**QMM 1001 Case Study 1 [20%]**

**Name Here**

**GOODNEWS AGBADU**

**Student Number Here**

**A00238219**

**Due: October 22, 2021, at 11:59 PM**

# Introduction

For this case study, personal data was about how I spent my day as an international student in Canada, I order to collect these data, I used a personalized data sheet to record my daily activities from September 10th, 2021 to October 15th, 2021. I assigned a total of nine (9) variables such as; Date, hours spent on zoom for virtual learning, hours it takes to sleep each day, how often I left the house, time spent on news, time spent on playing games, how stress I am each day and the count of calls I receive each day. To record my variables, I used a scale of 5-1 to measure how stressed I was daily and for Zoom, Study, Sleep and House, I used numbers in hours from 0-9 to measure the amount of time spent on each variable. While for Calls, I used a count from 0-7 to measure how many times I left my home. Finally, I used “yes” and “No” to measure the categorical variable Game and News.

Having a glance through my collected personalized data, I started with recording four (4) quantitative variables. Usually, I always feel that after my Zoom class, my day seems to have come to an end. I want to understand what portion of my day was spent on classes as well as studying. Remembering I had to sleep each day gives me joy because it enables me to relax my mind and be prepared for the day ahead. One thing I notice about myself is that I do not leave my house unnecessarily which has affected my social life. The last quantitative variable I collected data on was how many calls I take each day. So it is interesting to see if I spend more time daily talking over the phone.

To better understand my daily activities, I collected two nominal variables as well as an ordinal categorical variable. The nominal variables were to understand if I listen to the news to familiarize myself with the political, economical, and social aspects of Canadian life and whether I play a game each day as I did when I was living in Nigeria. Finally, I am researching how stressed are international students during COVID, I used a categorical ordinal variable to represent my stress level and I am particularly interested to understand how stressed I am maybe due to prolonged zoom classes or less sleeping time.

To answer the question what do you do in a day? from my dataset, I observed from the variables that some activities such as reading, sleeping, taking calls, and stress are all part of my daily life while some such as zoom, games, leaving the house are sometimes not part of what I do daily. This is surprising because I notice I was stressed all which may be as a result of not leaving my house often to catch some fun and I would say I did not spend the day the way I expected it because I may plan to spend some time outside my house but I end up using up all the time for prolonged studies. At the end of this data analysis, I seek to have a better understanding of how these variables affect my lifestyle and the relationships between them.

# Data Analysis

**Categorical Variables:**

Goodnews Stress Level:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1: Goodnews Level of Stress in percentages:** | | | | | | |
| **Stress Level** | 0 | 1 | 2 | 3 | 4 | 5 |
| **Frequency** | 0% | 19.4% | 25% | 19.4% | 19.4% | 16.7% |

Looking at this nominal categorical variable, which shows my daily stress level. I used a scale of 0 to 5 to show the degrees of my stress, 0 represents days that I was not stressed while 5 represents days that I was highly stressed. As shown in **Table 1** and **Figure 1**, that I spent most of my day in the stress level with equal percentages in levels one, three, and four at 19.4% each. Though level five was the lowest with 16.7%, while level two was significant with 25%. Achieving a stress level one, two, and four at 19.4% may likely be associated with Wednesday zoom classes which have more than 7 hours in zoom time and time spent on assignments. The results of the analysis are important to me because I was stressed more, on the three extreme levels which may imply a health risk. I have only spent 0% of my time feeling no stress at level zero, which shows that I have not managed my time well during the data collection period and the need to make out more time to relax for a balanced lifestyle.

Chart, bar chart

Description automatically generated

Figure 1: Bar Plot Showing the Number my Stress Level

Time Spent Listening to News:

Chart, pie chart

Description automatically generatedThe second variable to be analyzed is an ordinal categorical variable which looks at the amount of time I spend listening to the news. To present responses, ‘yes’ were used to represent days, I listened to the news and ‘no’ for days in which I did not listen to the news. Table 2 and Figure 2 shows that I used 41.7% of the days to listen to the news while 58.3% shows days which I did not listen to the news. After a closer investigation on the analysis, I discovered from my data set that I listen to news mostly on weekends when I don’t have zoom classes.

|  |  |
| --- | --- |
| **Table 2: Time Spent Listening to News in Percentages:** | |
| **Yes** | **No** |
| 41.7% | 58.3% |

Also, I spend a lot of time on social media reading world headlines so if the frequency of hours spent on social media reading news were included, it would have significantly increased the amount of time spent listening to the news.

Figure 2: Frequency I Watched the News

Time Spent on Games:

|  |  |
| --- | --- |
| **Table 3: Time Spent on Games in Percentages:** | |
| **Yes** | **No** |
| 36.1% | 63.9% |

Chart, pie chart

Description automatically generatedThe third categorical variable is ordinal in Table 3 and Figure 3, which looked at the frequency that I played games with ‘yes’ at 36.1% response representing the amount of time spent on playing games while ‘no’ at 63.9% represents the amount of time I did not play a game in days. This result is not surprising because my preferred game on PlayStation is FIFA 2020 and I always like to play against a human opponent. As a result of this, I can only play my game when my friends are around thereby reducing the time of play

Figure 3: Frequency I Watched the News

**Contingency Table:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Table 4: Contingency Table Comparing Time Spent on Games and my stress Level** | | | | | | | |
| **Did I play a game?** | |  | | **Stress Level for game** | | | | |
| **1** | **2** | | **3** | **4** | **5** | **TOTAL** |
| **Yes** | | 2.8% | 11.1% | | 19.4% | 13.9% | 16.7% | **63.9%** |
| **No** | | 16.7% | 13.9% | | 0% | 5.6% | 0% | **36.1%** |
| **TOTAL** | | **19.4%** | **25%** | | **19.4%** | **19.4%** | **16.7%** | **100%** |

The relationships between two categorical variables can be easily represented using a contingency table. The categorical variable as relates to my stress level and whether I played a daily were compared using a contingency table as seen in Table 4. From the table, it can be seen that there is a relationship between whether I played a game and my stress level daily. For the days I ranked my stress level 3, 4, and 5, there was a significant difference between the frequencies of paying a game or not daily. For example, if you look at the stress level 3, in the contingency table (my most significant level of stress), 19.4% of those days, I was playing games, and 0% I was not.

The relationship appears to be the same at my highest level of stress which is 5, I played games 16.7% of those days while I did not play a game at 0%. On the other end of days that I feel less stressed at all (i.e levels 1 and 2) at stress level 1, 16.7% shows days I did not pay any game at all and 2.8% for days I play games. This shows a significant difference of about 13.9% at level one and 2.8 at level two.

From these analyses, I can infer that there is a relationship between the level of stress and whether I played a game or not because at my highest rank of stress, I played more games, and at my lowest scale of stress, I played fewer games with an extremely significant difference.

Chart, histogram

Description automatically generated

**Quantitative Variables:**

Time Spent in Zoom:

Chart, box and whisker chart

Description automatically generated This quantitative variable from my dataset focuses on the total number of hours I spent in zoom class daily. In **Figure 4**, shows a histogram that illustrates the variable. It can be seen in the graph that my time spent in class showed an asymmetric bimodal distribution with a slight right skew and no outliers. The mode (highest peak) was between 0-1 hour and a (lower peak) 3-4 This is because of weekends where I spent only 0-1 hour on zoom for group assignments and did not attend classes and Mondays where I have an average of 1.5 hours in zoom class for the lower peak.

Figure 4: Histogram Time Spent Zoom

The boxplot in **Figure 5** also shows the skew and my data with the median not being centered between upper and lower quartiles. Also, the box plot contains a fence above and below the lower and upper quartile of my data. Because the histogram is asymmetric, the interquartile range (IQR) is used to measure to determine the center and dispersion. **Table 5** illustrates the summary statistics for this variable. It can be seen from the table that the median was 4, which indicates that the middle value in my data was 4 hours in zoom classes while the IQR for this data was 4.

Figure 5: Boxplot of Hours Spent in Zoom

|  |  |
| --- | --- |
| **Table 5: Summary Statistics for Hours Spent on Zoom Each Day** | |
| **Summary Statistic** | **Hours** |
| Minimum | 0.00 |
| Median | 4.00 |
| Mean | 3.417 |
| IQR | 4 |
| Maximum | 7.00 |
| Standard Deviation | 2.297 |

To determine dispersion, the IQR was used and also to determine the outliers. For this analysis, we defined outliers as values that are 1.5 times the IQR. Using this method and looking at the boxplot in Figure 5, it can be concluded that there are no outliers in my data. Having no outliers could be attributed to my pretty consistent amount of time on Zoom each day.

**Table 9** shows the correlation between the time spent in zoon and other quantitative variables from my dataset. It can be seen from the table that the strongest relationship is between the hours spent in zoom and the hours spent outside the house. These two variables had a coefficient correlation of -0.46386211, which is the strongest correlation however, it is weak and has a negative relationship which I can attribute to the fact that I only join zoom meetings when I am at home.

**Time Spent Studying:**

Chart, histogram

Description automatically generatedThis quantitative variable from my dataset focuses on the total number of hours I spent studying daily. In **figure 6,** shows a histogram that illustrates the variable. It can be seen in the graph that my time spent in class showed a slight asymmetric unimodal distribution with a right skew and no outliers. The mode (highest peak) was between 2 and 4 hours. This is because of weekends where I spend an average of 2 hours studying daily

Figure 6: Histogram Time Spent on Study

Chart, box and whisker chart

Description automatically generatedThe boxplot in **Figure 6** also shows the skew and my data with the median not centered between upper and lower quartiles. Also, the box plot contains a fence above and below the lower and upper quartile of my data.

Because the histogram is asymmetric, the interquartile range (IQR) is used to measure to determine the center and dispersion. **Table 6** illustrates the summary statistics for this variable. It can be seen from the table that the median was 4, which indicates that the middle value in my data was 4 hours in study time, while the IQR for this data was 4.

Figure 7: Boxplot of Hours Spent on Study

To ascertain if there were any outliers in the data the z-scores were calculated to see if any data received a z-score of +/- 3, for such data to be classified as an outlier. It was determined that were no outliers in this data set. Using this method and looking at the boxplot in Figure 6, it can be concluded that there are no outliers in my data. This could be attributed to the lack of variations in my dataset, my study time is a prerogative for passing through the challenging analytics program at Cambrian College.

|  |  |
| --- | --- |
| **Table 6: Summary Statistics for Hours Spent Studying Each Day** | |
| **Summary Statistic** | **Hours** |
| Minimum | 1.00 |
| Median | 4.00 |
| Mean | 3.972 |
| IQR | 3 |
| Maximum | 9.00 |
| Standard Deviation | 2.006 |

Table 9 shows the correlation between the time spent studying and other quantitative variables from my dataset. It can be seen from the table that the strongest relationship is between the hours spent studying and the hours spent outside the house. These two variables had a coefficient correlation of 0.34930492, which is the strongest correlation. However, it is a weak and a positive relationship which I can attribute to the fact that I can only study when I am at home.

Chart, histogram

Description automatically generated**Time Spent Sleeping:**

This quantitative variable from my dataset focuses on the total number of hours I spent sleeping daily. **Figure 8** shows a histogram that illustrates the variable. It can be seen in the graph that time spent in class showed a symmetric distribution with no outliers. The mode (highest peak) was between 5 and 7 hours. This result may imply I have a balanced sleeping habit.

Chart, box and whisker chart

Description automatically generatedThe boxplot in **Figure 9** also shows that the median is slightly centered between upper and lower quartiles at 6.5 hours. Also, the box plot contains a fence above and below the lower and upper quartile of my data.

Figure 8: Histogram Time Spent Sleeping

Because the histogram is symmetric, the standard deviation and mean will be used to determine dispersion and centrality. It can be seen in **Table 7,** the average amount of sleeping per day was 6.472 hours. The standard deviation is 1.502

|  |  |
| --- | --- |
| **Table 7: Summary Statistics for Hours Spent Sleeping Each Day** | |
| **Summary Statistic** | **Hours** |
| Minimum | 3.00 |
| Median | 6.5 |
| Mean | 6.472 |
| IQR | 3 |
| Maximum | 9.00 |
| Standard Deviation | 1.502 |

To ascertain if there were any outliers in the data the z-scores were calculated to see if any data received a z-score of +/- 3, for such data to be classified as an outlier. It was determined that were no outliers in this data set. Using this method and looking at the boxplot in **Figure 9,** it can be concluded that there are no outliers in my data. As I expected this was not surprising because, over the years, I have maintained a balanced sleeping habit

Figure 9: Boxplot of Hours Spent Sleeping

**Table 9** shows the correlation between the time spent sleeping and other quantitative variables from my dataset. It can be seen from the table, that the strongest relationship is between the hours spent sleeping and the hours spent answering calls. These two variables had a coefficient correlation of 0.3375842, which is the strongest correlation. However, it is a weak and a positive relationship which I cannot infer any relationship between the variables.

**Time Spent Outside my Home:**

Chart, box and whisker chart

Description automatically generatedChart, histogram

Description automatically generatedThis quantitative variable from my dataset focuses on the total number of times (count) I spent outside my home daily. **Figure 10** shows a histogram that illustrates the variable. It can be seen in the graph that my time spent outside my home showed an asymmetric unimodal distribution with a right skew and no outliers. The mode (highest peak) was between 0 and 1 hour. This is because I don’t have a job and I spend most of my time indoors studying or attending zoom classes.

Figure 10: Histogram Time Spent Outside the House.

The boxplot in **Figure 11** also shows the skew and my data with the median not being centered between upper and lower quartiles. Also, the box plot contains a fence above and below the lower and upper quartile of my data. Because the histogram is asymmetric, the interquartile range (IQR) is used to measure to determine the center and dispersion. **Table 8** illustrates the summary statistics for this variable. It can be seen from the table that the median was 2, which indicates that the middle value in my data was 2 hours, while the IQR for this data was 2 also.

Figure 11: Boxplot of Hours Spent Outside the House

|  |  |
| --- | --- |
| **Table 8: Summary Statistics for Number of Times I Left the House Each Day** | |
| **Summary Statistic** | **Times(Count)** |
| Minimum | 0.00 |
| Median | 2.00 |
| Mean | 2.25 |
| IQR | 2 |
| Maximum | 6.00 |
| Standard Deviation | 1.762 |

To determine dispersion, the IQR was also used to determine the outliers. For this analysis, we defined outliers as values that are 1.5 times the IQR. Using this method and looking at the boxplot in Figure 8, it can be concluded that there are no outliers in my data. This is because the hours I speed outside my house are pretty uniform, I mostly go out on the weekends maybe for Walmart to restock or to church on Sundays. These trends were consistent throughout this data collection.

**Table 9** shows the correlation between the time spent outside my home and other quantitative variables from my dataset. It can be seen from the table that the strongest relationship is between the hours spent in zoom and the hours spent outside the house. These two variables had a coefficient correlation of -0.4638621, which is the strongest correlation however, it is weak and has a negative relationship which I can attribute to the fact that I only join zoom meetings when I am at home.

**Chart, histogram

Description automatically generated**

**Number of Times I Answered Calls:**

This quantitative variable from my dataset focuses on the total number of hours I spent answering calls daily. In **Figure 12,** shows a histogram that illustrates the variable. It can be seen in the graph that my time spent taking calls showed a right-skewed unimodal distribution with a right skew and no outliers on the histogram. The mode (highest peak) was between 1 and 2 times daily. This is because I am new to Canada and have not made friends to talk to over the phone.

Chart, box and whisker chart

Description automatically generatedThe boxplot in **Figure 13** also shows the skew and my data with the median centered between upper and lower quartiles. Also, the box plot contains a fence above and below the lower and upper quartile of my data.

Figure 12: Histogram on Number of Calls

Because the histogram is asymmetric, the interquartile range (IQR) is used to measure to determine the center and dispersion. **Table 8** illustrates the summary statistics for this variable. It can be seen from the table that the median was 3, which indicates that the middle value in my data was 3 times, while the IQR for this data was 2

|  |  |
| --- | --- |
| **Table 9: Summary Statistics for Hours Spent Calls Each Day** | |
| **Summary Statistic** | **Hours** |
| Minimum | 1.000 |
| Median | 3.000 |
| Mean | 3.139 |
| IQR | 2 |
| Maximum | 7.000 |
| Standard Deviation | 1.725 |

To determine dispersion, the IQR was used. For this analysis, we defined outliers as values that are 1.5 times the IQR. Using this method and looking at the boxplot in **Figure 13**, it can be concluded that there are no outliers in my data. Having no outliers for my count of calls may imply that my data was consistent with no extreme values. This is as a result of video calls with family members and friends.

Figure 13: Boxplot of Number of Calls

**Table

Description automatically generatedTable 9** shows the correlation between the number of times I answered calls and other quantitative variables from my dataset. It can be seen from the table that the strongest relationship is between the number of times I answered calls and the hours spent sleeping. These two variables had a coefficient correlation of 0.33758420, which is the strongest correlation. However, it is a weak and a positive relationship.

**TWO quantitative variables with the highest correlation:**

It can be seen in Table 10 that the two quantitative variables in my data set with the highest correlation were the hours I spent outside my home and hours I spent in zoom classes or meetings. The correlation coefficient was -0.46386211, which represents a weak, negative relationship between the variables. This will suggest that I spend more time in zoom classes and I can only be out of my home when I am attending a zoom class or meeting for group assignments.

Table 10: Correlation between all Quantitative Variables

Table

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generatedThere are multiple conditions of regression that need to be met. The first condition states that both variables need to be quantitative, which is true in this case as the values collected for both variables were unique integers. The second condition of regression to be met is the linearity of the data, as correlation can only measure the strength of linear relationships. The plot of residuals vs fitted values in **Figure 14** shows that the condition of linearity is not being met as the red line is not horizontal. The next condition that needs to be met is the outlier condition. **Figure 16** is a unique plot that can help determine if this condition is met. Since the points slightly follow the straight line on this plot, the condition is met. The final condition is the equal spread condition. **Figure 15** illustrates the scale-location plot, and it can be used to determine if this condition is met. The line is close to horizontal, so this condition is met. Finally, in **Figure 17**, A residuals vs. leverage plot compares the standardized residuals with the influence every point has (called leverage). It can be used to check the outlier condition. Outliers would appear beyond the dashed lines. For this analysis, the condition is met.

Figure 14: Residuals vs Fitted Plot

Figure 15: Scale-Location Plot

Chart, scatter chart

Description automatically generatedChart, line chart

Description automatically generated

Figure 17: Residuals Vs Leverage

Figure 16: Normal Q-Q Plot

**The equation of the regression line is as follows:**

**Number of times I left the house (count) = -0.3558863 (Time in Zoom Meetings(Hours)) +3.4659450**

Slope = -**0.3558863**

**Interpretation:** As the number of hour spent in class increased by 1, the number of times I leave my house decreased by **-0.3558863**

Chart, scatter chart

Description automatically generatedY-intercept = **3.4659450**

**Interpretation:** This value tells us that when time spent in zoom is 0 (hour), the number of times I leave my house is = **3.4659450**

Figure 18: The Relationship between Hours Spent on Zoom and Outside the House

I chose the count of times I left my house independent variable and my zoom hours as the dependent variable for this analysis. I did this because I can only be able to leave my house when there are no zoom meetings to attend, so going out to Walmart or for movies are dependent on my zoom meetings.

The regression line allows us to calculate hypothetical values for the dependent variable by inputting a value for the independent value. I could determine what the predicted value could be for the number of times I left the house if I were to have 5.5 hours of zoom meeting. By looking at the regression line above, it could be estimated that the number of times I left the house could be approximately 1.59 times. Using the equation of the line could allow for a more precise value to be given.

The number of times I left the house (count) = -0.3558863 (5.5) +3.4659450

The number of times I left the house (count) = 1.50857

# Conclusion

After analyzing the data collected from September 10th to October 15th, I can better understand how I spend my days.

Analysis of the categorical variables resulted in the following conclusions. that I spent most of my day in the stress level with equal percentages in levels one, three, and four at 19.4% each. Though level five was the lowest with 16.7%, while level two was significant with 25%. Achieving a stress level one, two, and four at 19.4% may likely be associated with Wednesday zoom classes which have more than 7 hours in zoom time coupled with time spent on assignments. This was not surprising owing to the stressful nature of the analytics program. I used 41.7% of the days to listen to the news while 58.3% shows days which I did not listen to the news. I played games at 36.1% response representing the amount of time spent on playing games while 63.9% represents the amount of time I did not play a game in a day. It could be seen that there was a strong relationship between the time I spent playing games and my stress level which may imply that playing games increase my stress level after a stressful week of academic activities. It was surprising to understand that playing games increased my level of stress per day.

Analysis of the quantitative variables showed the following results. I spent an average of 3.417 hours in zoom and 3.972 hours studying per day. I slept an average of 6.472 hours per night and I went outside my house 2.25 times daily. Going outside of my house 2.25 times daily was surprising because I initially thought that I was not going out at all. Though on a closer look at my dataset, I noticed added when days that I took a minute walk outside my house to the data. I only answered calls 3.1 hours daily. Evaluation of the correlation coefficients revealed that there are no strong correlations between any of my quantitative variables either positive or negative. Though the quantitative variable House and Zoom had the highest negative relationship at -0.638621. this was expected because zoom meeting is a prerogative that is independent.

Overall, I am indifferent with my time management skill though I still need to improve I some areas to reduce my stress level to a lower scale. I have made some personal recommendations to improve how I spend my day for the remaining semester:

1. I will reduce the number of hours I spend talking to family members by 50%
2. I will only pay games when I am totally relaxed
3. I will improve my social and networking skills.

**MARKING GUIDE** The case study is worth 20% of your final grade and will be evaluated based on the guide below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Level 1 | Level 2 | Level 3 | Level 4 | Comments |
| REPORT | Introduction to the report is unclear, explaining few variables investigated, questions, and methods | Introduction to the report is somewhat clear, explaining some variables investigated, questions, and methods | Introduction to the report is mostly clear, explaining most variables investigated, questions, and methods | Introduction to the report is written clearly, summarizing variables investigated, explaining all questions in the case study and methods used |  |
| Data analysis section uses no relevant examples relating to personal activities | Data analysis section uses few relevant examples relating to personal activities (less than 3) | Data analysis section uses some relevant examples relating to personal activities (between 3 – 5) | Data analysis section uses multiple relevant examples relating to personal activities (more than 5) |  |
| Conclusion to the report clearly restates none of the questions investigated and few of the main findings | Conclusion to the report is missing two of the two questions investigated or the main findings | Conclusion to the report is missing one of the two questions investigated or the main findings | Conclusion to the report clearly restates two of the questions investigated and the main findings |  |
| Errors in word selection, usage, sentence structure, spelling, punctuation and capitalization detract from meaning | Many errors in word selection, usage, sentence structure, spelling, punctuation and capitalization | Some errors in word selection, usage, sentence structure, spelling, punctuation and capitalization | Few errors in word selection, usage, sentence structure, spelling, punctuation and capitalization |  |
| Notation and formatting of tables, graphs, and text is inconsistent, and the report does not look professional | Notation and formatting of few tables, graphs, and text is consistent and the report looks presentable | Notation and formatting of some tables, graphs, and text is consistent and the report looks presentable | Notation and formatting of all tables, graphs, and text is consistent and the report looks professional |  |
|  | **0 marks** | **1 mark** | **2 marks** | **3 marks** | **Comments** |
| Categorical | No frequency distributions are included | A correct frequency distribution is included and interpreted for 1/3 categorical variables | A correct frequency distribution is included and interpreted for 2/3 categorical variables | A correct frequency distribution is included and interpreted for all three categorical variables |  |
| No visual displays are included | A correct visual display and interpretation are included for 1/3 categorical variables | A correct visual display and interpretation are included for 2/3 categorical variables | A correct visual display and interpretation are included for all categorical variables |  |
| No two-way contingency table | A contingency table and interpretation are included with at least one error | A contingency table and interpretation are included with one error | A correct contingency table and interpretation are included |  |
| Quantitative | No visual display is included | A properly labelled and correctly constructed histogram is included and interpreted for one or two quantitative variables | A properly labelled and correctly constructed histogram is included and interpreted for three or four quantitative variables | A properly labelled and correctly constructed histogram is included and interpreted for all five quantitative variables |  |
| No measures of centre are calculated | Measure of centre are calculated correctly for one or two quantitative variables | Measure of centre are calculated correctly for three or four quantitative variables | Measure of centre are calculated correctly for all five quantitative variables |  |
| No measures of spread are calculated | Measures of spread are calculated correctly for one or two quantitative variables | Measures of spread are calculated correctly for three or four quantitative variables | Measures of spread are calculated correctly for all five quantitative variables |  |
| No check for outliers | Outliers are calculated and explained for one or two quantitative variables | Outliers are calculated and explained for three or four quantitative variables | Outliers are calculated and explained for all five quantitative variables |  |
| Correlation & Regression | Correlation coefficients are not calculated | Correlation coefficients are calculated with at least one error | Correlation coefficients are calculated correctly for all pairs of variables | N/A |  |
| Direction and strength of relationship are not given | One of the direction and strength of the relationship are stated correctly | Both the direction and strength of the relationship are stated correctly | N/A |  |
| Conditions are not checked | Conditions are checked with more than one error | All conditions are checked with one error | All conditions are checked correctly |  |
| The equation of the regression line is not determined | The equation of the regression line is calculated with at least one error | The equation of the regression line is calculated with one error and/or is missing from the report | The equation of the regression line is calculated correctly and stated in the report |  |
| The slope and y-intercept are not interpreted | The slope and y-intercept are both interpreted with errors | One of the slope and y-intercept are interpreted correctly | The slope and y-intercept are interpreted correctly |  |
| No visual display is included | A visual display and interpretation are included with at least one error | A properly labelled and correctly constructed visual display is included and interpreted | A properly labelled and correctly constructed visual display is included and interpreted |  |
| No prediction is given | A prediction is included with at least one error | A correct prediction is given | N/A |  |
|  | **0 marks** | **1 mark** |  |  | **Comments** |
| Submission | One or more of the following components are missing  -Personalized data (.csv)  -Report in Word  -R script with code | All the following components are submitted:  -Personalized data (.csv)  -Report in Word  -R script with code | N/A |  |  |
| Comments |  | | | | |
|  |  |  |  | TOTAL | **/60** |