

Task 3: Statistical hypothesis tests for comparing two samples

In previous tasks we observed that dolphin skulls from the North Island sample are larger than those of the South Island sample. In this task we compare these two samples using a statistical hypothesis test, called t test.

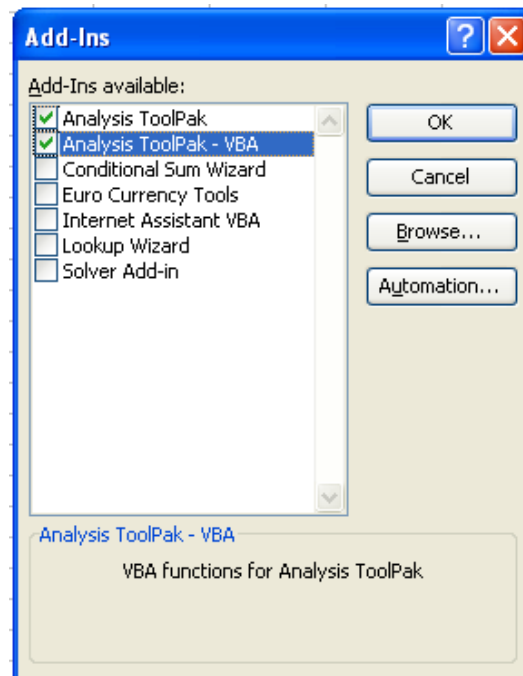
In this task, the hypotheses we are testing are: (1) the means of the two samples are the same (hence, there is no difference between the North Island and South Island samples) vs. (2) the mean of the North Island sample is larger than that of the South Island sample. We tests these hypotheses based on the evidence provided by the data.

- (1) Insert a new worksheet, and name it as “t test”.
- (2) Copy and paste all data from the “Dolphins data” worksheet to the “t test” worksheet.

The t test procedure is available in Excel only through its add-in package. Therefore, we next need to load this add-in package to Excel, as follows.

- (3) Load “Analysis tool” package by clicking on **Tools** menu and **Add-Ins...** option.

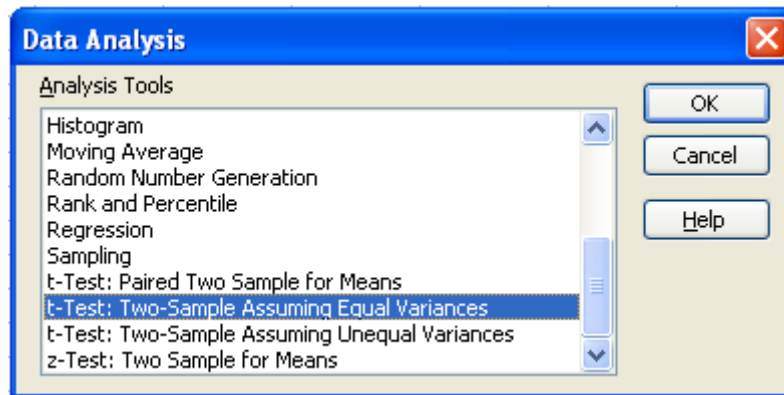
In the Add-Ins dialog box, select **Analysis ToolPak** and **Analysis ToolPak – VBA** as shown below, and click **OK**.



Now the add-in package is added to Excel.

- (4) Click on an empty cell on the “t test” worksheet and then, click the **Tools** menu, and select **Data Analysis...** option.

- (5) In the **Data Analysis** window, select the **t-Test: Two-Sample Assuming Equal Variances** option in the **Analysis Tools** box. Click **OK**.



(As you have seen above, there are two other options in *t* test procedures.

The **Paired Two Sample for Means** option should be used when the two samples tested are actually the same sample, for example, when the same subjects were treated in two different ways and we want to compare the mean effects of these two treatments on the subjects.

The **Two-Sample Assuming Unequal Variances** option should be used when we know that variances of the two samples tested are different.

In this task, we assume that the variances of our two samples are not different. This is because in Tasks 1 and 2 we have observed that the standard deviations and the spread of our two samples are similar in $X_1 \sim X_6$.

Alternatively, we can also use a statistical hypothesis test to see if our assumption of the equal variances here is appropriate or not. However, such testing procedures are unfortunately only taught at university level. Hence, for now we need to assume the equal variances based on our observation.)

- (6) We first compare the means of X_1 . In the following window,

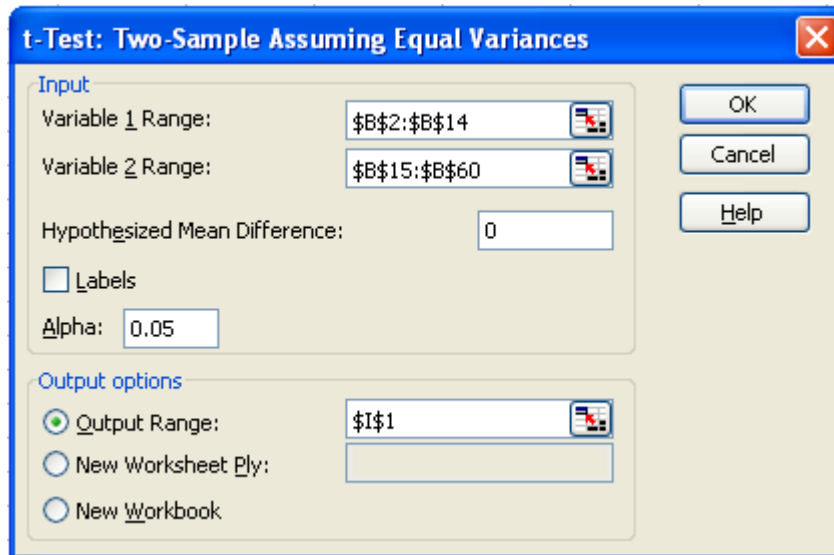
in the **Input** area,

select the 13 X_1 values of the North Island sample for the **Variable 1 Range** box, and the 46 X_1 values of the South Island sample for the **Variable 2 Range** box.

Enter **0** for the **Hypothesized Mean Difference** box and **0.05** for the **Alpha** box.

In the **Output options** area,

Click on the **Output Range** option and then, select an empty cell on the “t test” worksheet in the **Output Range** box, as shown on the following page.



Click **OK**. Then, the following result should appear.

	I	J	K
t-Test: Two-Sample Assuming Equal Variances			
		<i>Variable 1</i>	<i>Variable 2</i>
Mean		304.5385	281.6087
Variance		120.1026	99.48792
Observations		13	46
Pooled Variance		103.8278	
Hypothesized Mean Difference		0	
df		57	
t Stat		7.164193	
P(T<=t) one-tail		8.47E-10	
t Critical one-tail		1.672029	
P(T<=t) two-tail		1.69E-09	
t Critical two-tail		2.002465	

Note that this t test used the pooled variance of 103.83 to estimate the equal variances of the two samples.

Because we are testing the hypotheses: (1) the means of the two samples are the same vs. (2) the mean of the North Island sample is larger than that of the South Island sample, we should look at the value of $8.47\text{E}-10$ ($= 8.47 \times 10^{-10} = 0.00000000847$) in the **P(T<=t) one-tail** row.

Then, because $8.47\text{E}-10 < 0.05$, we now know that our hypothesis (1), that is, the means of the two samples are the same, must be rejected. In other words, the data provides strong evidence to support our hypothesis (2), that is, the mean of the North Island sample is larger than that of the South Island sample.

(7) Repeat the steps (4) ~ (6) for X2, X3, X4, X5 and X6.

Congratulations! You have successfully completed this task.

Do not forget to save all the worksheets you created, as we need to use them later for other tasks.

Name your worksheets appropriately, so that you can recognise them later.