Worksheets Exercises

Unit 8

Exercise 8.1 – Sample size, Mean, Standard Deviation

Diet	Wtloss				
A	3.709				
A	7.087		Diet A	n	50
A	6.754			Mean	5.341
A	8.994			SD	2.536
A	9.077				
A	6.413				
A	5.877				
A	2.572				
A	7.520				
A	6.881				
A	7.265				
A	3.477				
A	3.755				
A	8.760				
A	7.032				
A	9.052				
A	10.062				
A	4.840				
A	6.449				
A	9.019				
A	-1.715				
A	4.718		Diet B	n	50
A	4.007			Mean	3.710
A	7.241			SD	2.769
A	2.128				

To find the sample size → =COUNT(B52:B101).

To find the sample mean → =AVERAGE(B52:B101).

To find the standard deviation → =STDEV(B52:B101).

The sample mean weight loss for Diet B is $\bar{x}$ = 3.710. The average weight loss for those individuals who undertook Diet B is 3.710 kg, so the diet appears to have been effective.

The sample standard deviation of the weight loss for Diet B is s = 2.769 kg. Since the mean weight loss is a little larger than 1 standard deviation, then a high proportion of those individuals on Diet B had a positive weight loss, again emphasising the effectiveness of the diet.

Both diets have the same sample size, hence they can be compared. The mean of diet A is higher than the mean of diet B, therefore diet A is more effective. Additionally, diet's A s.d. is smaller than diet's B s.d., meaning that diet's A results are closer to their mean.

Exercise 8.2 – Median, Quartiles, Interquartile Range

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
30	A	2.680				

To find the median → =MEDIAN(B52:B101).

To find the 1st quartile → =QUARTILE(B52:B101,1).

To find the 2nd quartile → =QUARTILE(B52:B101,3).

To find the interquartile range → =F28-F27 (Note: Q3-Q1).

The sample median weight loss for Diet B is M = 3.745 kg, therefore the diet appears to have been effective.

The sample IQR of the Diet B weight loss is 3.451 kg. A several individuals on Diet B had a positive weight loss. The median of diet A is higher than the median of diet B, therefore diet A seems to be more effective. Additionally, diet’s A IQR. is smaller than diet’s B IQR, meaning that diet’s A results are closer with each other.

Exercise 8.3 – Conditional Counting and Sums

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

For frequencies:

=COUNTIF(B72:B161,"A") + same for B and Other by replacing A with B and Other, respectively.

Total → =SUM(F6:F8)

For percentages:

=100*F6/F$9 + Dragging the formula for the other two brands.

Total → =SUM(F15:F17)

The least favourite cereal brand in both areas is A. The most favourable cereal brand in both areas is cereal falling into the Other category. In Area 2 more people prefer cereal A or B compared to Area 1, whereas in Area 1 more people prefer Other compared to area 2.

Exercise 8.4 – Two-tailed t-test

	A	B	C	D	E	F	G	H
1	**Batch**	**Agent1**	**Agent2**		t-Test: Paired Two Sample for Means			
2	1	7.7	8.5					
3	2	9.2	9.6			*Agent1*	*Agent2*	
4	3	6.8	6.4		Mean	8.25	8.683333333	
5	4	9.5	9.8		Variance	1.059090909	1.077878788	
6	5	8.7	9.3		Observations	12	12	
7	6	6.9	7.6		Pearson Correlation	0.901055812		
8	7	7.5	8.2		Hypothesized Mean Difference	0		
9	8	7.1	7.7		df	11		
10	9	8.7	9.4		t Stat	-3.263938591		
11	10	9.4	8.9		P(T<=t) one-tail	0.003772997		
12	11	9.4	9.7		t Critical one-tail	1.795884819		
13	12	8.1	9.1		P(T<=t) two-tail	0.007545995		
14					t Critical two-tail	2.20098516		
15								
16					Difference in Means	-0.433333333		
17								
18								
19								

The obtained related samples t = -3.264 with 11 degrees of freedom.

The associated two-tailed p-value is p = 0.008, so the observed t is significant at the 5% level (two-tailed).

The sample mean numbers of Agents 1 and 2 were, respectively 8.250 and 8.683. Hence, the data constitute significant evidence that the underlying mean number of agents was greater for Agent 2, by an estimated 8.683 – 8.250 = 0.433 agents. The results suggest that Agent 2 is preferable.

Exercise 8.5 – One-tail t-test

	A	B	C	D	E	F	G	H
1	**Batch**	**Agent1**	**Agent2**		t-Test: Paired Two Sample for Means			
2	1	7.7	8.5					
3	2	9.2	9.6			*Agent1*	*Agent2*	
4	3	6.8	6.4		Mean	8.25	8.683333333	
5	4	9.5	9.8		Variance	1.059090909	1.077878788	
6	5	8.7	9.3		Observations	12	12	
7	6	6.9	7.6		Pearson Correlation	0.901055812		
8	7	7.5	8.2		Hypothesized Mean Difference	0		
9	8	7.1	7.7		df	11		
10	9	8.7	9.4		t Stat	-3.263938591		
11	10	9.4	8.9		P(T<=t) one-tail	0.003772997		
12	11	9.4	9.7		t Critical one-tail	1.795884819		
13	12	8.1	9.1		P(T<=t) two-tail	0.007545995		
14					t Critical two-tail	2.20098516		
15								
16					Difference in Means	-0.433333333		
17								
18								
19								

H_0: $\mu_1 \geq \mu_2$

H_1: $\mu_1 < \mu_2$

The sample mean numbers of Agent 1 and 2 were, respectively 8.250 and 8.683, so that the data are indeed consistent with H_1 .

As before, the obtained related samples t = -3.264 with 11 degrees of freedom.

The associated one-tailed p-value is p = 0.004, so the observed t is significant at the 1% level (one-tailed).

The data therefore constitute strong evidence (on a one-tailed test) that the underlying mean number of agents was greater for Agent 2, by an estimated 8.683 – 8.250 = 0.433 agents. The results continue to suggest that Agent 2 should be preferred.

Although broadly similar conclusions were reached as before, a higher level of significance was obtained with the one-tailed test.

Exercise 8.6 – F-test (Part 1)

E18 fx

	A	B	C	D	E	F	G	H	I	J	K
1	**Sex**	**Income**									
2	M	40.6									
3	M	54.6						F-Test Two-Sample for Variances			
4	M	38.6									
5	M	58.2							*Variable 1*	*Variable 2*	
6	M	34.6						Mean	52.91333333	44.23333333	
7	M	42.9						Variance	233.1289718	190.1758192	
8	M	67.5						Observations	60	60	
9	M	79.8						df	59	59	
10	M	54.4						F	1.225860221		
11	M	47.3						P(F<=f) one-tail	0.21824624		
12	M	66.4						F Critical one-tail	1.539956607		
13	M	69.0									
14	M	62.0						p2	0.43649248		
15	M	52.5									
16	M	72.6									
17	M	52.4									

The observed F test statistic is F = 1.226 with 59 and 59 associated degrees of freedom, giving a two tailed p-value of $p = 0.436^{NS}$. The observed F ratio is *not significant*.

Exercise 8.6 – Two-tail t-test (Part 2)

14	M	62.0						p2	0.43649248		
15	M	52.5									
16	M	72.6									
17	M	52.4						t-Test: Two-Sample Assuming Equal Variances			
18	M	59.5									
19	M	59.1							*Variable 1*	*Variable 2*	
20	M	36.7						Mean	52.91333333	44.23333333	
21	M	54.6						Variance	233.1289718	190.1758192	
22	M	52.1						Observations	60	60	
23	M	49.9						Pooled Variance	211.6523955		
24	M	52.0						Hypothesized Mean Difference	0		
25	M	47.1						df	118		
26	M	40.8						t Stat	3.267900001		
27	M	36.5						P(T<=t) one-tail	0.000709735		
28	M	57.1						t Critical one-tail	1.657869522		
29	M	54.1						P(T<=t) two-tail	0.00141947		
30	M	32.4						t Critical two-tail	1.980272249		
31	M	34.9									
32	M	64.1						Difference in Means	8.68		
33	M	54.0									
34	M	51.5									

The obtained independent samples t = 3.268 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 are, respectively, 52.913 and 44.233.

The data therefore constitute strong evidence that the underlying mean income was greater for male, by an estimated 52.913 – 44.233 = 8.68. The results strongly suggest that male income is higher.

Unit 9

Exercise 9.1 – Clustered Column Chart with One Variable

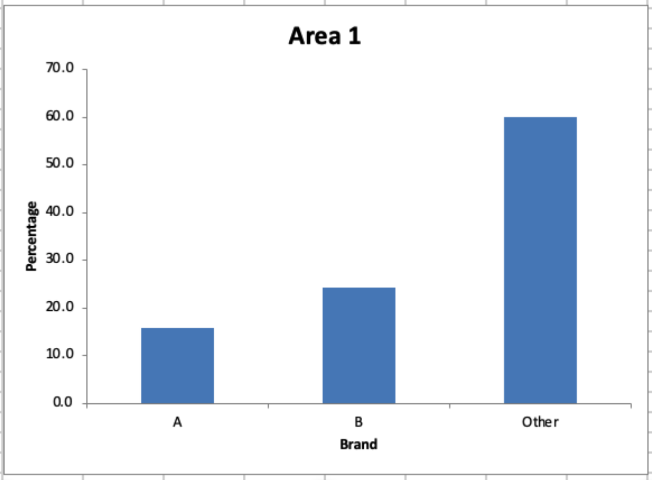
	Area	Brand
1	Area	Brand
2	1	B
3	1	Other
4	1	A
5	1	B
6	1	Other
7	1	A
8	1	Other
9	1	Other
10	1	Other
11	1	Other
12	1	B
13	1	Other
14	1	Other
15	1	A
16	1	A
17	1	A
18	1	B
19	1	A
20	1	Other
21	1	B
22	1	A
23	1	B
24	1	Other
25	1	Other
26	1	B
27	1	B
28	1	Other
29	1	Other
30	1	Other
31	1	Other
32	1	Other
33	1	B
34	1	B
35	1	Other
36	1	Other
37	1	B
38	1	B
39	1	B
40	1	Other
41	1	Other
42	1	B
43	1	Other
44	1	Other
45	1	Other
46	1	Other
47	1	Other
48	1	Other
49	1	Other
50	1	Other
51	1	Other
52	1	A
53	1	Other

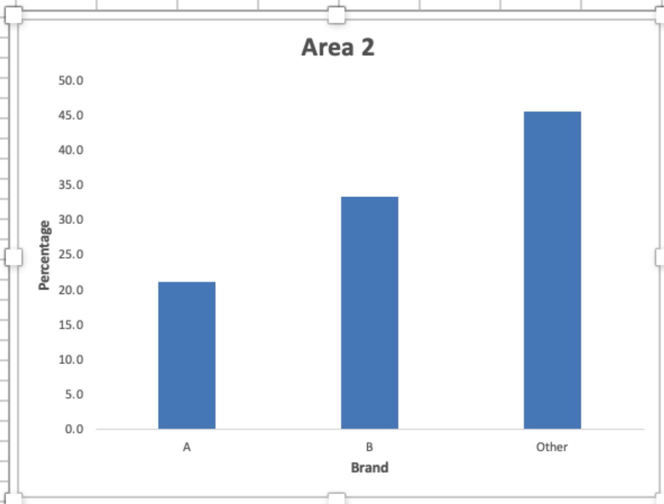
Frequencies

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

Percentages

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





Results are the same as in Area 1. In Area 2 the least favourite brand is A, followed by brand B, while the majority of the population in Area 2 prefers some other brand.

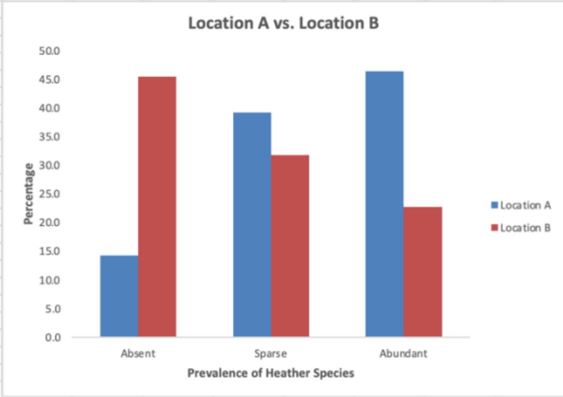
Exercise 9.2 – Clustered Column Chart with Two Variables

Frequencies

	Location A	Location B
Absent	8	20
Sparse	22	14
Abundant	26	10
Total	**56**	**44**

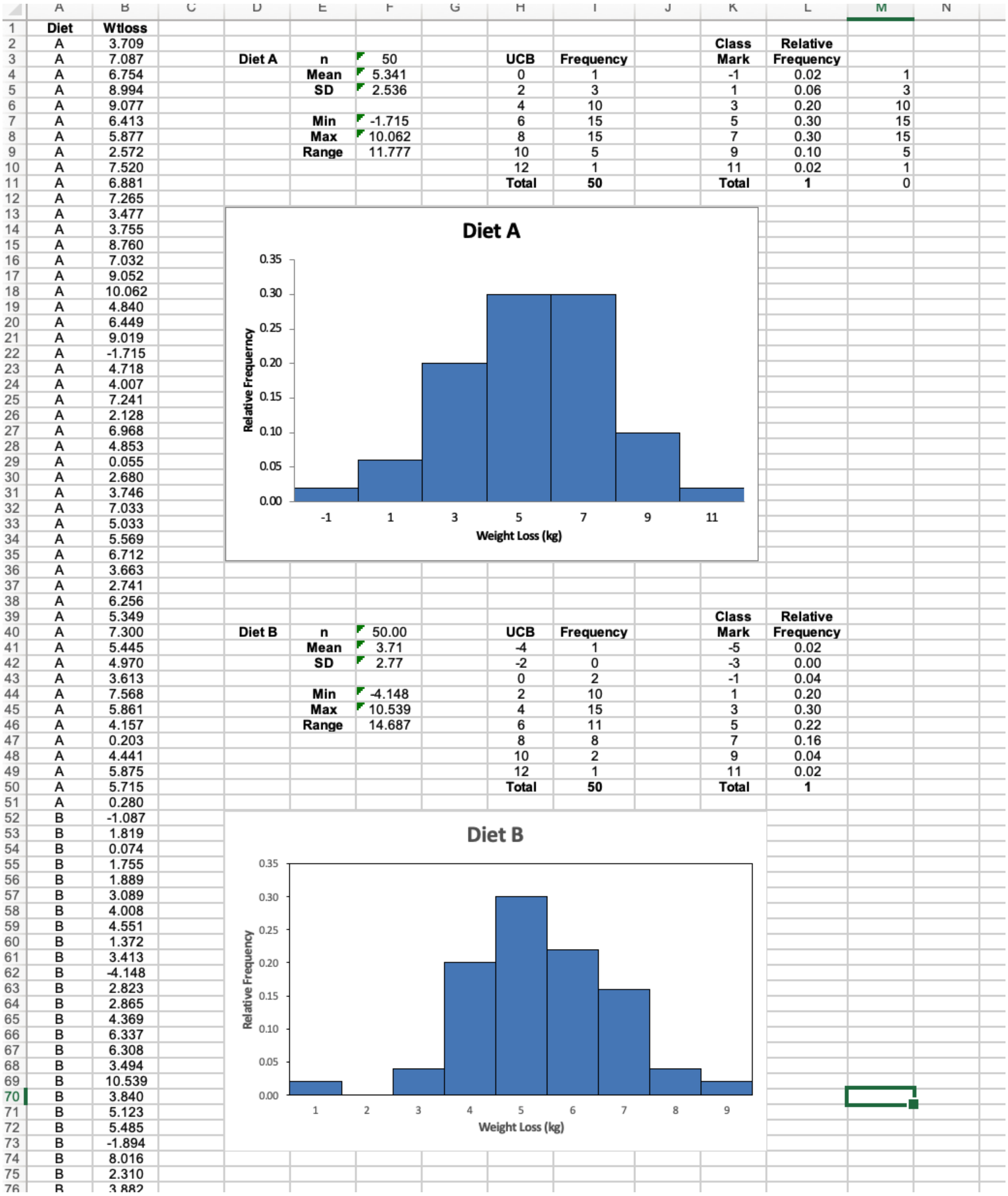
Percentages

	Location A	Location B
Absent	14.3	45.5
Sparse	39.3	31.8
Abundant	46.4	22.7
Total	**100**	**100**



The results of the two locations are opposite with each other. In location A heather species are mostly abundant and the lowest percentage belongs to the absent category. On the contrary, in location B the heather species are mostly absent, with the lowest percentage reflecting abundant heather species.

Exercise 9.3 - Histogram



	Diet	Wtloss
1	Diet	Wtloss
2	A	3.709
3	A	7.087
4	A	6.754
5	A	8.994
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
24	A	4.007
25	A	7.241
26	A	2.128
27	A	6.968
28	A	4.853
29	A	0.055
30	A	2.680
31	A	3.746
32	A	7.033
33	A	5.033
34	A	5.569
35	A	6.712
36	A	3.663
37	A	2.741
38	A	6.256
39	A	5.349
40	A	7.300
41	A	5.445
42	A	4.970
43	A	3.613
44	A	7.568
45	A	5.861
46	A	4.157
47	A	0.203
48	A	4.441
49	A	5.875
50	A	5.715
51	A	0.280
52	B	-1.087
53	B	1.819
54	B	0.074
55	B	1.755
56	B	1.889
57	B	3.089
58	B	4.008
59	B	4.551
60	B	1.372
61	B	3.413
62	B	-4.148
63	B	2.823
64	B	2.865
65	B	4.369
66	B	6.337
67	B	6.308
68	B	3.494
69	B	10.539
70	B	3.840
71	B	5.123
72	B	5.485
73	B	-1.894
74	B	8.016
75	B	2.310
76	B	3.882

Diet A		
n	50
Mean	5.341
SD	2.536
Min	-1.715
Max	10.062
Range	11.777

UCB	Frequency	Class Mark	Relative Frequency	
0	1	-1	0.02	1
2	3	1	0.06	3
4	10	3	0.20	10
6	15	5	0.30	15
8	15	7	0.30	15
10	5	9	0.10	5
12	1	11	0.02	1
Total	50	Total	1	0

Diet B	
n	50.00
Mean	3.71
SD	2.77
Min	-4.148
Max	10.539
Range	14.687

UCB	Frequency	Class Mark	Relative Frequency
-4	1	-5	0.02
-2	0	-3	0.00
0	2	-1	0.04
2	10	1	0.20
4	15	3	0.30
6	11	5	0.22
8	8	7	0.16
10	2	9	0.04
12	1	11	0.02
Total	50	Total	1

For individuals who underwent diet B, the weight loss distribution is unimodal and rather symmetrical, with no skew.