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?

Answer:

Diet A

size = 50, mean = 5.341, SD = 2.536

Diet B

size = 50, mean = 3.710, SD = 2.769

Comparing Diet A and Diet B, we can infer that Diet A may be more effective for weight loss. With the same sample size, Diet A has a higher average weight loss, and its standard deviation is smaller than that of Diet B. This suggests that individuals on Diet A have a narrower range of weight loss variability compared to those on Diet B. In other words, individuals following Diet A are more likely to achieve weight loss closer to the average weight loss of Diet A.

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?

Answer:

Diet A

size = 50, mean = 5.341, SD = 2.536, Median = 5.642, Q1 = 3.748, Q3 = 7.033, IQR = 3.285

Diet B

size = 50, mean = 3.710, SD = 2.769, Median = 3.745, Q1 = 1.935, Q3 = 5.404, IQR = 3.451

Both Diet A and Diet B have medians greater than their means, indicating that their data distributions are left-skewed.

Comparing Diet A and B, Diet A has a higher mean weight loss, suggesting that it is more effective in promoting weight loss than Diet B. Diet A's IQR is shorter than Diet B's, which implies that Diet A has less variability in weight loss results than Diet B. In other words, Diet A may provide more consistent and stable weight loss outcomes than Diet B.

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?

Answer:

area 1: A = 11 (15.7%), B = 17 (24.3%), Other = 42 (60%), Total = 70

area 2: A = 19 (21.1%), B = 30 (33.3%), Other = 41 (45.6%), Total = 90

In the first area, 60% of the people prefer "Other" brand of breakfast cereal, while 15.7% prefer brand "A" and 24.3% prefer brand "B". The total number of people surveyed in this area is 70.

In the second area, 45.6% of the people prefer "Other" brand of breakfast cereal, while 21.1% prefer brand "A" and 33.3% prefer brand "B". The total number of people surveyed in this area is 90.

It can be noted that in Area 2, more than half of the people prefer either brand "A" or brand "B". Out of these two, brand "B" is preferred by a higher percentage of 33.3% compared to 21.1% for brand "A". This gap is wider compared to the preference for these two brands in Area 1.

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.

Answer:

t-Test: Paired Two Sample for Means		
	Agent1	*Agent2*
Mean	8.25	8.683333
Variance	1.059091	1.077879
Observations	12	12
Pearson Correlation	0.901056	
Hypothesized Mean Difference	0	
df	11	
t Stat	-3.26394	
P(T<=t) one-tail	0.003773	
t Critical one-tail	1.795885	
P(T<=t) two-tail	0.007546	
t Critical two-tail	2.200985	
Difference in Means	-0.43333	

The data set shows a strong positive linear relationship between the amount of impurity remaining after filtration with Agent 1 and Agent 2, with a Pearson correlation coefficient of 0.901055812.

The null hypothesis (H0) in the paired two-sample t-test assumes that there is no significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, while the alternative hypothesis (Ha) assumes that there is a significant difference between the two means.

Both the p-value (0.007545995 for the two-tailed test and 0.003772997 for the one-tailed test) is less than 0.05 and the absoluate of t-statistic is greater than t critical value, and we can make rejection of the null hypothesis. Therefore, we can conclude that there is a statistically significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, and Agent 1 is more effective in removing impurities than Agent 2.

The data set shows a strong positive linear relationship between the amount of impurity remaining after filtration with Agent 1 and Agent 2, with a Pearson correlation coefficient of 0.901055812.

The null hypothesis (H0) in the paired two-sample t-test assumes that there is no significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, while the alternative hypothesis (Ha) assumes that there is a significant difference between the two means.

Both the p-value (0.007545995 for the two-tailed test and 0.003772997 for the one-tailed test) is less than 0.05 and the absoluate of t-statistic is greater than t critical value, and we can make rejection of the null hypothesis. Therefore, we can conclude that there is a statistically significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, and Agent 1 is more effective in removing impurities than 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?

Answer:

The null hypothesis (H0) in the paired two-sample t-test assumes that there is no significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, while the alternative hypothesis (Ha) assumes that there is a significant difference between the two means.

Based on the provided t-test analysis, the one-tailed p-value of 0.003772997 is less than the significance level of 0.05, indicating strong evidence against the null hypothesis. Additionally, the calculated t-statistic of -3.263938591 is less than the critical t-value of 1.795884819, further supporting the rejection of the null hypothesis.

Therefore, we can conclude that there is a statistically significant difference between the mean amount of impurity remaining in the product after filtration with Agent 1 and Agent 2, with Agent 1 having a more effective filtering process. The difference in means of -0.4333 further confirms this conclusion.

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?

Answer:

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	0.43649248	

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 Means	8.68	

	Male
Mean	52.91333333
Standard Error	1.97116282
Median	52.05
Mode	54.6
Standard Deviation	15.26856155
Sample Variance	233.1289718
Kurtosis	0.470640827
Skewness	0.724943349
Range	69.9
Minimum	31
Maximum	100.9
Sum	3174.8
Count	60

	Female
Mean	44.23333333
Standard Error	1.780336201
Median	38.15
Mode	33.4
Standard Deviation	13.79042491
Sample Variance	190.1758192
Kurtosis	0.35112373
Skewness	1.09975931
Range	52.9
Minimum	30
Maximum	82.9
Sum	2654
Count	60

Because the t Stat (3.2679) is greater than the t Critical one-tail (1.65787), we can reject the null hypothesis. We can conclude that there is a significant difference in income between males and females in the population, with males having higher incomes than females.

We made some assumptions for this analysis, including that the data is normally distributed and that the samples are unbiased. We performed an F-Test Two-Sample for Variances to validate the assumption of equal variances between the two groups. With a p-value of 0.21824624, which is greater than the 0.05 significance level, we can assume that the variances of the two groups are equal.

To further assess the normality of the data, we can examine the descriptive statistics (you can see them as below). Kurtosis and skewness values are relatively close to those expected for a normal distribution, but it would be more appropriate to perform a normality test, such as visually inspecting the data using a histogram to confirm normality.