

Independent t-test -- Analysis of 2-Between-Group Data with a Quantitative Response Variable

Application: Compare the means of a quantitative variable obtained from 2 independent groups.

H0: The two populations represented by the two conditions of the qualitative variable have the same mean scores on the quantitative variable (often called the response variable).

To reject H0: is to say that the two populations have different means on the quantitative response variable.

The data: This analysis involves the qualitative variable **reptdept** (1 = not separate reptile department, 2 = separate department), and the quantitative (response) variable **reptgood** (rating of reptile quality - 1-10 scale) . Below are the scores for the 12 stores (reptdept, reptgood).

1,2 2,8 2,9 2,7 1,4 1,7 2,4 1,4 1,5 2,9 2,7 1,2

Research Hypothesis: The researcher hypothesized that stores with separate reptile departments would have reptiles of better overall quality than stores which did not have separate reptile departments.

H0: for this analysis: Pet shops that do not have separate reptile departments have the same mean reptile quality ratings as shops that do have separate reptile departments.

Assemble the data for analysis. Rearrange the data so that the data from each group (defined by the qualitative variable) in a separate column. Label the quantitative variable as X, to simplify the presentation and use of the formulas below. (Please note: These formulas can also be applied when there are different numbers of subjects in each group.)

not separate dept.

X

2

4

7

4

5

2

separate dept.

X

8

9

7

4

9

7

Compute the square of each score and place it in an adjacent column.

not separate dept.

X

X²

2

4

4

16

7

49

4

16

5

25

2

4

separate dept.

X

X²

8

64

9

81

7

49

4

16

9

81

7

49

Compute $\sum X$, $\sum X^2$, the mean and sum of squares for each group (following the formulas shown earlier in the Mean and Standard Deviation section) and complete the following table.

	not separate dept.	separate dept.
$\sum X$	24	44
$\sum X^2$	114	340
\bar{X}	4.0	7.33
SS	18.0	17.33

Compute the mean, SS, and std for each condition.

Computational examples using data from not separate department condition:

$$\text{Mean} = \frac{\sum X_{k1}}{n_{k1}} = \frac{24}{6} = 4.0 \quad \text{SS} = \sum X_{k1}^2 - \frac{(\sum X_{k1})^2}{n_{k1}} = 340 - \frac{44^2}{6} = 17.33$$

$$\text{Standard deviation} = \sqrt{\frac{\text{SS}}{n_{k1}-1}} = \sqrt{\frac{17.33}{6-1}} = 1.86$$

Compute the pooled error variance

$$\text{pooled } s^2 = \frac{\text{SS}_1 + \text{SS}_2}{n_1 + n_2 - 2} = \frac{18.00 + 17.33}{6 + 6 - 2} = \frac{35.33}{10} = 3.53$$

Compute the standard error of the difference between two independent means (Standard Error of Mean Differences - SEMD)

$$\text{SEMD} = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} = \sqrt{\frac{3.53}{6} + \frac{3.53}{6}} = \sqrt{(.59 + .59)} = \sqrt{1.18} = 1.09$$

Compute t

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SEMD}} = \frac{4.0 - 7.33}{1.09} = \frac{-3.33}{1.09} = -3.06$$

Determine the degrees of freedom for the t-test

$$df = n_1 + n_2 - 2 = 6 + 6 - 2 = 10$$

Use Table t to determine the critical value of t, for a two-tailed test with $\alpha = .05$ and the appropriate degrees of freedom.

Sometimes, with larger df, the table doesn't include the df you are looking for. When this happens, just use the next smaller df that is included on the table. For example, if you had df = 33, you would use the t for df = 30.

$$t(10, \alpha = .05) = 2.23$$

Compare the obtained t and critical t values and determine whether to retain or reject the null hypothesis (that there is no mean difference on the response variable between the populations represented by the groups. Remember that t values can be positive or negative, and so we will compare the **absolute value** of the obtained t to the critical t.

-- if the **absolute value** of the obtained t is less than the critical t then retain the null hypothesis and conclude that there is no mean difference on the response variable between the populations represented by the conditions.

-- if the **absolute value** of the obtained t is greater than the critical t , then reject the null hypothesis and conclude that there is a mean difference on the response variable between the populations represented by the conditions.

For the example data, we would decide to reject the null hypothesis, because the absolute value of the obtained t (3.06) is larger than the critical t (2.23).

IF you reject the null hypothesis, determine whether the data support or do not support the research hypothesis.

-- IF you reject the null hypothesis AND the direction of the mean difference between the conditions agrees with the direction of the research hypothesis, then the research hypothesis is supported.

-- IF you retain the null hypothesis OR if you reject the null BUT the direction of the mean difference disagrees with the direction of the research hypothesis, then the research hypothesis is not supported.

For the example data, we would decide that the research hypothesis is supported, because we rejected the null hypothesis, and because, as hypothesized, the stores with separate reptile departments had higher mean reptile quality than the stores without separate reptile departments.

By the Way: Usually the researcher hypothesizes that there is a difference between the conditions of the qualitative grouping variable. Sometimes, however, the research hypothesis is that there is NO mean difference between the conditions. If so, the research hypothesis and H_0 are the same! When this is the case, retaining H_0 provides support for the research hypothesis, whereas rejecting H_0 provides evidence that research hypothesis is incorrect.

Reporting the results of the independent samples t -test -- be sure to include the following

- Tell the conditions of the qualitative grouping variable
- Name quantitative variable and tell its mean and standard deviation for each group (defined by the qualitative variable)
- The t -value, df (in parentheses) and p -value ($p < .05$ or $p > .05$).
- Describe the pattern of the data (which group has the larger mean, if there is a significant difference)
- Whether or not the results support the research hypothesis

Those stores without separate reptile departments displayed reptiles with a mean quality rating of 4.0 ($S = 1.90$), whereas those that did have separate departments had a mean rating of 7.33 ($S = 1.86$). As hypothesized, pet stores with separate reptile departments had significantly higher mean ratings than those without separate departments, $t(10) = -3.06$, $p < .05$.