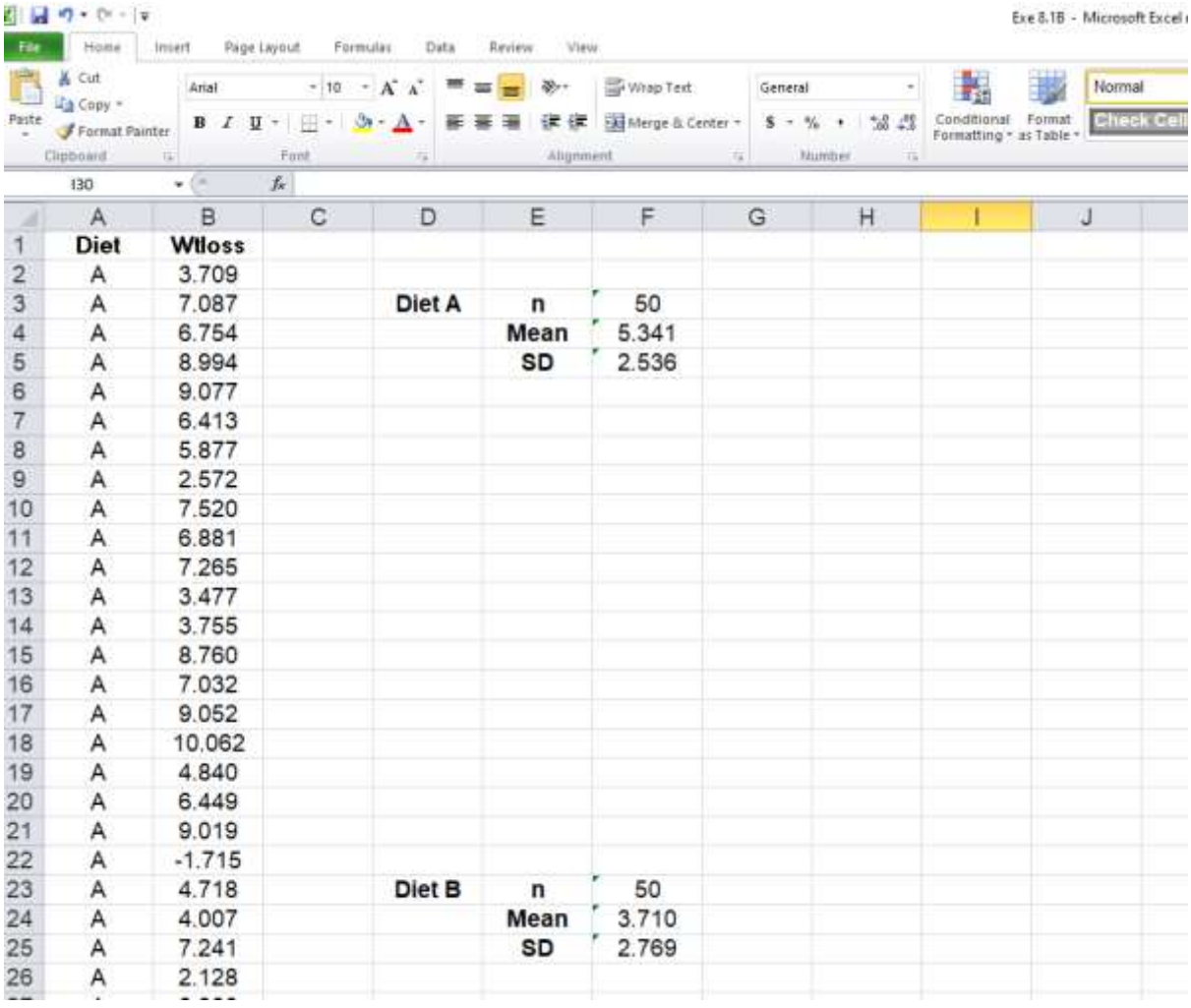
Exercise 8.1

Open the Excel workbook in **Exe 8.1B.xlsx** from the Exercises folder. Obtain the sample size, sample mean weight loss and the sample standard deviation of the weight loss for Diet B. Place these results in the block of cells F23 to F25, using the same format as that employed for the Diet A results in the above example.

Briefly interpret your findings. What do these results tell you about the relative effectiveness of the two weight-reducing diets?

Solution 8.1:



	A	B	C	D	E	F	G	H	I	J
1	**Diet**	**Wtloss**								
2	A	3.709								
3	A	7.087		**Diet A**	**n**	50				
4	A	6.754			**Mean**	5.341				
5	A	8.994			**SD**	2.536				
6	A	9.077								
7	A	6.413								
8	A	5.877								
9	A	2.572								
10	A	7.520								
11	A	6.881								
12	A	7.265								
13	A	3.477								
14	A	3.755								
15	A	8.760								
16	A	7.032								
17	A	9.052								
18	A	10.062								
19	A	4.840								
20	A	6.449								
21	A	9.019								
22	A	-1.715								
23	A	4.718		**Diet B**	**n**	50				
24	A	4.007			**Mean**	3.710				
25	A	7.241			**SD**	2.769				
26	A	2.128								

The following data summary pertains to the weight loss (in kg) results of two distinct groups, each consisting of 50 individuals, who followed either of the two distinct weight-reduction diets, labelled as A and B.

Diet A	
n	50
Mean	5.341
SD	2.536
Diet B	
n	50
Mean	3.710
SD	2.769

Diet A has a higher mean weight reduction (5.341kg) compared to Diet B (3.710 kg). This suggests that, on average, people who followed Diet A lost more weight than those who followed Diet B.

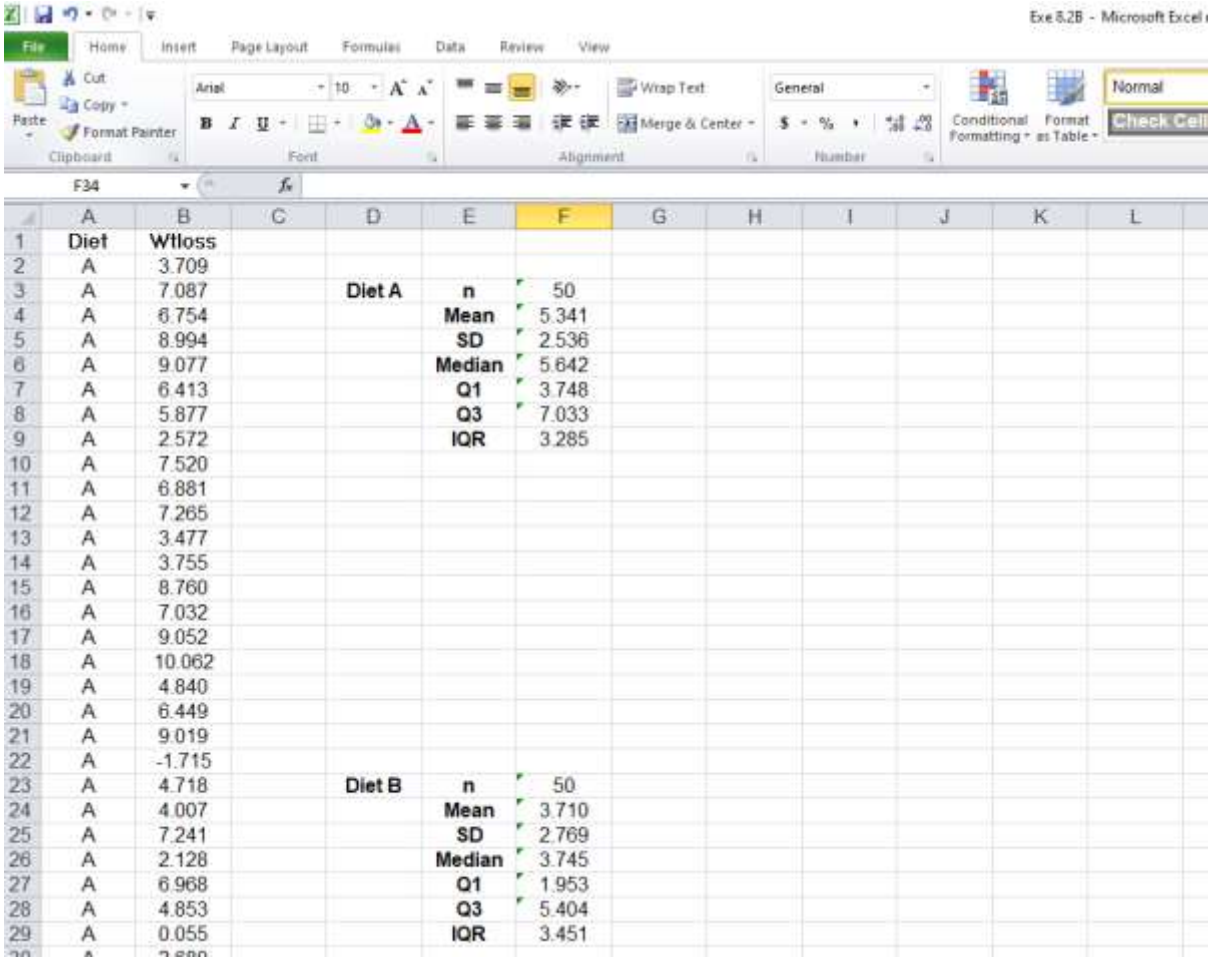
Diet A has a lower standard deviation (2.536) compared to Diet B (2.769). A lower standard deviation indicates that the weight reductions in Diet A are more consistent or have less variation compared to Diet B.

Exercise 8.2

Open the Excel workbook in **Exe 8.2B.xlsx** from the Exercises folder. Obtain the sample median, first and third quartiles and the sample interquartile range of the weight loss for Diet B. Place these results in the block of cells F26 to F29, using the same format as that employed for the Diet A results in the above example.

Briefly interpret your findings. What do these results tell you about the relative effectiveness of the two weight-reducing diets?

Solution 8.2:



	A	B	C	D	E	F	G	H	I	J	K	L
1	Diet	Wtloss										
2	A	3.709										
3	A	7.087		Diet A	n	50						
4	A	6.754			Mean	5.341						
5	A	8.994			SD	2.536						
6	A	9.077			Median	5.642						
7	A	6.413			Q1	3.748						
8	A	5.877			Q3	7.033						
9	A	2.572			IQR	3.285						
10	A	7.520										
11	A	6.881										
12	A	7.265										
13	A	3.477										
14	A	3.755										
15	A	8.760										
16	A	7.032										
17	A	9.052										
18	A	10.062										
19	A	4.840										
20	A	6.449										
21	A	9.019										
22	A	-1.715										
23	A	4.718		Diet B	n	50						
24	A	4.007			Mean	3.710						
25	A	7.241			SD	2.769						
26	A	2.128			Median	3.745						
27	A	6.968			Q1	1.953						
28	A	4.853			Q3	5.404						
29	A	0.055			IQR	3.451						

Following are based on same data as in Exercise 8.1. In addition provide Median, first quartile, third quartile and interquartile range for Diet A and Diet B

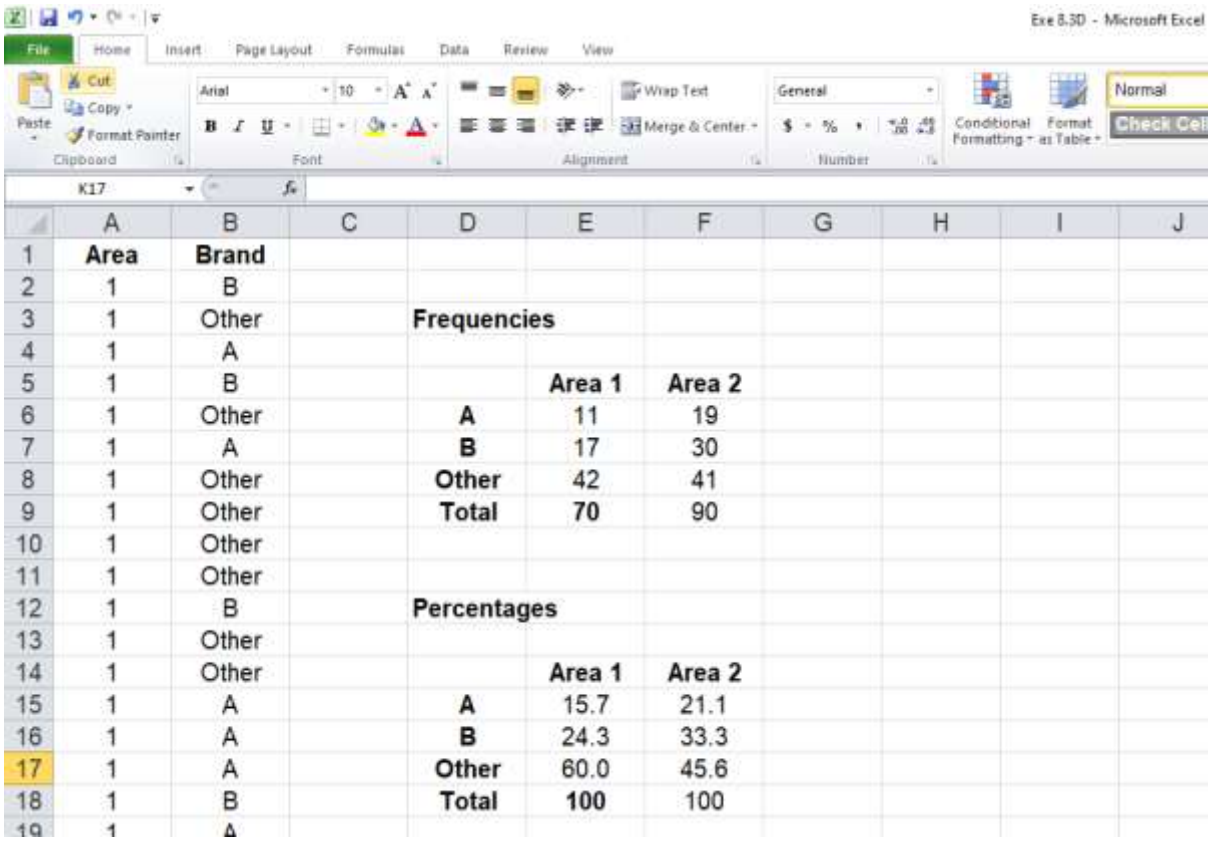
	Diet A	Diet B
n	50	50
Mean	5.341	3.710
SD	2.536	2.769
Median	5.642	3.745
Q1	3.748	1.953
Q3	7.033	5.404
IQR	3.285	3.451

Diet A appears to have a higher mean weight loss and a wider interquartile range compared to Diet B, suggesting that Diet A may have more variability in weight loss outcomes. Diet B, on the other hand, has a lower mean weight loss and a narrower interquartile range, indicating that it may have more consistent but lower weight loss outcomes compared to Diet A

Exercise 8.3
Open the Excel workbook in **Exe 8.3D.xlsx** from the Exercises folder. Obtain the frequencies and percentage frequencies of the variable Brand, but this time for the Area 2 respondents, using the same format as that employed for the Area1 results in the above example.

Briefly interpret your findings. What do these results tell you about the patterns of brand preferences for each of the two demographic areas?

Solution 8.3:



	A	B	C	D	E	F	G	H	I	J
1	**Area**	**Brand**								
2	1	B								
3	1	Other		**Frequencies**						
4	1	A								
5	1	B			**Area 1**	**Area 2**				
6	1	Other		**A**	11	19				
7	1	A		**B**	17	30				
8	1	Other		**Other**	42	41				
9	1	Other		**Total**	**70**	90				
10	1	Other								
11	1	Other								
12	1	B		**Percentages**						
13	1	Other								
14	1	Other			**Area 1**	**Area 2**				
15	1	A		**A**	15.7	21.1				
16	1	A		**B**	24.3	33.3				
17	1	A		**Other**	60.0	45.6				
18	1	B		**Total**	**100**	100				
19	1	A								

Following are the results of a marketing study, samples of individuals in each of two different demographic areas were asked to state their brand preferences for a certain type of breakfast cereal. Of particular interest were two brands (A and B) made by a certain manufacturer.

	Frequencies		Percentages	
	Area 1	**Area 2**	**Area 1**	**Area 2**
A	11	19	15.7	21.1
B	17	30	24.3	33.3
Other	42	41	60.0	45.6
Total	**70**	90	**100**	100

From these results, we can see that in Area 1, there is a significant preference for "Other" breakfast cereals, with the majority of respondents choosing options other than Brands A and B. Brand B has a higher preference compared to Brand A, but neither Brand A nor Brand B is the dominant choice in this area.

In Area 2, there is a stronger preference for both Brand A and Brand B compared to Area 1. However, similar to Area 1, the majority of respondents in Area 2 still prefer "Other" breakfast cereals over Brands A and B.

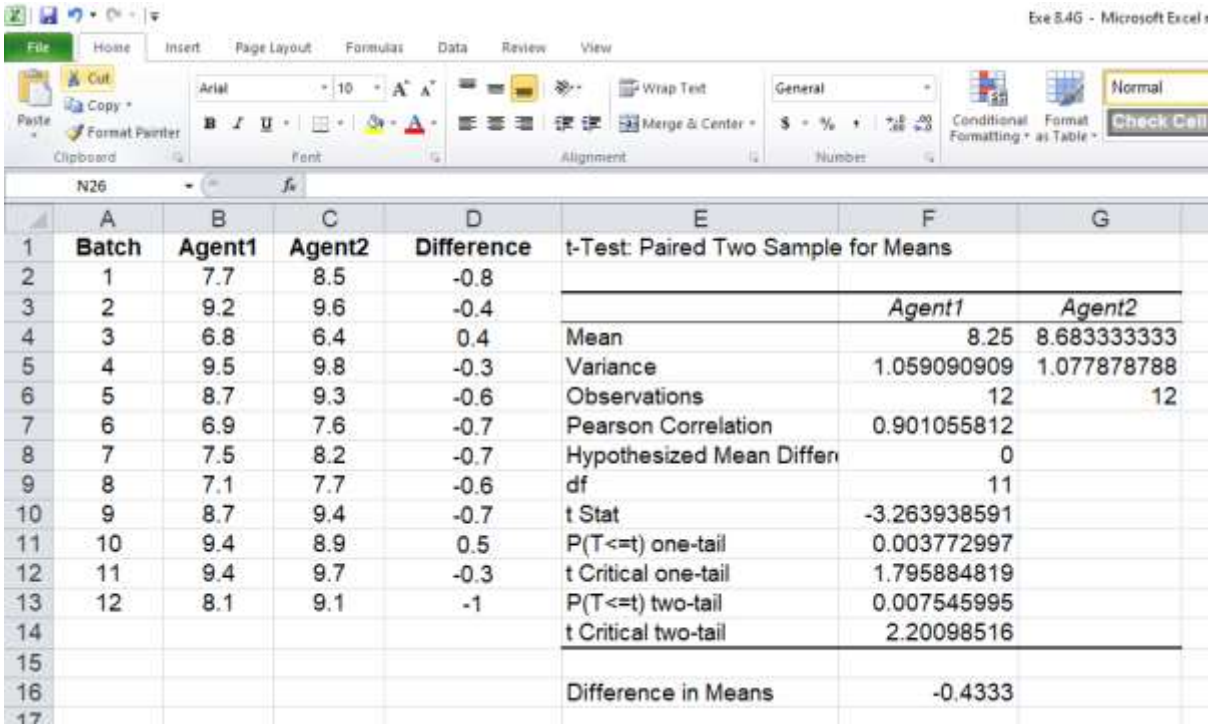
In summary, the patterns of brand preferences differ between the two demographic areas. In Area 1, there is a significant preference for "Other" breakfast cereals, with neither Brand A nor Brand B dominating. In contrast, in Area 2, there is a stronger preference for both Brand A and Brand B, but "Other" cereals still have a significant share of preferences. These patterns suggest that consumer preferences for breakfast cereals vary between the two areas, with Area 2 showing a relatively higher preference for Brands A and B.

Exercise 8.4

Consider the filtration data of Data Set G. Open the Excel workbook **Exe8.4G.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.

Solution 8.4:



	A	B	C	D	E	F	G
1	**Batch**	**Agent1**	**Agent2**	**Difference**	t-Test: Paired Two Sample for Means		
2	1	7.7	8.5	-0.8			
3	2	9.2	9.6	-0.4		*Agent1*	*Agent2*
4	3	6.8	6.4	0.4	Mean	8.25	8.683333333
5	4	9.5	9.8	-0.3	Variance	1.059090909	1.077878788
6	5	8.7	9.3	-0.6	Observations	12	12
7	6	6.9	7.6	-0.7	Pearson Correlation	0.901055812	
8	7	7.5	8.2	-0.7	Hypothesized Mean Differ	0	
9	8	7.1	7.7	-0.6	df	11	
10	9	8.7	9.4	-0.7	t Stat	-3.263938591	
11	10	9.4	8.9	0.5	P(T<=t) one-tail	0.003772997	
12	11	9.4	9.7	-0.3	t Critical one-tail	1.795884819	
13	12	8.1	9.1	-1	P(T<=t) two-tail	0.007545995	
14					t Critical two-tail	2.20098516	
15							
16					Difference in Means	-0.4333	

The last step in manufacturing a chemical product is the filtration process, designed to eliminate unwanted side products and impurities. The production manager aimed to assess the efficiency of two potential filter agents, referred to as Agent 1 and Agent 2. A total of 12 product batches were prepared, with each batch divided into two halves. One half was subjected to filtration using Agent 1, while the other half underwent filtration using Agent 2. After filtration, the remaining impurity level in the product was measured in parts per 1000 by weight.Impurity data between two filtration agents 1 and 2

Batch	Agent1	Agent2	Difference(Agent1-Agent2)
1	7.7	8.5	-0.8
2	9.2	9.6	-0.4
3	6.8	6.4	0.4
4	9.5	9.8	-0.3
5	8.7	9.3	-0.6
6	6.9	7.6	-0.7
7	7.5	8.2	-0.7
8	7.1	7.7	-0.6
9	8.7	9.4	-0.7
10	9.4	8.9	0.5
11	9.4	9.7	-0.3
12	8.1	9.1	-1

t-Test: Paired Two Sample for Means		
	Agent1	*Agent2*
Mean	8.25	8.683333333
Variance	1.059090909	1.077878788
Observations	12	12
Pearson Correlation	0.901055812	
Hypothesized Mean Difference	0	
df	11	
t Stat	-3.263938591	
P(T<=t) one-tail	0.003772997	
t Critical one-tail	1.795884819	
P(T<=t) two-tail	0.007545995	
t Critical two-tail	2.20098516	
Difference in Means	-0.4333	

Null Hypothesis (H0): The null hypothesis is that there is no significant difference in the population means of impurity levels between the two filtration agents. H0: Mean Agent1 = MeanAgent2

Alternative Hypothesis (H1): The alternative hypothesis is there is a significant difference in the population means of impurity levels between the two filtration

agents. H1: either Agent1 has a higher mean impurity level than Agent2 or that Agent1 has a lower mean impurity level than Agent2.

p-Values:

The two-tailed p-value is 0.0075

Significance Level (α)= 0.05.

The two-tailed p-value (0.0075) is less than α, indicating statistical significance for a two-tailed test. In this case, it suggests that there is evidence of a significant difference in impurity levels between the two agents hence we reject the null hypothesis

The data shows that the mean impurity of Agent1 is less than that of Agent2 which suggests that Agent1 should be preferred.

Exercise 8.5
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?

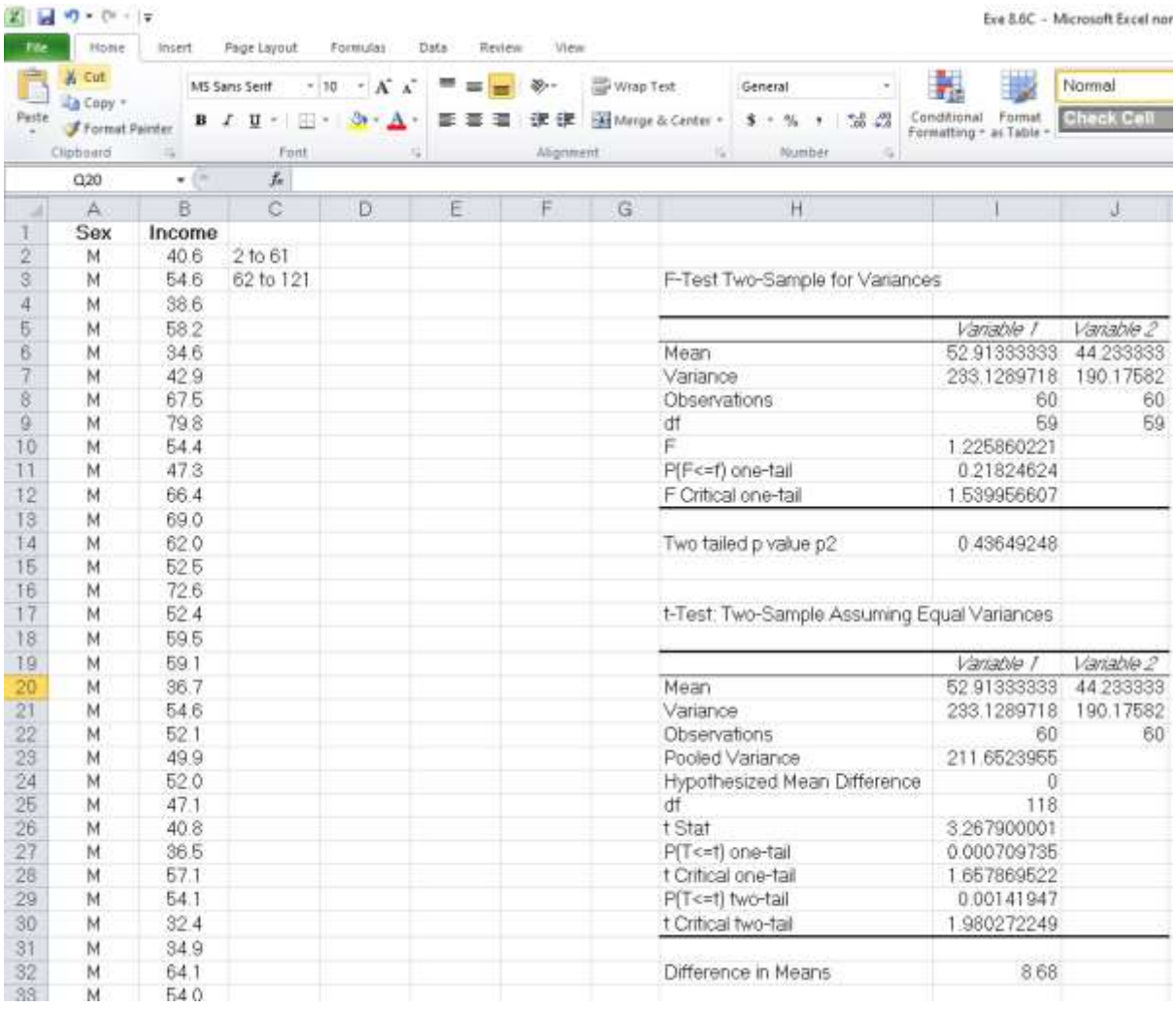
Solution 8.5: If a one-tailed test had been conducted specifically to determine whether filter agent 1 was more effective than filter agent 2 in reducing impurity levels, the conclusion would depend on the difference in sample means which shows that mean impurities of Agent 1 is lesser by 0.4333 than that of Agent2, which means Agent1 is more effective.

Exercise 8.6

Consider the bank cardholder data of Data Set C. Open the Excel workbook **Exe8.6C.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?

Solution 8.6:



	A	B	C
1	Sex	Income	
2	M	40.6	2 to 61
3	M	54.6	62 to 121
4	M	38.6	
5	M	58.2	
6	M	34.6	
7	M	42.9	
8	M	67.5	
9	M	79.8	
10	M	54.4	
11	M	47.3	
12	M	66.4	
13	M	69.0	
14	M	62.0	
15	M	52.5	
16	M	72.6	
17	M	52.4	
18	M	59.5	
19	M	59.1	
20	M	36.7	
21	M	54.6	
22	M	52.1	
23	M	49.9	
24	M	52.0	
25	M	47.1	
26	M	40.8	
27	M	36.5	
28	M	57.1	
29	M	54.1	
30	M	32.4	
31	M	34.9	
32	M	64.1	
33	M	54.0	

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	52.91333333	44.233333
Variance	233.1289718	190.17582
Observations	60	60
df	59	59
F	1.225860221	
P(F<=f) one-tail	0.21824624	
F Critical one-tail	1.539956607	

Two tailed p value p2 0.43649248

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	52.91333333	44.233333
Variance	233.1289718	190.17582
Observations	60	60
Pooled Variance	211.6523955	
Hypothesized Mean Difference	0	
df	118	
t Stat	3.267900001	
P(T<=t) one-tail	0.000709735	
t Critical one-tail	1.657869522	
P(T<=t) two-tail	0.00141947	
t Critical two-tail	1.980272249	

Difference in Means 8.68

The bank cardholder data has income of male and female employees in thousands of pounds. Since we are comparing separate samples male vs female, independent samples t test is appropriate to perform.

The independent samples t test result is

F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	52.913333	44.23333
Variance	233.12897	190.1758
Observations	60	60
df	59	59
F	1.2258602	
P(F<=f) one-tail	0.2182462	
F Critical one-tail	1.5399566	
Two tailed p value p2	0.4365	

The variance in income in males = 233.12897 and in females = 190.1758. The observed F test statistic is F= 1.2259 with 59 and 59 associated degrees of freedom, giving a two tailed p value of p = 0.4365

The observed F ratio is not significant. The data are consistent with the assumption that the population variances in the income of male and female employee do not differ, and we therefore proceed to use the equal variances form of the unrelated samples t test.

To test if the population mean income differ between male and female employees, a two-tailed t test is performed.

Two-tailed t test results:

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	*Variable 2*
Mean	52.91333333	44.2333333
Variance	233.1289718	190.175819
Observations	60	60
Pooled Variance	211.6523955	
Hypothesized Mean Difference	0	
df	118	
t Stat	3.267900001	
P(T<=t) one-tail	0.000709735	
t Critical one-tail	1.657869522	
P(T<=t) two-tail	0.00141947	
t Critical two-tail	1.980272249	
Difference in Means	8.68	

The obtained independent samples t = 3.3.2679 with 118 degrees of freedom.

The associated two-tailed p-value is p = 0.0014, so the observed t is significant at the 1% level (two-tailed).

The sample mean income for male and female employees were, £52,913 and £44,233 respectively.

The data therefore constitute strong evidence that the underlying mean income was higher for male employees, by an estimated £52,913 –£ 44,233 = £8680. The results strongly suggest that male employees' income is higher than that of female employees.