Inferential Statistics Exercises

8.1

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?

The mean for diet A is higher than diet B, implying that when people did lose weight, losses were higher on diet A than B. Standard deviation to mean ratio was lower for diet B which I am interpreting as weight loss results were less consistent. 28 people lost more than the mean on diet A, 25 lost more than the mean on diet B, a 6% difference yet the mean to stdev ratio was 2.1 to 1.34, considerably different.

8.2

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?

Those on diet A did lose about 50% more weight than those on diet B but both diets did work for those that experienced wight loss. The median and mean values for both diet A and diet B are close, 5.341/5.642 and 3.710/3.745 implying there are not likely a large number of outliers in the data, or weight loss is fairly consistent.

8.3

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?

60% of people in area 1 preferred a cereal other than brands A or B, in area 2 Brand B has 33% market share and 55% of people in area two preferred brands B or A over any other choice. Brand B is consistently more popular in both areas.

8.4

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.

Referencing the text from example *8.4* *"*The associated two-tailed p-value is p = 0.018, so the observed t is significant at the 5% level (two-tailed)."

There is no explanation in the exercise notes why 0.018 is significant at the 5% level, the assumption is the student understands statistics terms and concepts which isn't necessarily the case, so this is purely a guess.

- *Is the two tail P value less than 0.05?* yes

- *Are the filtration agents different?* I have no idea because the difference in means is minimal, whereas the example worked through with the container displays was almost 10% of the sales volume. Not sure if I am even looking in the right area.

8.5

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?

Similar explanation as in 8.4,

- Is the one tail P value less than 0.01, yes, it is 0.0038%
- Unsure how this relates to the difference in means which is very low so I am still assuming this implies the two agents are roughly the same.

8.6

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?

The Null hypothesis would be true if the mean income for male cardholders exceeds the mean income for female cardholders.

Averaging the mean incomes of each sample groups shows arithmetically that the male sample group had a higher income, 52,900 pounds VS. 44,200 pounds.

There was considerably more variance between high and low incomes in the male sample group which could be skewing the mean higher. As a result, the median was calculated for each sample identifying that both sample groups were positively skewed so the sample sets have similar shape. With similar sample shapes the mean value can be considered a strong indicator of average annual income.

Male Cardholders	**n**	60
	Mean	52.9
	SD	15.26856
	Median	52.1
Female Cardholders	**n**	60
	Mean	44.2
	SD	13.79042
	Median	38.2

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	52.91333	44.23333
Variance	233.129	190.1758
Observations	60	60
df	59	59
F	1.22586	
P(F<=f) one-tail	0.218246	
F Critical one-tail	1.539957	
P2	0.436492	

A second test was performed, t-Test Two-Sample Assuming Equal Variances. According to the course textbook, "Basic Business Statistics", if the t Stat value exceeds the t Critical value the null hypothesis can not be rejected. A review of the tested output below confirms this to be the case, further evidence that the mean value assessment in the first test is correct.

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	
P2	0.002838941	