Statistical Analysis of Time Allotted for Morning Routines

Chances are, when someone is running late, it is because they didn't reserve enough time for themselves to get ready for their obligation. Being a student that is constantly slipping into morning classes right as the bell rings, I undoubtedly feel the effects of requiring too much time to get ready in the morning. However, I wanted to see exactly where I stood in respect to my peers. And so I asked a group of thirty students, fifteen girls and fifteen boys: How long do you take to get ready in the morning (excluding breakfast)? I received the following responses (in minutes):

5, 10, 10, 10, 10, 10, 15, 15, 20, 20, 20, 20, 20, 20, 25, 25, 25, 30, 30, 30, 30, 35, 45, 45, 45, 45, 45, 60, 90, 110, 120

The minimum value for the resulting 30-value data set was 5 minutes, and the maximum value was 120 minutes, resulting in a range of 115 minutes \((120 - 5 = 115\)). The value of Q₁, or the number for which 25% of the data falls below, is 20 minutes. This value was determined by dividing the sample size (30) by 4. To determine where the first 25% of the data fell, \(\frac{30}{4} = 7.5\), so the value of Q₁ was the average of the 7th and 8th value in the dataset. The value of Q₃, or the number for which 75% of the data falls below, was determined in the same way, except this value would be the average of the 22nd and 23rd values. The value of Q₃ is 45 minutes. Using Q₁ and Q₃, I found the value of the interquartile range, or the IQR, which is simply the difference between Q₁ and Q₃. The IQR of the dataset is 25. The mean of the dataset was determined by finding the sum of all values (1005 minutes),
and dividing this total by the number of terms in the dataset (30). \( \frac{1005}{30} = 33.5 \)

minutes. I found the median of the dataset by finding the two values in the middle of the data. The two middle values were 25, which means that the median of the data is 25 minutes.

The variance of the data was determined by first, finding the difference between each of the 30 values from the mean (33.5). I then squared all of the 30 resulting values and found the sum of those squared figures. The sum was then divided by 30, to arrive at a variance of 767.29. Using this value, I found the standard deviation. The standard deviation is a measure of how the data is clustered around the mean. The standard deviation for this dataset is 27.7.

I then determined whether or not any of my data points could be considered outliers (Section D of graph paper) and found that the recorded times of 90, 110, and 120 minutes were all outliers in the data set.

The histogram, boxplot, and stemplot provide a visual representation using the previously calculated summary statistics. (Section A of graph paper).

In the next part of the data analysis I added 100 to all values in the data set. The minimum value for the modified data is 105 minutes, and the maximum value is 220 minutes making the range of the dataset 115. \( Q_1 \) is 120, and \( Q_3 \) is 145 making the IQR 25. The median of the modified data is 125 minutes, and the mean is 133.5. The standard deviation for the modified data is 27.7 and the variance is 767.29. The mean and median of the modified dataset are both 100 minutes larger than the original mean and median of the unmodified data. The standard deviation, however,
remained the same. The histogram, boxplot, and stemplot (Section B of graph paper) are visual representations of the data after 100 minutes was added to each value.

The next step in the data analysis was increasing all values in the data set by 50%. The minimum value for this modified data is 7.5 minutes, and the maximum value is 180 minutes making the range of the dataset 172.5. Q₁ is 30, and Q₃ is 67.5 making the IQR 37.5. The median of this modified data is 37.5 minutes, and the mean is 50.25. The standard deviation for the modified data is 41.66 and the variance is 1735.56. The median and mean of this dataset were both 50% larger than the median and mean of the first dataset. The standard deviation also increased by 50% from the original value. The (Section C) histogram, boxplot, and stemplot are visual representations of the data after 100 minutes was added to each value.

The next part of the data analysis was to assume the data fit a normal curve and analyze the characteristics of that data. Five minutes above the mean of my original data set would be 38.5 minutes. 43% of the data would fall below this figure. To be in the highest 10% of the data, one would have to take 69 minutes to get ready in the morning.

By examining the histograms, boxplots, and stemplots, you can see that the data is skewed right, meaning that most students estimate their time to get ready in the morning to be below an hour. This fact could be the result of peoples preconceptions as to what an acceptable number would have been, making them report a smaller amount of time. A data set with more data points would have shown a larger spread in the data, and been a more accurate representation of the amount of time people actually take to get ready. However, according to this 30-
value dataset, most people take between 20 and 45 minutes to get ready in the morning, and almost all of the people surveyed (except for 3 individuals) take an hour or less.