Exercise 7.1

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?

The INDEPENDENT Samples T Test

Consider again Data Set B, 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 8.6B.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

Consider the bank cardholder data of Data Set C. 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?