Mini-Project duo group 7

Venkata Bapanapalli

Karthik Meyyappan

Contributions: Both members equally contributed to analytically solve and implement the code of the given two questions.

**Problem 1**

Chart, bar chart

Description automatically generated

Figure : Bar graph to compare Maine and Away runners

From the graph above it can be concluded that the majority of the runner population consists of runners from Maine. The total population of runners is 5875, out of which 4458 are from Maine and 1417 are Away runners. This equates to 75.9 % Maine runners and 24.1% Away runners.

Chart, histogram

Description automatically generated

Figure : Histogram showing the comparison between time and frequency for Maine runners.

Chart, histogram

Description automatically generated

Figure : Histogram showing the comparison between time and frequency for Away runners.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| From | Min | 1st Qu | Median | Mean | 3rd Qu | Max | SD | IQR | Range |
| Maine | 30.57 | 50.00 | 57.03 | 58.20 | 64.24 | 152.17 | 12.18 | 14.24 | 30.567 152.167 |
| Away | 27.78 | 49.15 | 56.92 | 57.82 | 64.83 | 133.71 | 13.83 | 15.67 | 27.782 133.710 |

Table 1: Shows the summary of statistics for Maine and Away runners.

From observing the graphs above it can be concluded that both distributions are skewed to the right. Runners from Maine have a higher min, mean, and max time compared to away runners. Away runners also have a smaller range compared to runners from Maine.

Chart, box and whisker chart

Description automatically generated

Figure : Shows the box plot representation of the histograms in the previous part.

Chart, box and whisker chart

Description automatically generated

Figure : Box plot to compare male and female runners’ age.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| From | Min | 1st Qu | Median | Mean | 3rd Qu | Max | SD | IQR | Range |
| Male | 9.00 | 30.00 | 41.00 | 40.45 | 51.00 | 83.00 | 13.99 | 21 | 9  83 |
| Female | 7.00 | 28.00 | 36.00 | 37.24 | 46.00 | 86.00 | 12.26 | 18 | 7  86 |

Table 2: Shows the summary of statistics for male and female runners.

From the graph and table above, it can be observed that male runners have greater statistics compared to female runner except the maximum age.

**Problem 2**

Chart

Description automatically generated

Figure : Box plot for fatal accidents.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Qu | Median | Mean | 3rd Qu | Max | SD | IQR | Range |
| Fatal Accidents | 0.00 | 6.00 | 13.50 | 17.02 | 23.00 | 60.00 | 13.81 | 17 | 0  60 |

Table 3: Shows the summary of statistics for fatal accidents.

To find the outliers the 25th and 75th quantiles were calculated. Any value that is more that 1.5 times the IQR from the 25th and 75th quantiles are outliers. Upon calculating the lower and upper bounds, it was discovered that the counties with the highest motorcycle fatale accidents are Greenville and Horry. This is likely due to a larger group of young drivers and negligent drivers.