S_{j2}^2 \]

\[ F_j = \frac{S_{j2}^2}{S_{j1}^2} \quad \text{if} \quad S_{j1}^2 < S_{j2}^2 \]

Degrees of freedom based on pooled variance estimate:

\[ D_p = \max (n_{i1}-1,0) + \max (n_{i2}-1,0) \]

\[ t \] based on pooled variance estimate:

\[ t = \frac{X_1 - X_2}{\sqrt{\left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_1^2 (n_1 - 1) + \frac{S_2^2(n_2 - 1)}{n_1} - 1 + n_2 - 1}} \]
t based on separate variance estimate:

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

Degrees of freedom based on separate variance estimate:

\[ D = \frac{1}{\left(\frac{1}{n_{j1} - 1}\right) \left(\frac{1}{(SE_{j1})^2 + (SE_{j2})^2}\right) + \left(\frac{1}{n_{j2} - 1}\right) \left(\frac{1}{(SE_{j1})^2 + (SE_{j2})^2}\right)} \]

- In case of pooling, the pooled point estimate and pooled standard error are calculated. In case of the separate variance estimate, the difference between the two groups is used as the data. The mean and standard error of the difference are computed.

- F-value is the ratio of the mean square error (mean of the sum of squares) for each of the two groups.

- The P-value is the significance level corresponding to the computed F-value. If this P-value is less than or equal to the significance level of the test, then we reject the null hypothesis.

- The two groups are defined by the cut-off value selected. The cases are assigned to the two groups as per the cut-off value.

Footnotes

Student (1908), "The Probable Error of a Mean", Biometrika, 6:1.

The normal probability distribution is defined by the equation:

\[ f(x) = \frac{1}{2\sigma} e^{-\left(\frac{1}{2}\right)\left[\frac{(x-\mu)}{\sigma}\right]^2} \quad \text{and} \quad s_x = \frac{\Sigma(x - \bar{x})^2}{\sqrt{n - 1}} \]
We’re Here to Help!

Qualtrics.com provides the most advanced online survey building, data collection (via panels or corporate / personal contacts), real-time view of survey results, and advanced “dashboard reporting tools”.

If you are interested in learning more about how the Qualtrics professional services team can help you with a conjoint analysis research project, contact us at research@qualtrics.com.