# Practical

## Inferential Statistics

***Objectives***

At the end of this practical, you should be able to:

* conduct common statistical tests to test the difference between sample means
* interpret the results generated from t-test and paired t-test

## Main Tasks

**Difference in sample means (t-test)**

##### Use of Excel

For example, suppose we are interested in comparing SCORES across GROUPS, where there are two groups. The purpose is to determine if the mean SCORE on a test is different for the two groups tested (*i.e*., control and treatment groups). The example data is shown here:

|  |  |
| --- | --- |
| **Group** | **Scores** |
| 1 | 20 |
| 1 | 23 |
| 1 | 32 |
| 1 | 24 |
| 1 | 25 |
| 1 | 28 |
| 1 | 27.5 |
| 2 | 25 |
| 2 | 46 |
| 2 | 56 |
| 2 | 45 |
| 2 | 46 |
| 2 | 51 |
| 2 | 34 |
| 2 | 47.5 |

In this example, GROUP contains two values, 1 or 2, indicating which group each subject was in. The t-test will be performed on the values in the variable (column) named SCORE.

An independent group t-test is done in two steps:

**Step 1:** Decide if the variances are equal in both groups, which determines the type of t-test to perform (one that assumes equal variances or one that doesn’t make that assumption.) A conservative approach suggested in some texts is to always assume unequal variances. Another approach is to do a statistical test to determine equality.

**Step 2:** Depending on your decision about the equality of variances you either perform the version of the t-test that assumes equality of variances or other one that doesn’t make that assumption.

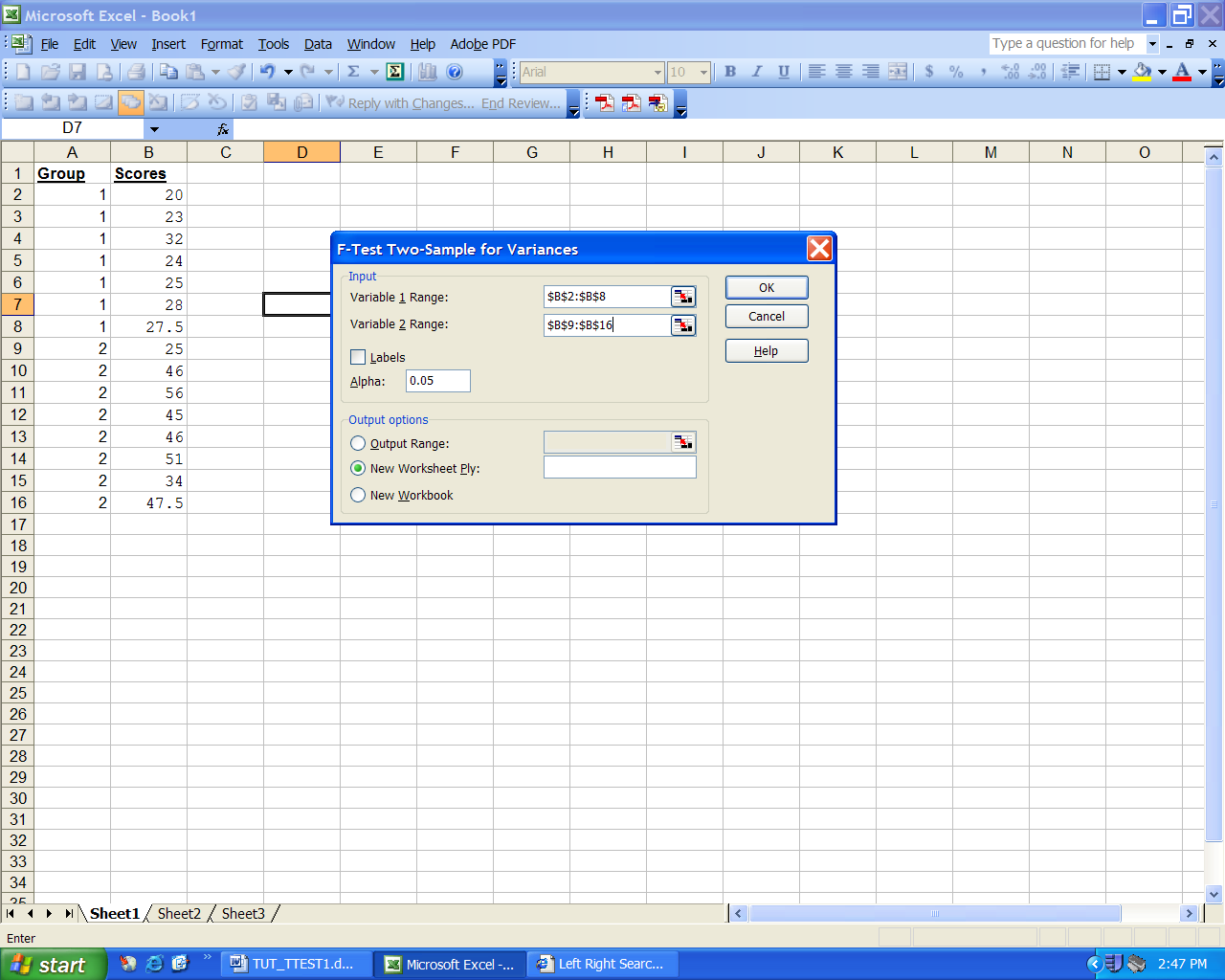
**Determine Equality of Variance**

If you take the conservative approach, skip this test and proceed to the version of the t-test that does not assume equality of variance.

To do a statistical test to determine equality of variance, follow these instructions. (The test for equality of variances is an F-test.)

1.     In Excel, select **Data/ Data Analysis / F-Test Two Sample for Variance**.

2.     In the F-Test Two Sample for Variance dialog box: For the Input Range for Variable 1, highlight the seven values of Score in group 1 (values from 20 to 27.5). For the input range for Variable 2, highlight the eight values of Score in group 2 (values from 25 to 47.5). Leave the other items at their default selections. This dialog box is shown below. Click OK.



3.     The following results are produced by Excel:

|  |  |  |
| --- | --- | --- |
| F-Test Two-Sample for Variances |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 25.64285714 | 43.8125 |
| Variance | 15.22619048 | 96.42410714 |
| Observations | 7 | 8 |
| df | 6 | 7 |
| F | 0.157908545 |  |
| P(F<=f) one-tail | 0.019378053 |  |
| F Critical one-tail | 0.23771837 |  |

Notice the highlighted probability p=0.01937. This is a one-tail p-value associated with the test for equality of variance. Generally, **if this value is less than 0.05 you assume that the variances are NOT equal.**

a.      If the variances are assumed to NOT be equal, proceed with the t-test that assumes non-equal variances.

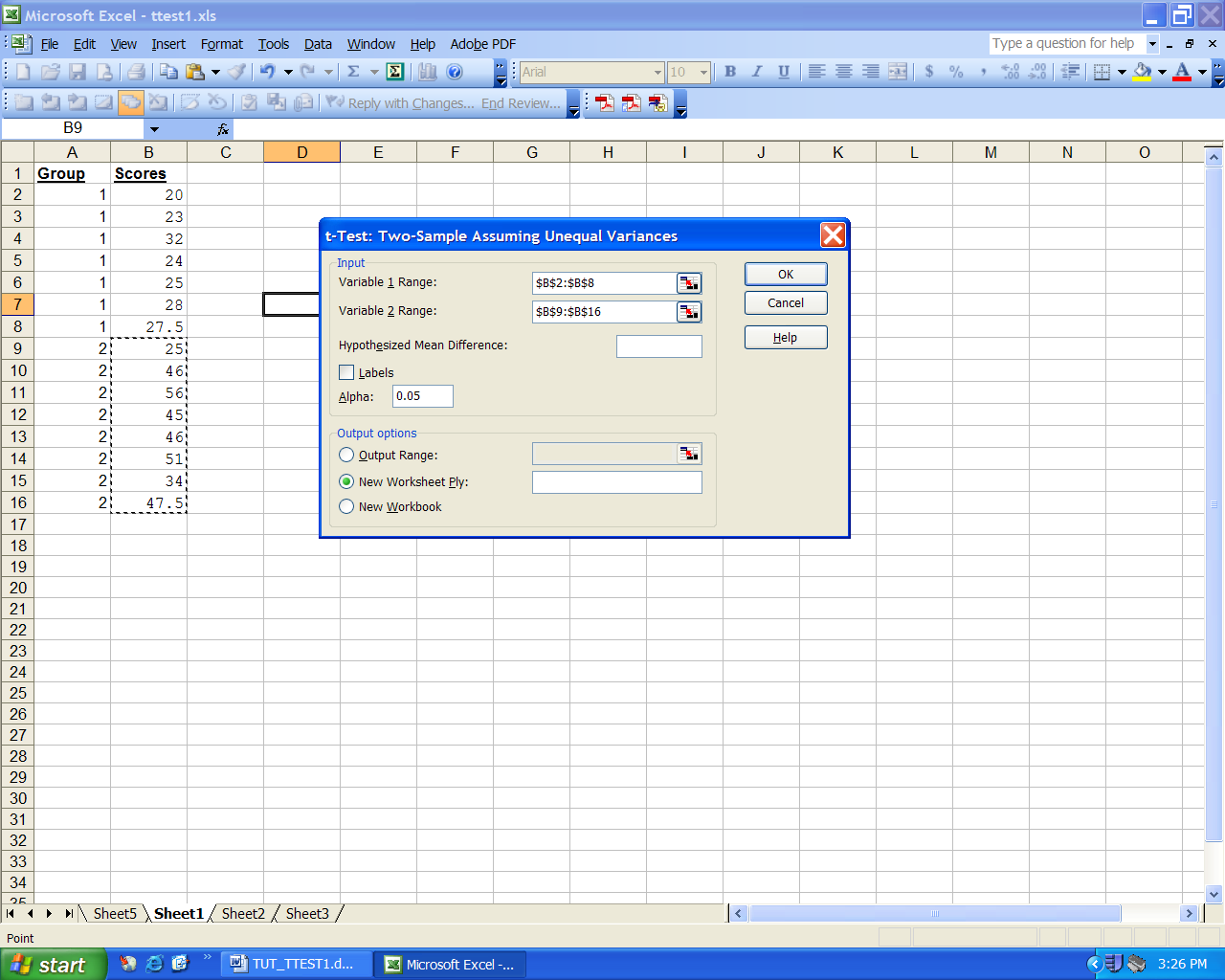
b.     If the variances are assumed to be equal, proceed with the t-test that assumes equal variances.

**Perform the t-test**

The process of doing the t-test in Excel is similar for both the equal and unequal variances case – the main difference is which version you select from the menu.  Suppose you select the unequal version of the two-sample t-test – this is how you proceed:

1.     Select **Data/ Data Analysis/ t-Test: Two Sample assuming Unequal Variances**

2.     For the Input Range for Variable 1, highlight the seven values of Score in group 1 (values from 20 to 27.5). For the input range for Variable 2, highlight the eight values of Score in group 2 (values from 25 to 47.5). Leave the other items at their default selections. This dialog box is shown below. Click OK.



3.     The following output is created:

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  |  |
|  |  |  |
|  | *Variable 1* | *Variable 2* |
| Mean | 25.64285714 | 43.8125 |
| Variance | 15.22619048 | 96.42410714 |
| Observations | 7 | 8 |
| Hypothesized Mean Difference | 0 |  |
| Df | 9 |  |
| t Stat | -4.816944724 |  |
| P(T<=t) one-tail | 0.000475506 |  |
| t Critical one-tail | 1.833112923 |  |
| P(T<=t) two-tail | 0.000951012 |  |
| t Critical two-tail | 2.262157158 |  |

Notice that the two sample mean values (variance) are 25.64(15.23) and 43.81(96.42). The two tailed calculated t-statistic is 4.82 and the highlighted p-value for this test is p=0.001. (0.000951012) Since, the p-value is less than 0.05, this provides evidence to reject the null hypothesis of equal means.

##### Exercises

1. Open the tab for “Exercise 1”. This contains the profit data of a company. Your task is to conduct the appropriate statistical test to check if there is any difference in the means for 2012 and 2013.
2. Open the tab for “Exercise 2”. This data contains the consumer’s sentiment of a product before and after watching a related advertisement. Conduct an appropriate statistical test to check if there is any difference in the means.

**~~END~~**