ADA1

Assignment 1: Due August 30 Thursday in class

For assignment 1, please do all calculations by R.

"cars" is a built in dataset in R with 50 observations on two variables:

- "speed" gives the speed of cars (mph) and

- "dist" gives the distances taken to stop (ft)

- Please refer to columns in the data using the $ sign, if we use d1 < −cars, then "speed" column and dist column will be

  `d1$speed and d1$dist repectively`

- (a) [5pt] Plot dist (y axis) vs speed (x axis), have you seen a pattern?

  mention that we have seen a linear relationship, with the increase of speed, the distance needed to stop the car also increase.

- (b)[5pt] Compute the mean, median, standard deviation, and interquartile range for the dist data.

  mean= 42.98, median = 36, sd = 25.77, interquartile range= 56-26=30

- (c)[5pt] Make a stem-and-leaf display, histogram, and boxplot for the dist data. Is there much difference between the mean and median? Discuss, briely, whether the size and the direction of the difference is sensible, given the graphical summaries.

  there is not much difference between the mean and median, it is 6.98. Compared to sd 25.77, the difference is small. the plots indicate right skewed distribution, this is consistent with the finding that mean is bigger than the median, and is consistent with the boxplot and histogram.

- (d)[5pt] Using the graphical summaries, describe the shape of the dist distribution. Discuss modality, presence/absence of outliers, whether skewness is present, and if

so, in what direction, and whether it would be reasonable to assume that the dist distribution is normal.

right skewed distribution, unimodal, one outlier detected by boxplot, which is the maximum value of the dist, 120, go back to data, speed is 24mph, and dist needed to stop is 120. This may be a recording error. since the data is right skewed, it is not appropriate to assume a normal distribution.