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?

Diet A

Diet A	n	50
	Mean	5.341
	SD	2.536

Diet B

Diet B	n	50
	Mean	3.710
	SD	2.769

Diet A has a higher average of weight loss compared to Diet B. This therefore suggests that Diet A is more effective than Diet B for weight loss. The standard deviation of Diet B is higher than Diet A which there for suggests that the results of the weight loss for the individuals on Diet B are spread out compared to Diet A (National Library of Medicine, 2023). As a result, individual result may vary more in Diet B than in Diet A.

Reference:

National Library of Medicine, 2023. *National Library of Medicine.* [Online]
Available at: https://www.nlm.nih.gov/oet/ed/stats/02-900.html#:~:text=A%20standard%20deviation%20(or%20%CF%83,data%20are%20more%20spread%20out.

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?

Diet A

Diet A	n	50
	Mean	5.341
	SD	2.536
	Median	5.642
	Q1	3.748
	Q3	7.033
	IQR	3.285

Diet B

Diet B	n	50
	Mean	3.710
	SD	2.769
	Median	3.745
	Q1	1.879
	Q3	5.404
	IQR	3.525

The average of Diet A (5.341) is higher than Diet B (3.710 kg) which suggests that Diet A is more effective in weight loss compared to Diet B. Both the diets demonstrate a comparable range of weight loss outcomes, with Diet B showing slightly higher variability. The IQR is similar between both diets with a slight difference of 0.24kg where Diet B is slightly higher. As Diet B has a higher IQR than Diet A, it indicates that the data is slightly further spread out compared to Diet A (Frost, 2022).

Reference:

Frost, J., 2022. *Interquartile Range (IQR): How to Find and Use It.* [Online]
Available at: https://statisticsbyjim.com/basics/interquartile-range/
[Accessed 28 September 2023].

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 Area 1 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?

Frequencies

	Area 1	Area 2
A	11	19
B	17	30
Other	42	41
Total	70	90

Percentage

	Area 1	Area 2
A	15.7	21.1
B	24.3	33.3
Other	60.0	45.6
Total	100	100

The results from the two demographics show that Brand B is more popular than Brand A. Whereas the other brands are higher in Area 1 than in Area 2.

In Area 1, there were a total of 70 respondents in which 11 (15.7%) preferred Brand A, 17 (24.3%) preferred Brand B, and the remaining 42 (60%) preferred some other brand of breakfast cereal.

In Area 2, there were a total of 90 respondents in which 19 (21.1%) preferred Brand A, 30 (33.3%) preferred Brand B, and the remaining 41 (45.6%) preferred some other brand of breakfast cereal.

Exercise 8.4

Consider the filtration data of Data Set G. Open workbook **Exe8.4G.xlsx** which contains these data from the Exercises folder.

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

Data

Batch	Agent1	Agent2
1	7.7	8.5
2	9.2	9.6
3	6.8	6.4
4	9.5	9.8
5	8.7	9.3
6	6.9	7.6
7	7.5	8.2
8	7.1	7.7
9	8.7	9.4
10	9.4	8.9
11	9.4	9.7
12	8.1	9.1

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	-0.433333333	

The mean values for Agent 1 is 8.25 and Agent 2 is 8.68 where the difference between the means is -0.43. This indicates that Agent 1 has a lower mean impurity compared to Agent 2. As the Pearson Correlation value is 0.90, there is a strong positive linear relationship measured by the 2 agents, Agent 1, and Agent 2.

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?

When testing whether Filter Agent 1 is more effective than Filter Agent 2, the hypotheses would be:

- H_0: $\mu_1 \geq \mu_2$
- H_1: $\mu_1 < \mu_2$

Results from two-tailed test from **Exercise 8.4**

t Stat	-3.263938591
One Tail p-value	0.003772997
t Critical one-tail	1.795884819

The mean of Agent 1 is 8.25 and the mean of Agent 2 is 8.68 which indicates that the data is consistent with H_1. The associated p-value is 0.003 which is less than 0.05 (5%) while the t stat value is less than the t critical. if a one-tailed test had been conducted to determine whether Filter Agent 1 was more effective, H0 would be rejected in the favour of H1. As a result, my conclusion would be that there is appropriate evidence to suggest that Filter Agent 1 is more effective than Filter Agent 2 at the 0.05 significance level (Zach, 2022).

Reference:

Zach, 2022. [Online]

Available at: https://www.statology.org/one-tailed-test-example-problems/#:~:text=Example%201%3A%20Factory%20Widgets&text=However%2C%20one%20engineer%20believes%20that,Hypothesis)%3A%20%CE%BC%20%3C%2020%20grams

[Accessed 28 September 2023].

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?

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	52.91333333	44.23333333
Variance	233.1289718	190.1758192
Observations	60	60
df	59	59
F	1.225860221	
P(F<=f) one-tail	0.21824624	
F Critical one-tail	1.539956607	
p2	2.451720442	

t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	52.91333333	44.23333333
Variance	233.1289718	190.1758192
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 Mean	8.68	

I first performed the F-Test Two-Sample for Variances which the value of p-value for one-tail was 0.21 which is greater than 0.05. Therefore, there was no conclusive

evidence to reject H_0. As a result, I conducted the t-Test Two Sample Assuming Equal Variances.

The t-statistic value is 3.267 with 118 degrees of freedom which indicates that the two groups are different. The difference of the mean is 8.68 which therefore indicates that on average, Men have a higher income compared to Women.