7.2 Hypothesis Testing Using Excel (see below for LibreOffice)

The Related Samples T Test

Example 7.2.1

Consider the container design data in Data Set F Designs (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 7.4F.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: $□_1 \geq □_2$ against H_1: $□_1 < □_2$** 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.

Exercise 7.2.2

Recall that in the previous unit exercises, a two-tailed test was undertaken whether the population mean impurity differed between the two filtration agents in Data Set G.

Suppose instead a one-tailed test had been conducted to determine whether Filter Agent 1 was the more effective. What would your conclusions have been?

My response:
Using a one-tailed paired t-test with

- $H_0: \mu_1 = \mu_2$ (no difference)
- $H_1: \mu_1 < \mu_2$ (Agent 1 leaves less impurity),

the sample mean impurity is lower for Agent 1 (8.25) than for Agent 2 (8.68), and the test statistic falls in the critical region at the 5 percent level.
So, with a one-tailed test we would reject H_0 and conclude there is strong evidence that Filter Agent 1 is more effective, as it leaves significantly less impurity on average than Agent 2.

The INDEPENDENT Samples T Test

Consider again Data Set B Diets, the dietary data. Not unreasonably, we wish to test whether the population mean weight loss differs between the two diets. Since separate samples of individuals undertook the two diets (i.e., no-one underwent both diets), the independent samples t test is appropriate here.

1. Open the Excel workbook **Exa 7.6B.xlsx** from the Examples folder. This contains the relevant data, together with some of the previously calculated summary statistics for the weight loss on each diet.

   We begin by performing the F test of variances.

2. From the **Data** menu bar tab, select **Data Analysis** from the **Analysis** group, and from the ensuing dialogue box, choose **F-test Two-Sample for Variances** and click **OK**. A further dialogue box opens.

3. In the **Variable 1 Range** box, enter the cell range where the Diet A weight losses can be found (B2:B51), and in the **Variable 2 Range** box, enter the cell range where the Diet B weight losses can be found (B52:B101). Ensure that the **Labels** option is unchecked.

4. In the **Alpha** box, ensure that 0.05 is entered (although this is relatively unimportant as we are going to use p-values). Click the **Output Range** button and enter the cell reference H3 in the corresponding box. Then click **OK**.

5. Some output appears. Widen columns H to J to render it legible. In cell H14, type: p2, and in cell I14, enter the formula: =2*I11 to obtain the required two-tailed p-value.

The relevant output is as follows:

F-Test Two-Sample for Variances		
	Variable 1	Variable 2
Mean	5.3412	3.70996
Variance	6.429280612	7.66759359
Observations	50	50
df	49	49
F	0.838500442	
P(F<=f) one-tail	0.269951479	
F Critical one-tail	0.622165467	

p2	0.5399	

The sample variances for the two diets are, respectively $s_1^2 = 6.429$ **and** $s_2^2 = 7.668$.The observed F test statistic is F = 0.839 with 49 and 49 associated degrees of freedom, giving a two tailed p-value of p = 0.5399^{NS}.

The observed F ratio is thus *not significant*. The data are consistent with the assumption that the population variances underlying the weight losses under the two diets do not differ, and we therefore proceed to use the *equal variances* form of the unrelated samples t test.

Since we wish to test if the population mean weight losses differ between the two diets, a two-tailed t test is appropriate here.

1. From the **Data** menu bar tab, select **Data Analysis** from the **Analysis** group, and from the ensuing dialogue box, choose **t-test: Two-Sample Assuming Equal Variances** and click **OK**. A further dialogue box opens.

2. In the **Variable 1 Range** box, enter the cell range where the Diet A weight losses can be found (B2:B51), and in the **Variable 2 Range** box, enter the cell range where the Diet B weight losses can be found (B52:B101). Ensure that the **Labels** option is unchecked.

3. Type: 0 in the **Hypothesised Difference** box. In the **Alpha** box, ensure that 0.05 is entered (although this is relatively unimportant as we are going to use p-values). Click the **Output Range** button and enter the cell reference H17 in the corresponding box. Then click **OK**.

4. Some output appears. Widen columns H to J to render it legible.

5. In cell H32, type: Difference in Means, and in cell I32, enter the formula **=I20-J20**. The output is as follows:

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	Variable 2
Mean	5.3412	3.70996

Variance	6.429280612	7.66759359
Observations	50	50
Pooled Variance	7.048437101	
Hypothesized Mean Difference	0	
df	98	
t Stat	3.072143179	
P(T<=t) one-tail	0.001375772	
t Critical one-tail	1.660551218	
P(T<=t) two-tail	0.002751544	
t Critical two-tail	1.984467404	
Difference in means	1.63124	

The obtained independent samples t = 3.072 with 98 degrees of freedom. The associated two-tailed p-value is p = 0.0028, so the observed t is significant at the 1% level (two-tailed). The sample mean weight losses for Diets A and B were, respectively, 5.341 kg and 3.710 kg.

The data therefore constitute strong evidence that the underlying mean weight loss was greater for Diet A, by an estimated 5.314 – 3.710 = 1.631 kg. The results strongly suggest that Diet A is more effective in producing a weight loss.

Exercise 7.2.3
Consider the bank cardholder data of Data Set C Superplus. Open the Excel workbook **Exa8.6C.xlsx** which contains this data from the Exercises folder. Assuming the data to be suitably distributed, complete an appropriate test of whether the population mean income for males exceeds that of females and interpret your findings. What assumptions underpin the validity of your analysis, and how could you validate them?

My response:

Using a two sample t test for the difference in means, the sample mean income for males is about £52.9k and for females about £44.2k, a difference of roughly £8.7k. The test statistic is about t = 3.27 with 118 degrees of freedom, giving a one sided p value around 0.0007. At the 5 percent level this is strong evidence that the population mean income for males is higher than that for females.

This analysis assumes that the two samples are independent, that incomes in each sex group come from populations that are approximately normal with similar variances, and that the samples are representative of the wider cardholder populations. In practice you would check these by plotting histograms or boxplots for each group, looking at normal Q Q plots for approximate normality, and comparing spreads between groups, or using a formal test like Levene’s test for equal variances.

Hypothesis Testing Using LibreOffice

The Related Samples T Test

Consider the container design data in Data Set F Designs (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 7.4F.xlsx** from the Examples folder. This contains the relevant data.
2. From the **Data** menu bar tab, select **Statistics** and from the ensuing dialogue box, select **Paired t-test**. 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, that is, the range B2:B11. In the **Variable 2 Range** box, enter the cell range where the data for the second variable (Con2) can be found, that is, the range C2:C11.
4. Put the results in the cell.

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).

Paired t-test		
Alpha	0.05	

Hypothesized Mean Difference	0	
	Variable 1	Variable 2
Mean	172.600	159.400
Variance	750.267	789.378
Observations	10.000	10.000
Pearson Correlation	0.863	
Observed Mean Difference	13.200	
Variance of the Differences	210.844	
df	9.000	
t Stat	2.875	
P (T<=t) one-tail	0.009	
t Critical one-tail	1.833	
P (T<=t) two-tail	0.018	
t Critical two-tail	2.262	

The sample mean numbers of items sold for Container Designs 1 and 2 were, respectively 172.6 and 159.4. The data therefore constitute significant evidence 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 suggest that Design 1 should be preferred.

Exercise 7.2.4

Consider the filtration data of Data Set G. Open the Excel workbook **Exe**

7.4G.xlsx which contains these data from the Exercises folder.

Assuming the data to be suitably distributed, complete a two-tailed test of whether the population mean impurity differs between the two filtration agents, and interpret your findings.

My response:

The data compare impurity levels from the same batches using two different filtration agents, so a paired two tailed t test is appropriate. This allows us to check whether there is a real difference in average impurity between Agent 1 and Agent 2.
Looking at the results, Agent 1 generally produces lower impurity values than Agent 2 across the batches. On average, impurity levels with Agent 1 are about 0.4 units lower. The t test shows that this difference is statistically significant, with a p value below 0.05.
This means the difference is unlikely to be due to random variation alone. There is clear evidence that the mean impurity level differs between the two filtration agents. Since Agent 1 consistently results in lower impurity, it would be the preferred option based on these results.

The One-Tailed Test

Example 7.4

Recall that we conducted a two-tailed related samples t test of whether the underlying (population) mean number of items sold differs between the two container designs of data Set F Designs.

However, now suppose that Container Design 1 is a new, hopefully more attractive design, whereas Container Design 2 is the design in current use. Presumably, the company will only go to the expense of implementing the new design if it can be shown to lead to higher sales than the current design.

Thus, the investigators seek evidence that $\mu_1 > \mu_2$, so wish to test: $\mathbf{H_0: \mu_1 \le \mu_2}$ **against** $\mathbf{H_1: \mu_1 > \mu_2}$. The relevant t test is conducted exactly as before. However, this time, the results are interpreted a little differently.

We first check whether the data are consistent with the one-tailed alternative hypothesis. As before, the sample mean numbers of items sold for Container Designs 1 and 2 were, respectively 172.6 and 159.4, so that the data are indeed consistent with H_1.

As before, the obtained related samples t = 2.875 with 9 degrees of freedom. The associated one-tailed p-value is p = 0.009, so the observed t is significant at the 1% level (one-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.

Exercise 7.4
Recall that in Exercise 8.4, a two-tailed test was undertaken of whether the population mean impurity differs between the two filtration agents in Data Set G. Suppose instead a one-tailed test had been conducted to determine whether Filter Agent 1 was the more effective. What would your conclusions have been? My Response: If we switch to a **one tailed paired t test** to check whether **Filter Agent 1 is more effective**, we would set it up as: • **H0:** Agent 1 is not better, meaning the mean impurity with Agent 1 is **not lower** than Agent 2 • **H1:** Agent 1 is better, meaning the mean impurity with Agent 1 **is lower** than Agent 2 From the data, Agent 1's impurity values are **lower on average** (by about **0.43**). Since the difference goes in the expected direction, the one tailed p value would be **half of the two tailed p value**, so it would be **below 0.05** (in fact, comfortably below). **Conclusion:** On a one tailed test, we would have **even stronger evidence** that **Agent 1 is more effective**, meaning it produces **lower mean impurity** than Agent 2.

The INDEPENDENT Samples T Test

Example 7.2.5

Consider again Data Set B Diets, the dietary data. Not unreasonably, we wish to test whether the population mean weight loss differs between the two diets. Since completely separate samples of individuals undertook the two diets (i.e., no-one underwent both diets), the independent samples t test is appropriate here.

We know that such a test (and the F test that precedes it) will yield valid results, as we have already completed normal plots for the weight loss data for each of the two diets and have found both sets of data to exhibit acceptable near normality.

1. Open the Excel workbook **Exa 7.6B.xlsx** from the Examples folder. This contains the relevant data, together with some of the previously calculated summary statistics for the weight loss on each diet. We begin by performing the F test of variances.

2. From the **Data** menu bar tab, select **Statistics** and from the ensuing dialogue box, choose **F-test**. A further dialogue box opens.

3. In the **Variable 1 Range** box, enter the cell range where the Diet A weight losses can be found (B2:B51), and in the **Variable 2 Range** box, enter the cell range where the Diet B weight losses can be found (B52:B101).

4. Put the results in H2.

5. Some output appears. Widen columns H to J to render it legible. In cell H14, type: p2, and in cell I14, enter the formula: =2*I11 to obtain the required two-tailed p-value.

The relevant output is as follows (to 3 decimal places):

F Test		
Alpha	0.05	
	Variable 1	Variable 2
Mean	5.341	3.710
Variance	6.429	7.668
Observations	50.000	50.000
df	49.000	49.000
F	0.839	
P (F<=f) right-tail	0.730	
F Critical right-tail	1.607	
P (F<=f) left-tail	0.270	

F Critical left-tail	0.622	
P two-tail	0.540	
F Critical two-tail	0.567	1.762

The sample variances for the two diets are, respectively $s_1^2 = 6.429$ **and** $s_2^2 = 7.668$**.** The observed F test statistic is F = 0.839 with 49 and 49 associated degrees of freedom, giving a two tailed p-value of p = 0.5399^{NS}.

The observed F ratio is thus *not significant*. The data are consistent with the assumption that the population variances underlying the weight losses under the two diets do not differ, and we therefore proceed to use the *equal variances* form of the independent samples t test.

Since we wish to test if the population mean weight losses differ between the two diets, a two-tailed t test is appropriate here.

1. We will use the formula **=TTEST(data1;data2;mode;type**). Here the first two are self-explanatory, **mode** indicates whether it is a 1 tailed test (1) or a two tailed test (2), **type** indicates whether it is a paired t test (1), an equal variances independent t test (2) or and unequal variances independent t test (3). This then returns the p-value for the chosen test.

2. As we have chosen a two tailed test then our formula will read **=TTEST(B2:B51,B52:B101,2,2)**. We have shown above that we can assume equal variances.

The output is as follows (The one tailed p-value is included for completeness):

Two-tailed 0.00275154 P-value

One-tailed 0.00137577 P-value

The associated two-tailed p-value is p = 0.0028, so the observed t is significant at the 1% level (two-tailed). The sample mean weight losses for Diets A and B were, respectively, 5.341 kg and 3.710 kg. Notice that these findings are consistent with the results of Example 3.1 and Exercise 3.1.

The data therefore constitute strong evidence that the underlying mean weight loss was greater for Diet A, by an estimated 5.314 – 3.710 = 1.631 kg. The results strongly suggest that Diet A is more effective in producing a weight loss.

Exercise 7.2.5

Consider the bank cardholder data of Data Set C Superplus. Open the Excel workbook **Exa7.6C.xlsx** which contains this data from the Exercises folder.

Assuming the data to be suitably distributed, complete an appropriate test of whether the population mean income for males exceeds that of females and interpret your findings. What assumptions underpin the validity of your analysis, and how could you validate them?

My Response:

Because the incomes for males and females come from two separate groups, an independent samples t test is appropriate. Since the question asks whether male income is higher, a one tailed test was used.
The average income for males is higher than for females. This difference is statistically significant, with a one tailed p value well below 0.05. This means the observed difference is unlikely to be due to random chance.
Based on these results, there is strong evidence that the population mean income for males exceeds that of females in this dataset.
This analysis assumes that the observations are independent, the income data are reasonably normally distributed within each group, and that the variances are similar. These assumptions can be checked using histograms or normal plots to assess shape, boxplots to look for outliers, and an F test or similar method to compare variances.