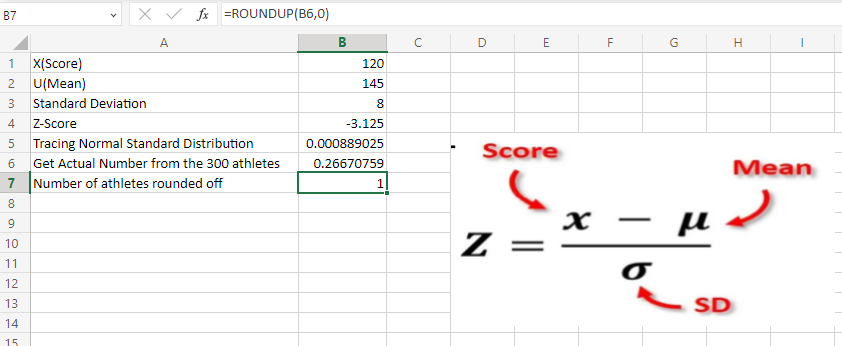
Assignment 1

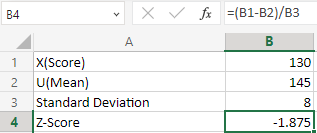
1. Question 1
   1. Giving examples distinguish between Correlation and Regression (4mks)
      1. Correlation shows the relationship between two variables. An example of this is say in the transport sector, the relationship between alcohol consumption before an individual drives a vehicle and the number of accidents caused because of this alcohol consumption. In the case of Regression, it allows us to see how one variable affects the other. An example of this is say two variables, x and y are related, with regression, a change in variable x leads to a change in variable y which on average is usually by a certain amount. Regression also gives the equation for a line or a curve.
   2. Thought Process
      1. The Overview
         1. 
         2. The formula used to get the Z-Score
         3. Graphical user interface, application

            Description automatically generated
         4. The Formula used to trace the Normal Standard Distribution
         5. Graphical user interface, text, application, email

            Description automatically generated
         6. The formula I used to get the actual number of athletes out of the total 300
         7. Graphical user interface, text, application, email

            Description automatically generated
         8. I rounded the number off to the nearest whole number since we are dealing with people as the data
         9. Graphical user interface, text, application, table

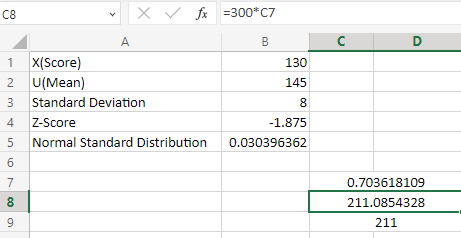
            Description automatically generated
         10. Finally, the number of athletes who completed the race in under 120 minutes is 1
      2. Thought Process
         1. The Overview
            1. Graphical user interface, text, application

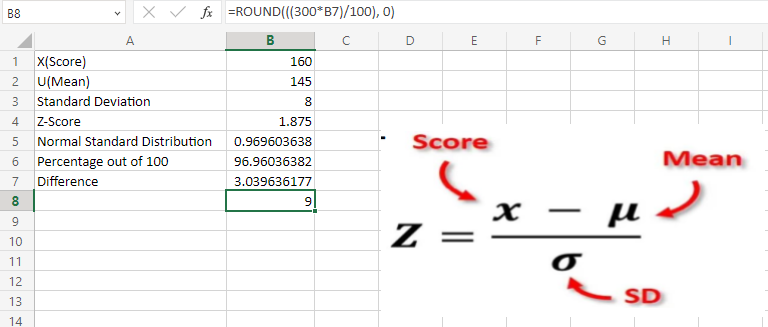
               Description automatically generated
            2. The Z-Score formula for 130 Score
            3. 
            4. The Normal Standard Distribution for 130
            5. Graphical user interface, application, table

               Description automatically generated
            6. The Z-Score formula for 150 Score
            7. Graphical user interface, application, table

               Description automatically generated
            8. The Normal Standard Distribution for 150 Score
            9. Graphical user interface, application, table

               Description automatically generated
            10. The difference in the distributions
            11. Graphical user interface, application, table, Excel

                Description automatically generated
            12. The actual number of athletes out of the 300
            13. 
            14. The number of athletes rounded off to the nearest whole number
            15. Table

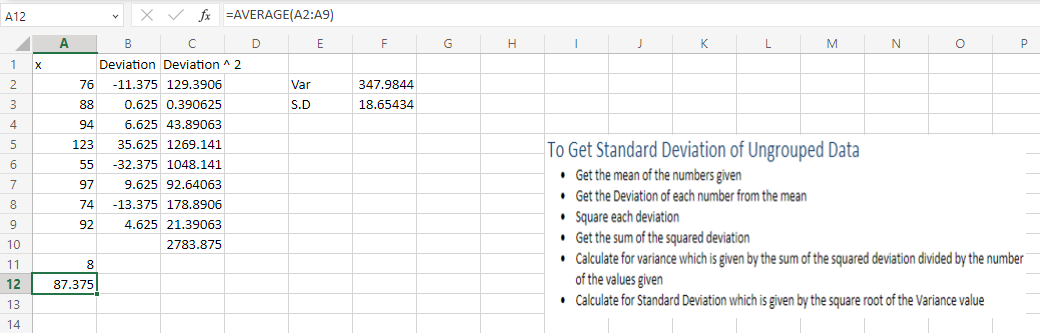
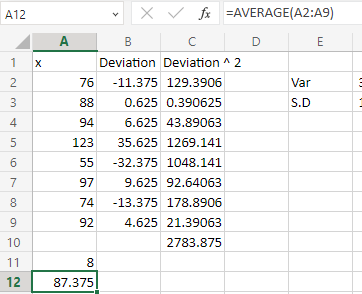
                Description automatically generated
            16. The number of athletes who ran between 130 minutes and 150 minutes is 211.
      3. Thought Process
         1. The Overview
            1. 
            2. Z-Score formula I used
            3. Graphical user interface, application, table

               Description automatically generated
            4. The formula I used to trace the Normal Standard Distribution
            5. Graphical user interface, application, table

               Description automatically generated
            6. How I got the percentage out of 100
            7. Graphical user interface, application, table

               Description automatically generated
            8. How I got the difference since we are calculating the number of athletes who spent more than 160 minutes
            9. Graphical user interface, application, table, Excel

               Description automatically generated
            10. The formula I used to get the whole actual number of athletes
            11. Graphical user interface, application, table, Excel

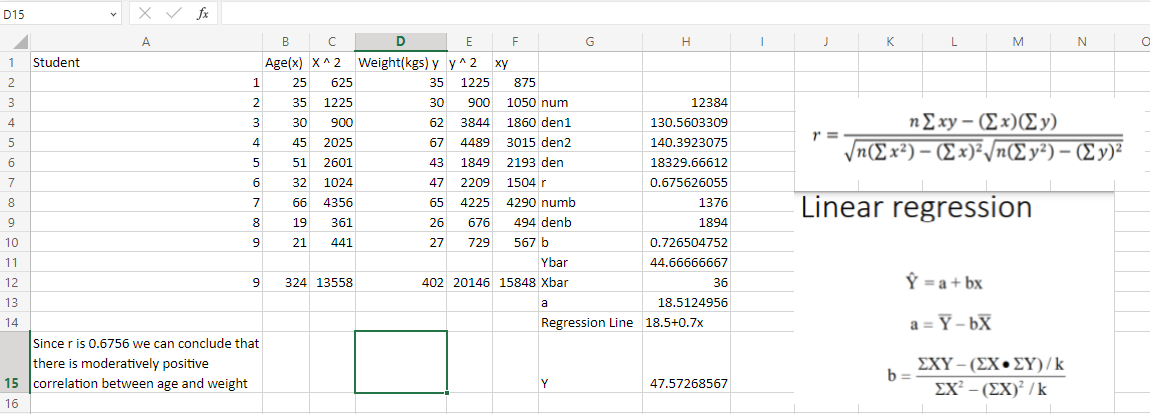
                Description automatically generated
            12. Hence making my conclusion that the number of athletes who spent more than 160 minutes in the race is 9.
   3. Thought Process
      1. The Overview
         1. 
      2. The Formula I used to get the mean
         1. 
      3. The Formula I used to get the deviation of each number from the mean
         1. Graphical user interface, application, table, Excel

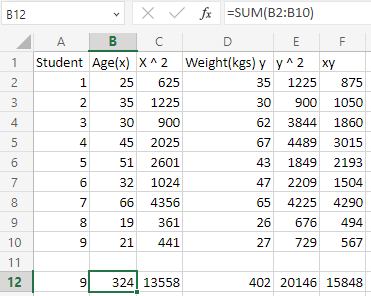
            Description automatically generated
      4. The Formula I used to Square each deviation
         1. Graphical user interface, application, table, Excel

            Description automatically generated
      5. The Formula I used to get the sum of the squared deviations
         1. Graphical user interface, application, table, Excel

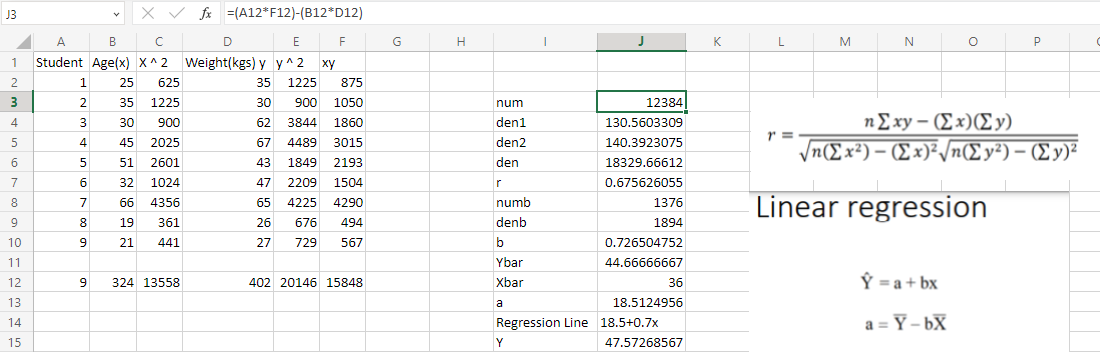
            Description automatically generated
      6. The Formula I used to calculate for variance
         1. Table, Excel

            Description automatically generated
      7. The Formula I used to calculate for Standard Deviation
         1. Graphical user interface, application, table, Excel

            Description automatically generated
      8. Hence having a standard deviation of 18.65
2. Coefficient
   1. Thought Process
      1. The Overview
         1. 
      2. The Formula I used to get the number of values
         1. Table, Excel

            Description automatically generated
      3. The Formula I used to get the summation of the Age, Age squared, weight, weight squared and the product between age and weight
         1. 
      4. The Function I used to get Age and Weight Squared
         1. Table, Excel

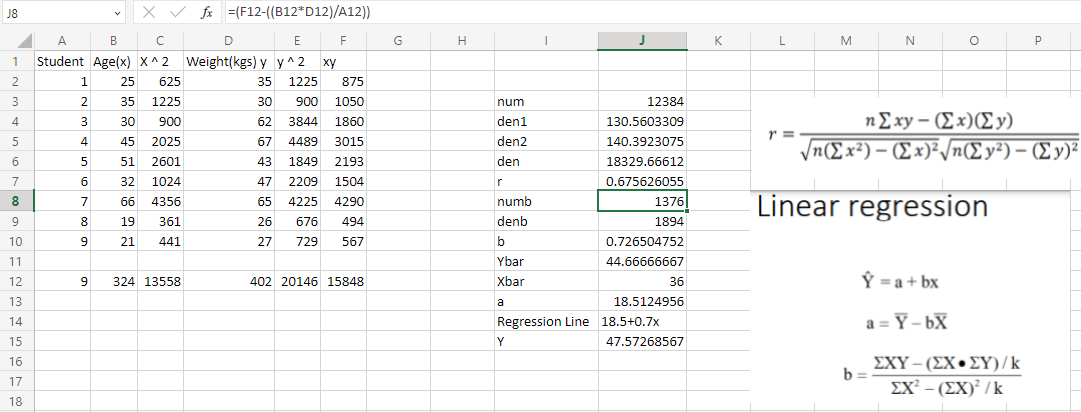
            Description automatically generated
      5. The Formula I used to get the product between Age and Weight
         1. Table, Excel

            Description automatically generated
      6. The Formula I used to get the numerator of r
         1. 
      7. The Formula I used to get the first denominator of r
         1. Graphical user interface, table, Excel

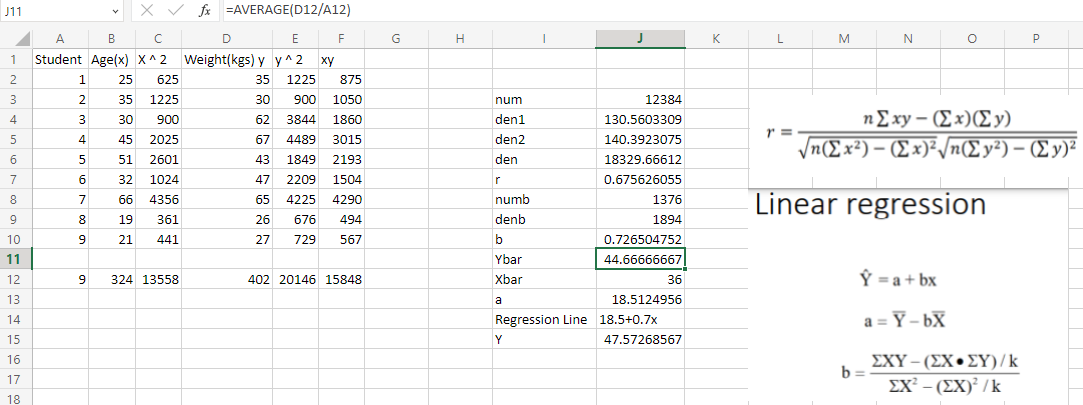
            Description automatically generated
      8. The Formula I used to get the second denominator of r
         1. A screenshot of a computer

            Description automatically generated with medium confidence
      9. The Formula I used to get the whole Denominator of r
         1. Graphical user interface, application, table

            Description automatically generated
      10. The Formula I used to calculate r
          1. Application, table, Excel

             Description automatically generated
      11. The Formula I used to get the numerator for b
          1. 
      12. The formula I used to get the denominator for b
          1. Graphical user interface, application, table, Excel

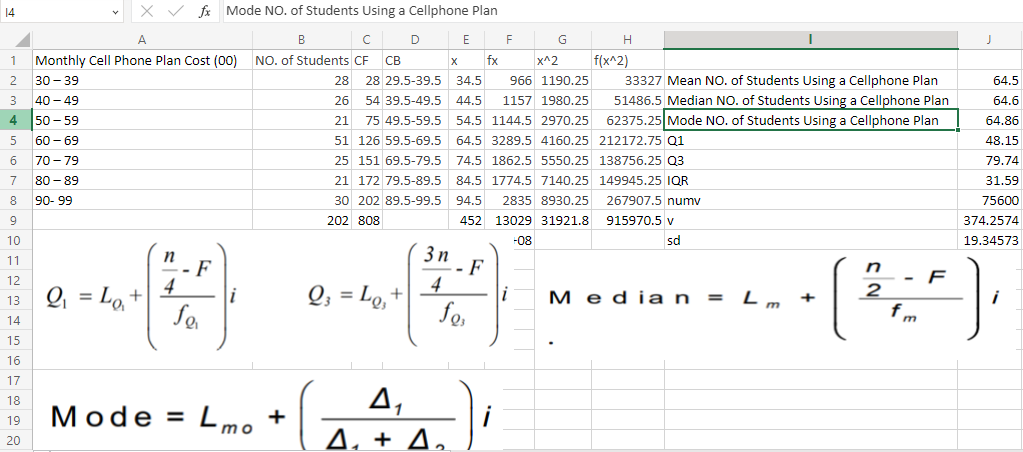
             Description automatically generated
      13. The formula I used to calculate for b
          1. Graphical user interface, application, table, Excel

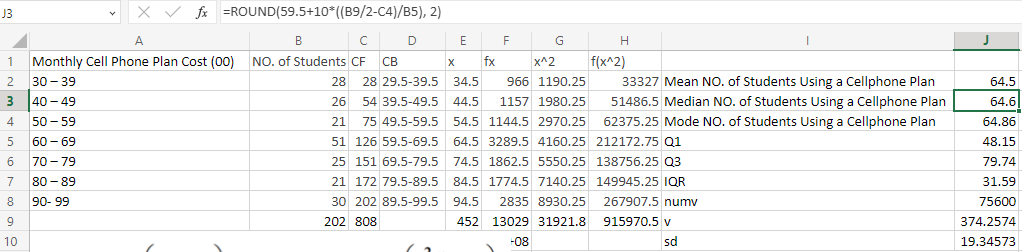
             Description automatically generated
      14. The formula I used to calculate for Ybar to be used in finding the value of a
          1. 
      15. The formula I used to calculate for Xbar to be used in finding the value of a
          1. Graphical user interface, application, table, Excel

             Description automatically generated
      16. The formula I used to calculate for a
          1. Graphical user interface, application, table, Excel

             Description automatically generated
      17. My regression line is given as show below
          1. Table

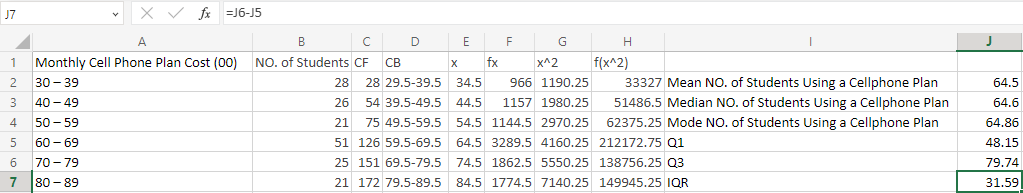
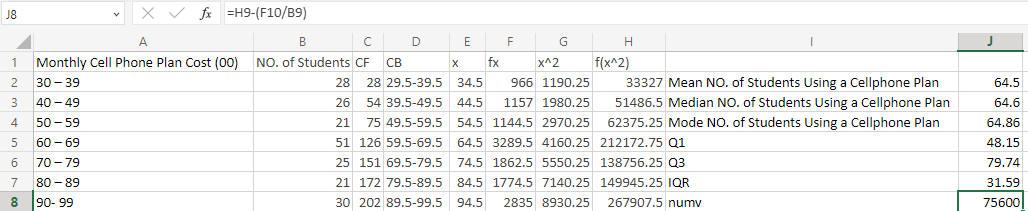
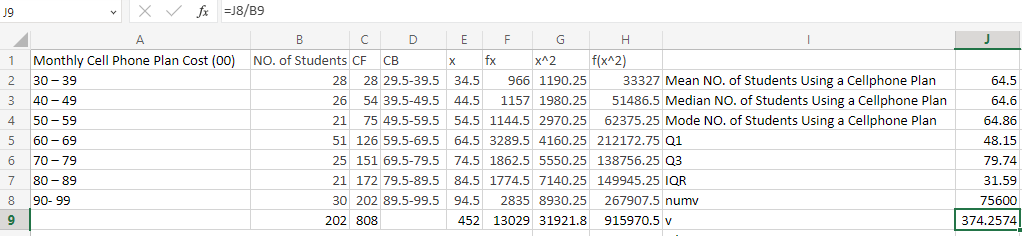
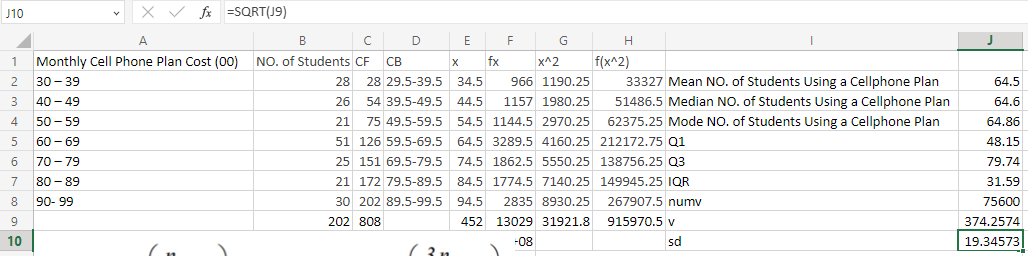
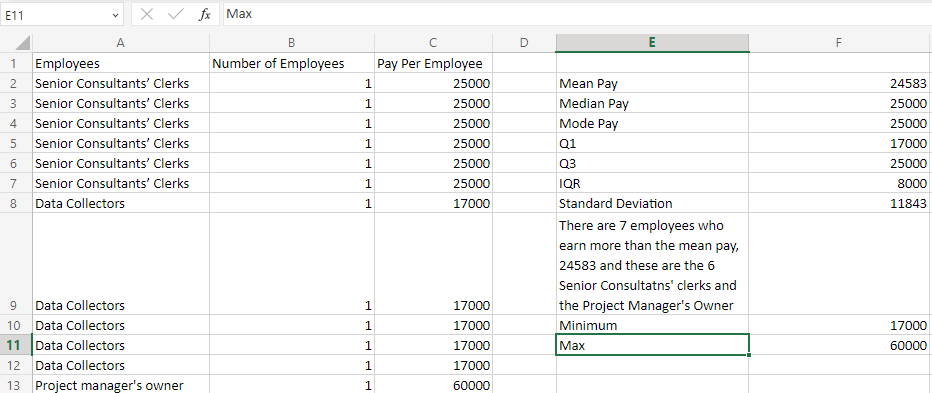
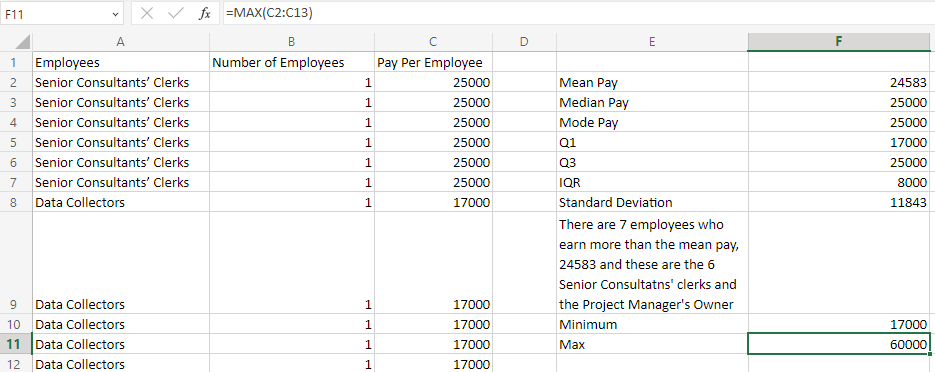
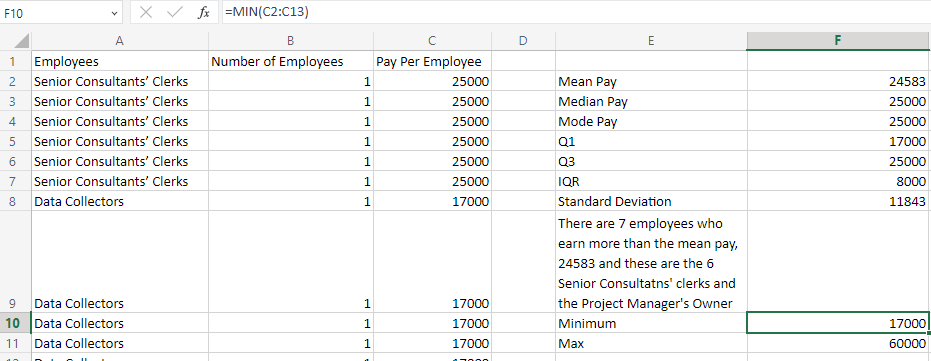
             Description automatically generated
      18. The value of Y when x is 40 is given by the formula below
          1. Graphical user interface, application, table, Excel

             Description automatically generated
   2. Differentiate the different types of data in statistics.
      1. When we receive data, we can classify it into two categories:
         1. Categorical data. This is also called qualitative data and includes data that is not numerical. Qualitative data can also be divided into:
            1. Nominal Data. This data labels variables without any type of quantitative value. An example of this can be in the case of Race, we can Black, White & Brown.
            2. Ordinal Data. This data represents the order of a number. It places data into some kind of order based off their position on a scale. An example is in a class, grading is done from A-F with A representing the highest score and F representing the lowest score in that scale and order.
         2. Numerical data. This is also called quantitative data and represents a numerical value, that is, how many, how often and how much. Quantitative data can also be divided into:
            1. Discrete data. This contains only a finite number of values, i.e., whole numbers like the number of students in a class.
            2. Continuous data. This contains only infinite number of values, i.e., numbers with decimal places such as the temperature of a classroom.
3. Mean, Mode, Median, IQR
   1. Thought Process
      1. Overview
         1. 
      2. Mean
         1. Graphical user interface, text

            Description automatically generated
      3. Median
         1. 
      4. Mode
         1. A picture containing text

            Description automatically generated
      5. Q1
         1. A picture containing table

            Description automatically generated
      6. Q3
         1. Text

            Description automatically generated
      7. IQR
         1. 
      8. Variance numerator
         1. 
      9. Variance
         1. 
      10. Standard Deviation
          1. 
   2. Overview
      1. 
      2. Max
         1. 
      3. Min
         1. 
      4. Q1
         1. Graphical user interface

            Description automatically generated with medium confidence
      5. Median
         1. Graphical user interface

            Description automatically generated with medium confidence
      6. Q3
         1. A picture containing graphical user interface

            Description automatically generated