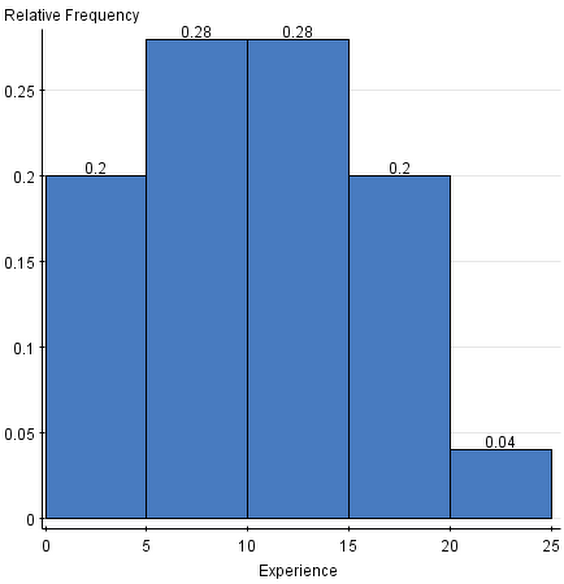
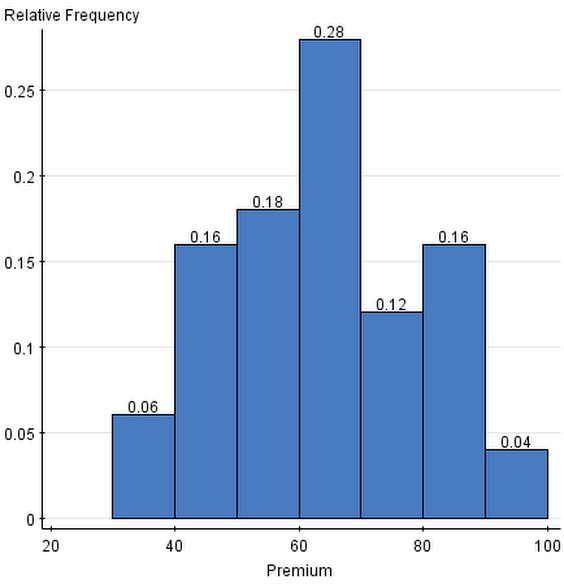
* + Experience – Quantitative, because it represents the total amount of years of experience when driving.
  + Gender – Categorical, because it represents the gender of each respondent.
  + Premium – Quantitative, because it represents the premium price that the respondent pays.

1. The categorical variable of gender represents either a 0 or a 1, with gender 0 having a frequency of 29, which represents 58% of respondents. Gender 1 has 21 respondents, representing 42% of all respondents. Though the distribution of the frequency is fairly similar, there is a larger distribution for gender 0.
2. The shape of both distributions are fairly symmetrical, and generally fall under the shape of a bell curve. There is a noticeable anomaly in this pattern in the Premium graph, between 80 and 90 where the relative frequency jumps up a bit. The center of the Premium seems to fall somewhere between 60 and 70, somewhere closer to the middle. Meanwhile, the center of the Experience graph falls between 10 and 15, probably closer to the 15.



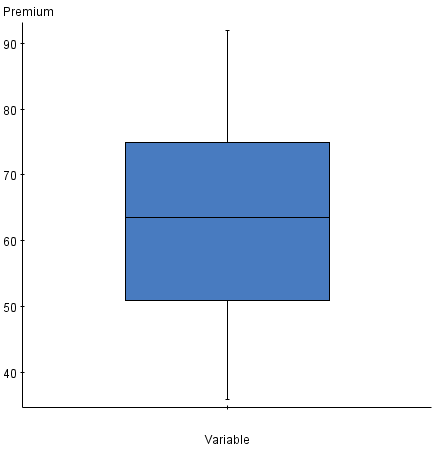
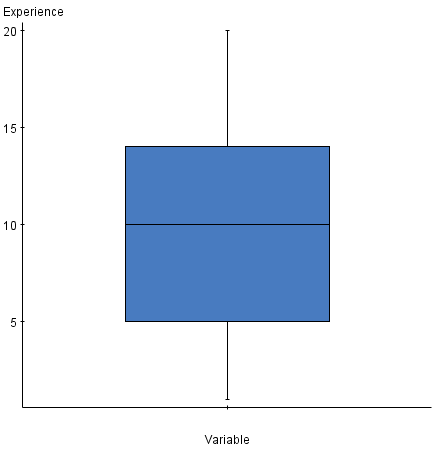


1. Data presented below in question 5.
2. The mean and median are very close to each other. Experience has a mean of 9.86 and a median of 10, which are very close together, while Premium has a mean of 62.98 and a median of 63.5. This means that the center of distribution should be in a relatively similar location for each graph. The data should be fairly symmetric as well, and should fall under the general shape of a bell curve. Though the data shouldn’t be exactly symmetrical, the graphs will be fairly symmetrical in their shape. The standard deviation illustrates that there will be a relatively decent spread of data, particularly for the Premium graph, which has a standard deviation of 15.1, though because there is a large range of data in the Premium data set, the spread of data won’t be incredibly large. The standard deviation of the Experience data set is only 5.4, which tells us that the spread of data will be relatively compact and concentrated.

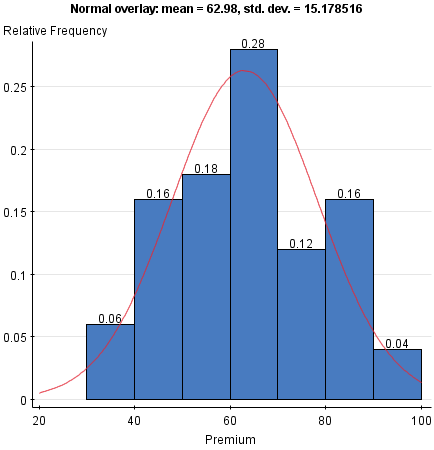
|  |  |  |
| --- | --- | --- |
| **Column** | **Mean** | **Std. Dev.** |
| Experience | 9.86 | 5.477263 |
| Premium | 62.98 | 15.178516 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column** | **Median** | **Min** | **Max** | **Q1** | **Q3** |
| Experience | 10 | 1 | 20 | 5 | 14 |
| Premium | 63.5 | 36 | 92 | 51 | 75 |

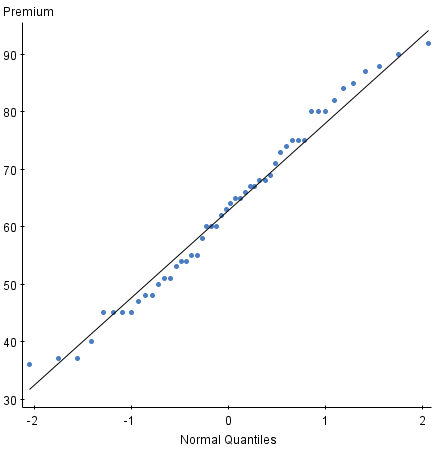
1. The spread between the two data sets are relatively similar. There is a smaller spread for the Premium box plot, but the spread is only barely smaller than the Experience plot. The spread of data for both of the data sets lie fairly equally above and below the median, though the spread tends to fall more heavily below the median.



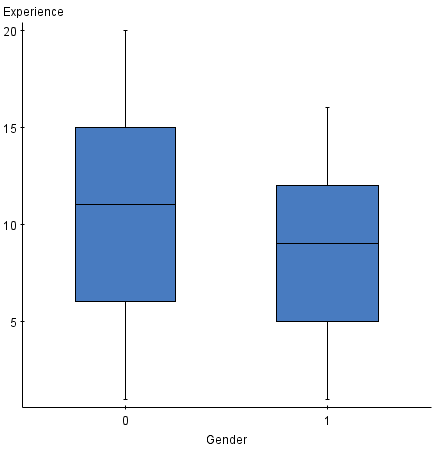
1. The distribution is relatively normal. Though the standard deviation is approximately 15, this deviation is relatively small in comparison to the range of Premiums; between 20 and 100. The graph is symmetrical for the most part, with symmetrical sides varying by a relative frequency 0.06 at most (this occurs between the 0.18 and 0.12 on opposite sides). It seems that the data above the curve will make up for some of the data that is missing within the curve, thus I believe that the variable Premium is a fairly normal graph.



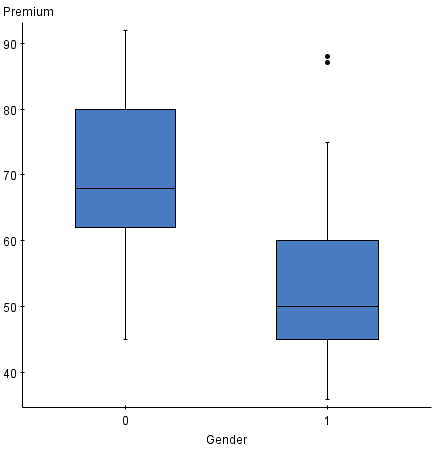
1. Premium is fairly normal. There is a large concentration of Premiums between 60 and 70, and though many data points do not directly touch the line, many of them are close to it. The data is not perfectly linear, but it is fairly linear, and thus even though the graph isn’t perfectly normal, the graph is still relatively normal; there do not appear to be and distinct outliers either.



1. Gender 0 has a greater spread of Experience than Gender 1 does. Though both genders have a similar minimum, Gender 0 has a greater maximum, median, Q3, and Q1. Both genders appear to have a larger portion of experience fall below the median, though Gender 0 appears to be more equal in that regard; an approximately equal amount of data falls both above and below the median, and between Q1 and Q3.



When comparing the Premium of both Genders, Gender 0 has a much higher premium in all aspects of the box plot when compared to Gender 1. The spread of data is more compact for Gender 1, and overall the Premium is much higher in all regards for Gender 0. Gender 1 does have two outliers however, and they are in the top range of Gender 0. This brings forth some questions. Why are the two outliers so large in comparison to the rest of the data for Gender 1? Was there an error in inputting the data? Does that data set belong in Gender 0? Or are these outliers simply a rare occurrence in the data set?



1. The five number summary (along with a mean) is an appropriate set of statistics to report on for each distribution. It allows for various aspects of the data to be analyzed. The mean and median of both genders can be compared to each other, which gives an idea of where the center of distribution is located for each gender. Comparisons can be made between the two genders to see whether the genders have a different center of distribution, and how different that center is. The five number summary helps in gaining an understanding in the spread of data between the two genders for each respective table. For example, the numbers in the five number summary for Experience are fairly similar between the two genders, while Premiums are quite different. We can see that Gender 0 has overall higher data in every regard, and the spread for Gender 0 is greater for both Experience and Premium. These statistics allows us to essentially understand the spread of data, the center of data, and the relative graphical shape of the data. Hence these statistics give us an accurate way to compare Gender 0 and Gender 1 by both Experience and Premium.

**Summary statistics for Experience:**   
Group by: Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gender** | **Mean** | **Median** | **Min** | **Max** | **Q1** | **Q3** |
| 0 | 10.896552 | 11 | 1 | 20 | 6 | 15 |
| 1 | 8.428572 | 9 | 1 | 16 | 5 | 12 |

**Summary statistics for Premium:**   
Group by: Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gender** | **Mean** | **Median** | **Min** | **Max** | **Q1** | **Q3** |
| 0 | 69.034485 | 68 | 45 | 92 | 62 | 80 |
| 1 | 54.61905 | 50 | 36 | 88 | 45 | 60 |