**Introduction**

The independent-samples t-test is used to determine if a difference exists between the means of two independent groups on a continuous dependent variable. More specifically, it will let you determine whether the difference between these two groups is statistically significant. This test is also known by a number of different names, including the independent t-test, independent-measures t-test, between-subjects t-test and unpaired t-test.

For example, you could use the independent-samples t-test to determine whether (mean) salaries, measured in US dollars, differed between males and females (i.e., your dependent variable would be "salary" and your independent variable would be "gender", which has two groups: "males" and "females"). You could also use an independent-samples t-test to determine whether (mean) reaction time, measured in milliseconds, differed in under 21 year olds versus those 21 years old and over (i.e., your dependent variable would be "reaction time" and your independent variable would be "age group", split into two groups: "under 21 year olds" and "21 years old and over").

## Null and alternative hypothesis

The null hypothesis is:

H0: µ = µ Scores are equal in the population. There is no difference between Groups/Conditions.

The alternative hypothesis is:

HA: µ ≠ µ Scores are not equal in the population. There is a difference between Groups/Conditions.

**Assumptions of the independent-samples t-test: Getting started**

When you choose to analyze your data using an independent-samples t-test, a critical part of the process involves checking to make sure that the data you want to analyze can actually be analyzed using this test. In fact, the independent-samples t-test has six assumptions that you have to consider. The first three of these assumptions are: (1) you have a continuous dependent variable; (2) your independent variable is categorical with two groups; and (3) you have independence of observations. If your study design does not meet these three assumptions, the independent-samples t-test is the incorrect statistical test to use to analyze your data.

The other three assumptions relate to the nature of your data and can be tested using SPSS. Since it is not uncommon for the data you have collected to violate (i.e., fail) one or more of these three assumptions, we show you different ways to proceed. This could include (a) making corrections to your data so that it no longer violates the assumptions, (b) using an alternative statistical test, or (c) proceeding with your analysis even when your data violates certain assumptions.

**Assumption #4**

There should be no significant outliers in the two groups of your independent variable in terms of the dependent variable. For both groups of the independent variable, if there are any scores that are unusual for that group, in that their value is extremely small or large compared to the other scores, these scores are called outliers (e.g., 8 participants in a group scored between 60-75 out of 100 in a difficult math test, but one participant scored 98 out of 100). Outliers can have a large negative effect on your results because they can exert a large influence (i.e., change) on the mean and standard deviation for that group, which can affect the statistical test results. Outliers are more important to consider when you have smaller sample sizes, as the effect of the outlier will be greater. Therefore, in this example, you need to investigate whether engagement has no outliers for each group of gender (i.e., you are testing whether the engagement score is outlier free for both the "Male" and "Female" groups). Due to the effect that outliers can have on your results, you have to choose whether to include these in your data when performing an independent-samples t-test in SPSS.

**Assumption #5**

Your dependent variable should be approximately normally distributed for each group of the independent variable. The assumption of normality is necessary for statistical significance testing using an independent-samples t-test. However, the independent-samples t-test is considered "robust" to violations of normality. This means that some violation of this assumption can be tolerated and the test will still provide valid results. Therefore, you will often hear of this test only requiring *approximately* normal data. Furthermore, as sample size increases, the distribution can be very non-normal and, thanks to the Central Limit Theorem, the independent-samples t-test can still provide valid results. Also, it should be noted that if the distributions are all skewed in a similar manner (e.g., all moderately negatively skewed), this is not as troublesome as when compared to the situation where you have groups that have differently-shaped distributions (e.g., the distribution of Group A is moderately 'positively' skewed, whilst the distribution of Group B is moderately 'negatively' skewed). Therefore, in this example, you need to investigate whether engagement is normally distributed. There are many different methods available to test this assumption. Other methods include skewness and kurtosis values and histograms.

Assumption #6

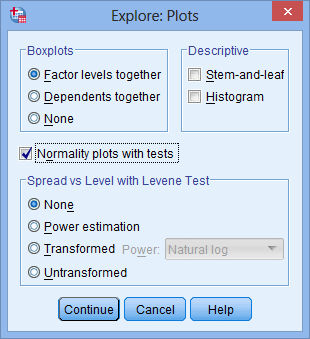
You have homogeneity of variances (i.e., the variance is equal in each group of your independent variable). The assumption of homogeneity of variances states that the population variance for each group of your independent variable is the same. If the sample size in each group is similar, violation of this assumption is not often too serious. However, if sample sizes are quite different, the independent-samples t-test is sensitive to the violation of this assumption. Either way, SPSS uses **Levene's test of equality of variances** and two differently-calculated independent-samples t-tests, which will give you a valid result irrespective of whether you met or violated this assumption (i.e., SPSS provides an independent-samples t-test that is calculated normally (with pooled variances) and another for when the assumption is violated that uses separate variances (i.e., non-pooled variances) and the Welch-Satterthwaite correction to the degrees of freedom). The Levene's test for equality of variances is automatically run at the same time as the main independent-samples t-test procedure.

**Example**

An Advertising Agency is commissioned to create a TV advert to promote a new product. Since the product is designed for men and women, the TV advert has to appeal to men and women equally. Before the company that commissioned the Advertising Agency spends $250,000 across a number of TV networks, it wants to make sure that the TV advert created by the Advertising Agency appeals equally to men and women. More specifically, the company wants to know whether the way that men and women engage with the TV advert is the same. To achieve this, the TV advert is shown to 20 men and 20 women, who are then asked to fill in a questionnaire that measures their engagement with the advertisement. The questionnaire provides an overall engagement score.

This overall engagement score is the dependent variable, which we have labelled engagement in SPSS Statistics. Our independent variable, which we have labelled gender in SPSS Statistics, contains two groups: "Male" and "Female". In variable terms, the Advertising Agency would like to know if the independent variable, gender, has an effect on the dependent variable, engagement (or expressed another way, if there are differences in engagement between levels of gender). In other words, is the mean engagement score different for males and females? Since the Advertising Agency needs the advertisement to be similarly engaging, they hope there is no difference!

The following instructions show you how to run tests to detect outliers and check if your data is normally distributed:

1. Click **Analyze > Descriptive Statistics > Explore.**

**2.** Transfer your variables into the Dependent List: box by clicking on them while holding down the shift-key, and then clicking the top https://statistics.laerd.com/premium/rma/img/right-arrow-button.pngbutton.

**3.** Click the https://statistics.laerd.com/premium/rma/img/plots-button.pngbutton and you will be presented with the **Explore: Plots** dialogue box.

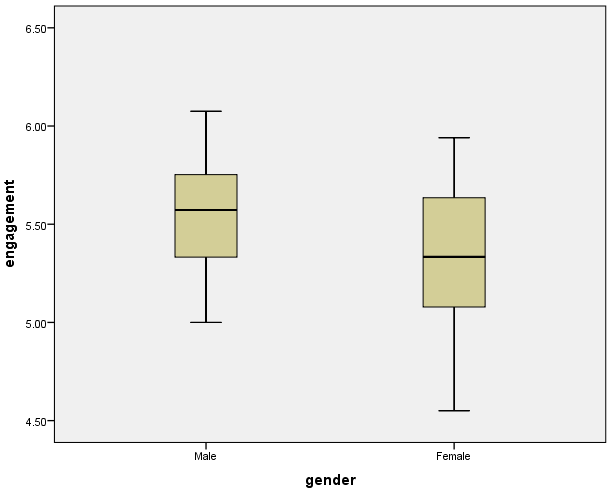
**4.** Select Dependents together in the -Boxplots- area, but deselect Stem-and-leaf in the -Descriptive- area and select Normality plots with tests, so that you end up with the following screen:

Note: The "tests" part of Normality plots with tests will generate the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. Selecting Dependents together results in one boxplot for all variables rather than a separate boxplot for each variable.

5. Click the https://statistics.laerd.com/premium/rma/img/continue-button.pngbutton. You will be returned to the **Explore** dialogue box.

6. Click the https://statistics.laerd.com/premium/rma/img/ok-button.pngbutton to generate the output.

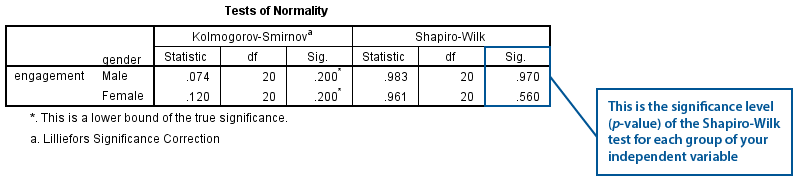
**1. Are there any outliers in our data?**



Note: Any data points that are more than 1.5 box-lengths from the edge of their box are classified by SPSS as outliers. These data points are illustrated as circular dots and labeled with their case number (i.e., their row number in the **Data View** window). If any data points are more than 3 box-lengths away from the edge of their box, they are classified as extreme points (i.e., extreme outliers) and illustrated as an asterisk (\*) with their case number labeled.

You can see that there are no outliers in our data, as evidenced by the lack of any circular points or asterisks.

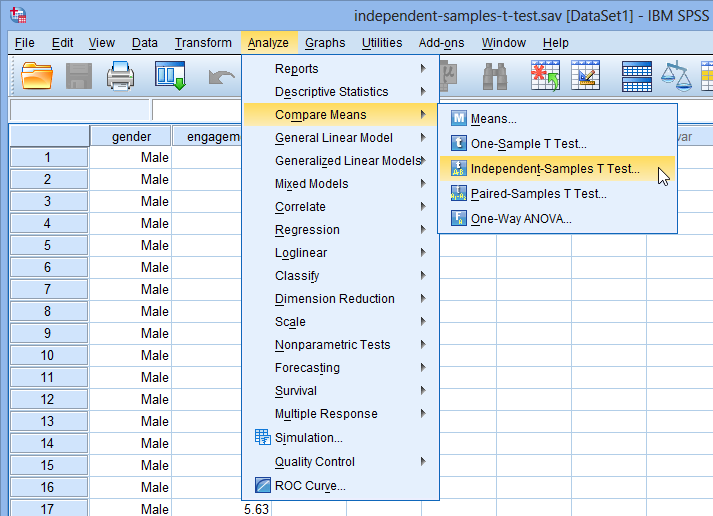
**2. Are our data normally distributed?**



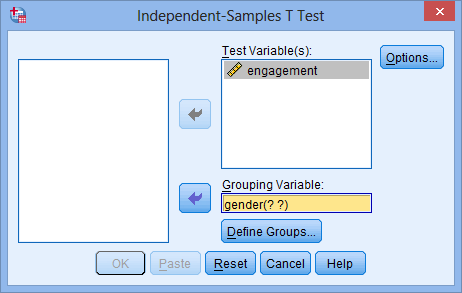
## Independent-samples t-test procedure

The following procedure shows you how to run an independent-samples t-test in SPSS. This procedure also calculates and presents Levene's Test for Equal Variances (i.e., to test for the assumption of homogeneity of variances). SPSS will produce output showing the t-test results on the basis that the assumption of homogeneity of variances has been met, as well as when it is violated (i.e., you are given two t-test results). We show you how to interpret this output after showing you how to run an independent-samples t-test in SPSS below:

Step 1: Click **Analyze > Compare Means > Independent-Samples T Test...** on the main menu:



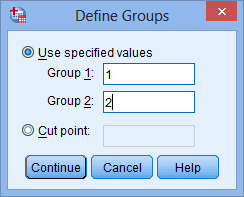
Step Two: Transfer your dependent variable into the Test Variable(s): box by selecting it (by clicking on it) and then clicking on the top https://statistics.laerd.com/premium/istt/img/right-arrow-button.pngbutton. Also transfer the independent variable into the Grouping Variable: box by selecting it and then clicking on the bottom https://statistics.laerd.com/premium/istt/img/right-arrow-button.pngbutton. You will end up with the following screen:



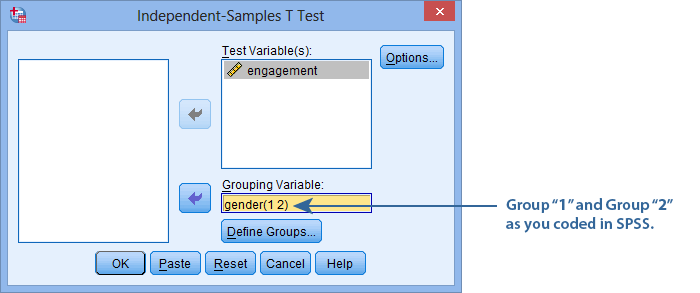
Explanation: The Test Variable(s): box is where you enter the dependent variable(s) you wish to analyze. You can transfer more than one dependent variable into this box to simultaneously analyze many dependent variables at the same time. The independent variable is referred to as the "Grouping Variable:" in SPSS for the independent-samples t-test procedure. If you transfer multiple dependent variables into the Test Variable(s): box, they will all be tested for differences between the same groups of the independent variable you transfer into the Grouping Variable: box.

Step Three: Click on the https://statistics.laerd.com/premium/istt/img/define-groups-button.pngbutton. You will be presented with the **Define Groups** dialogue box.

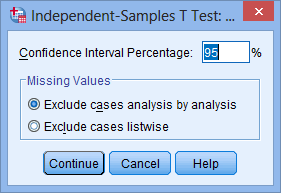
Enter "1" into the Group 1: box and "2" into the Group 2: box, as shown below:



Step Four: Click the https://statistics.laerd.com/premium/istt/img/continue-button.pngbutton and you will be returned to the **Independent-Samples T Test** dialogue box, but now with a completed Grouping Variable: box, as shown below:



Step Five: Click the https://statistics.laerd.com/premium/istt/img/options-button.pngbutton. You will be presented with the **Independent-Samples T Test: Options** dialogue box, as shown below:



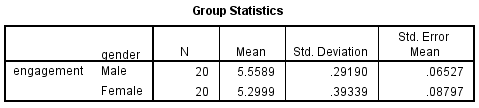
For this example, keep the default 95% confidence intervals and Exclude cases analysis by analysis in the –Missing Values– area.

Step Six: Click the https://statistics.laerd.com/premium/istt/img/continue-button.pngbutton and you will be returned to the **Independent-Samples T Test** dialogue box.

Step Seven: Click the https://statistics.laerd.com/premium/istt/img/ok-button.pngbutton to generate the output.

**Descriptive statistics**

SPSS will have generated a **Group Statistics** table containing some useful descriptive statistics for your two groups – "Male" and "Female" – which will help you get a "feel" for your data and will be used when you report your results. You can make an initial interpretation of your data using the **Group Statistics** table:



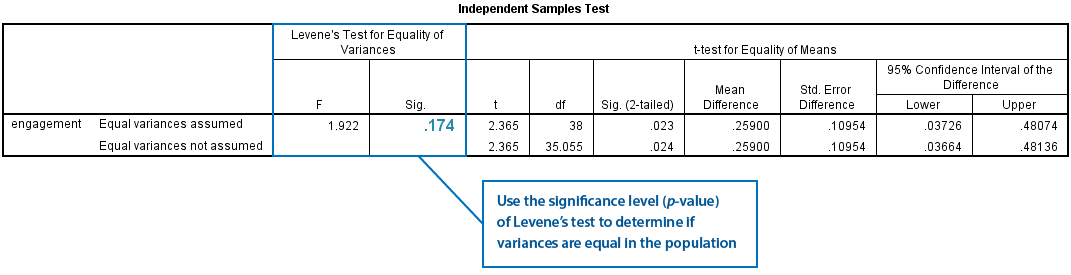
Reporting in APA format: There were 20 male and 20 female participants. The advertisement was more engaging to male viewers (M = 5.56, SD = 0.29) than female viewers (M = 5.30, SD = 0.39).

## Assumption of homogeneity of variances

An important assumption of the independent-samples t-test is that the two group's variances are equal in the population. Failure to adhere to this assumption (i.e., if variances are unequal), will generally increase the chance of making a Type I error. Equality of variances is often referred to as homogeneity of variances.

You interpret your results after running an independent-samples t-test depends on whether your data met or violated the assumption of homogeneity of variances.

* Homogeneity of variances was met: If your data has met the assumption of homogeneity of variances, you simply need to interpret the 'standard' independent-samples t-test output in SPSS.
* Homogeneity of variances was violated: If your has violated the assumption of homogeneity of variances, you can still continue with your analysis. However, you will have to interpret the results from a modified t-test that SPSS produces in its output. This modified t-test, often referred to as the **unequal variance t-test**, **separate variances t-test** or **Welch t-test** after its creator (Welch, 1947), can accommodate unequal variances and still deliver a valid test result.

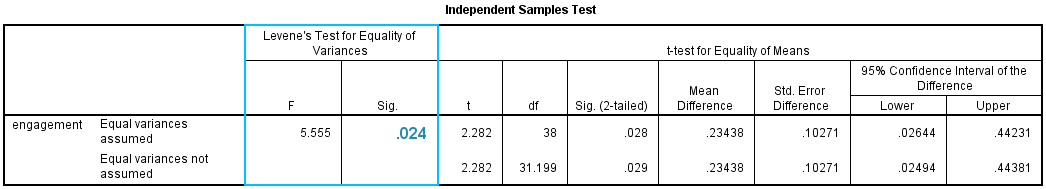


To check whether the population variances are equal, you need to consult the "**Sig.**" column located under the "**Levene's Test for Equality of Variances**" column (as highlighted above). In this example, the significance value is ".174" (i.e., p = .174). If the population variance of both groups is equal, this test will return a p-value greater than 0.05 (i.e., p > .05), indicating that you have met the assumption of homogeneity of variances. However, if the test returns a p-value less than 0.05 (p < .05), the population variances are unequal and you have violated the assumption of homogeneity of variances. In this example, the population variances of the engagement scores for both groups are equal because p = .174 (i.e., p > .05). Therefore, the assumption of homogeneity of variances is met.

Reporting in APA format: There was homogeneity of variances for engagement scores for males and females, as assessed by Levene's test for equality of variances (p = .174).

### f you have unequal variances

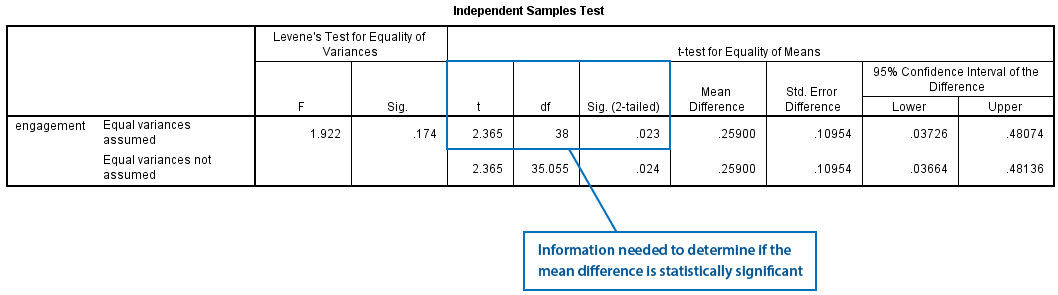
The example we have used so far met the assumption of homogeneity of variances. If you did not have homogeneity of variances, you would have a result similar to below:



As before, the "**Sig.**" column located under the "**Levene's Test for Equality of Variances**" column will inform you as to whether the assumption of equality of variances is met. In this new example, Levene's test returns a statistically significant result (i.e., p = .024, which is p < .05). Therefore, the assumption of homogeneity of variances is violated. This is sometimes referred to as heterogeneity of variances. However, in reports, it is common instead to refer to a violation of homogeneity of variances rather than state that there was heterogeneity of variances.

Reporting in APA Format: The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances (p = .024).

When the assumption of homogeneity of variances is met, you can interpret the results from the standard t-test that uses pooled variances in its calculations and requires no modification to the degrees of freedom. This is, for the most part, advantageous when we are trying to detect the mean difference between groups. The results for the standard independent-samples t-test you need to interpret are found on the "**Equal variances assumed**" row of the **Independent Samples Test**. This row is highlighted below:



You are presented with the observed t-value (the "**t**" column), the degrees of freedom (the "**df**" column), and the statistical significance (p-value) (the "**Sig. (2-tailed)**" column). If p < .05, this means that the mean difference between the two groups is statistically significant. Alternatively, if p > .05, you do not have a statistically significant mean difference between the two groups. In this example, p = .023 (i.e. p < .05). Therefore, it can be concluded that males and females have statistically significantly different mean engagement scores. Or, phrased another way, the mean difference in engagement score between males and females is statistically significant. What this result means is that there is a 23 in 1,000 chance (2.3%) of getting a mean difference at least as large as the one obtained if the null hypothesis was true (the null hypothesis stating that there is no difference between the group means). Remember, the independent-samples t-test is testing whether the means are equal in the population.

Note: If you see SPSS state that the "**Sig. (2-tailed)**" value is ".000", this actually means that p < .0005; it does not mean that the significance level is actually zero. Where possible, it is preferable to state the actual p-value rather than a greater/less than p-value statement (e.g., p = .023 rather than p < .05). This way, you convey more information to the reader about the level of statistical significance of your result.

Note: There is a link between the 95% confidence intervals of the mean difference and the statistical significance of the mean difference. If the confidence intervals do not contain the number zero, you have a statistically significant mean difference (p < .05). If they do contain zero, you do not have a statistically significant mean difference (p > .05). In this example, you discovered that the 95% confidence intervals were from 0.04 to 0.48, thus not including zero and indicating a statistically significant result.

Reporting in APA Format: There was a statistically significant difference in engagement scores between males and females, with males scoring higher than females, M = 0.26, 95% CI [0.04, 0.48], t(38) = 2.365, p < .05.

Putting it all together in APA Format: There were 20 male and 20 female participants. An independent-samples t-test was run to determine if there were differences in engagement to an advertisement between males and females. There were no outliers in the data, as assessed by inspection of a boxplot. Engagement scores for each level of gender were normally distributed, as assessed by Shapiro-Wilk's test (p > .05), and there was homogeneity of variances, as assessed by Levene's test for equality of variances (p > .05). The advertisement was more engaging to male viewers (M = 5.56, SD = 0.35) than female viewers (M = 5.30, SD = 0.35), a statistically significant difference, M = 0.26, 95% CI [0.04, 0.48], t(38) = 2.365, p < .05.