University of Essex

Online

7.1 Hypothesis Testing Using Excel (see below for LibreOffice)

The Related Samples T Test

Example 7.1

Consider the container design data in Data Set F (see the Data Annexe). Notice that the two variables Con1 and Con 2 indeed measure the same characteristic (the number of items sold), but under two different “conditions” (the two different container designs).

We conduct a two-tailed related samples t test of whether the underlying (population) mean number of items sold differs between the two container designs.

Strictly speaking, before undertaking the test we should calculate the differences **D = Con1 – Con 2** for each observation. A normal plot of these differences (i.e., of the values of the variable D) should then be constructed in order to check whether the data are acceptably near-normally distributed.

We will assume for now that the data are indeed so distributed so that the resulting t test is valid. You might want to construct the normal plot as an additional exercise. (Ensure you save your answers in the Exercise sheets for your submission.)

1. Open the Excel workbook **Exa 8.4F.xlsx** from the Examples folder. This contains the relevant data.

2. From the **Data** menu bar tab, select **Data Analysis** from the **Analysis** group, and from the ensuing dialogue box, select **t test: Paired Two Sample for Means**. A new dialogue box appears.

3. In the **Variable 1 Range** box, enter the cell range where the data for the first variable (Con1) can be found, including the variable name, that is, the range B1:B11. In the **Variable 2 Range** box, enter the cell range where the data for the second variable (Con2) can be found, including the variable name, that is, the range C1:C11. Ensure that the **Labels** box is checked.

4. Type:0 in the **Hypothesised Mean Difference** box. This represents the null hypothesis of no difference between the treatment means.

5. Ensure that the **Alpha** box contains the value 0.05. This is only of marginal relevance, as we shall make direct use of the p-value that will be output.

6. Select the **Output Range** button, and in the corresponding box, enter the cell reference E1. Click the **OK** button. Some output appears in your spreadsheet. Widen columns E, F and G so that all the text becomes readable.

7. In cell E16, type: Difference in Means, and in cell F16, enter the formula **=F4-G4**.

The resulting output is presented below.

Not all this output is relevant, so it need not all be discussed.

The obtained related samples $t = 2.875$ with 9 degrees of freedom.

The associated two-tailed p-value is $p = 0.018$, so the observed t is significant at the 5% level (two-tailed).

t-Test: Paired Two Sample for Means		
	Con1	Con2
Mean	172.6	159.4
Variance	750.2666667	789.3777778
Observations	10	10
Pearson Correlation	0.863335004	
Hypothesised Mean Difference	0	
df	9	
t Stat	2.874702125	
P(T<=t) one-tail	0.009167817	
t Critical one-tail	1.833112923	
P(T<=t) two-tail	0.018335635	
t Critical two-tail	2.262157158	
Difference in Means	13.2	

The data therefore constitute strong evidence (on a one-tailed test) that the underlying mean number of containers sold was greater for Design 1, by an estimated 172.6 - 159.4 = 13.2 items per store. The results continue to suggest that Design 1 should be preferred. Although broadly similar conclusions were reached as before, a higher level of significance was obtained with the one-tailed test.

Notice that if we had sought to test the alternative pair of one-tailed hypotheses **H_0: $\mu_1 \geq \mu_2$ against H_1: $\mu_1 < \mu_2$** we would have found the difference in sample means to be consistent with the *null hypothesis* that the population mean sales for Design 2 was no greater than that for Design 1. We would thus have declared the result to be not significant without even bothering to inspect the p-value.