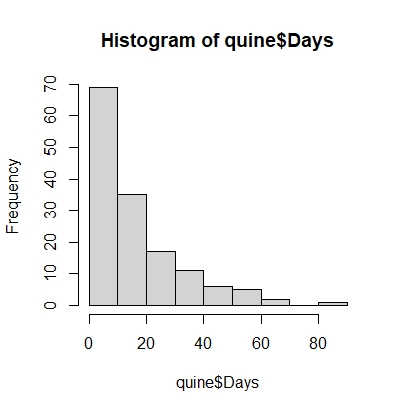
1. (3\*(mean(quine$Days)-median(quine$Days)))/sd(quine$Days)

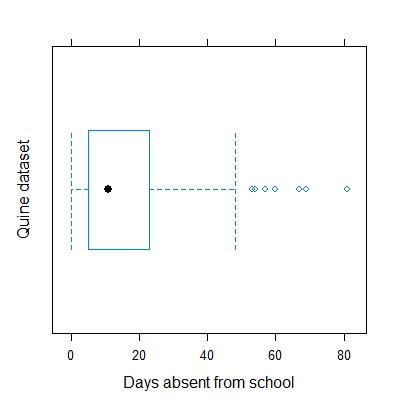
=1.007598, the data is skewed to the right

1. hist(quine$Days)



The histogram shows the Days graph is skewed to the right because less frequency of data is on the right. Question 1’s answer comes to the same conclusion.

1. bwplot(quine$Days, xlab = "Days absent from school", ylab = "Quine dataset")



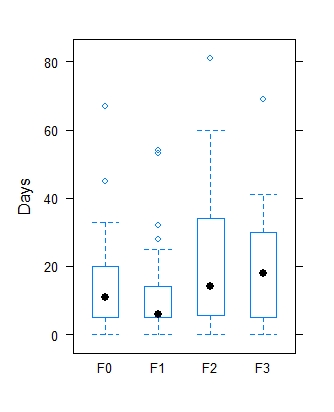
From the graph there seems to be 8 outliers on the higher end of days. The typical (50%) number of absences are the values fall within the box. From the looks of the graph, a typical student was absent between 5 and 22 days.

1. quantile(quine$Days)



This command shows the values each quarter mark of the data. 25% to 75% range is the typical range (the box in the boxplot). The days line up with what I gathered visually from the graph in question 4.

1. bwplot(Days~Age, data = quine)



Less than 25% of students in age group F1 had as many absences than the upper 50% of students in age group F2.

F2 age group students had the most absences.

More than 50% of students of age groups F0, F2, and F3 had more absences than 50% students of age group F1.

1. sd(Days~Sex+Age, data = quine)

A picture containing text

Description automatically generated

The most consistent grouped data set is males in age group F1 at 5.3060 standard deviation regarding day absences. The least consistent grouped data set is females in age group F2 at 23.1020 standard deviation.

1. quantile(Days~Lrn, data=quine, probs = seq(0.2, 0.8, 0.2))

quantile(Days~Sex, data=quine, probs = seq(0.2, 0.8, 0.2))

quantile(Days~Eth, data=quine, probs = seq(0.2, 0.8, 0.2))

Text

Description automatically generated

Since each variable has 2 states, perhaps the percentiles with more consistent values per percentile would be more accurate. Since if you had to guess if a student would be absent, the data closer to each other would give you a better chance to be right. Therefore, using learning status to predict absences may be more reliable.

1. options(digits = 2)

MaleStudents=filter(quine, quine$Sex == "Male")

primaryMaleStudents=filter(quine, quine$Age == "F0")

percent\_rank(primaryMaleStudents$Days)

Text, letter

Description automatically generated

1. zScores=as.data.frame(scale(quine$Days))

Graphical user interface, application

Description automatically generated

1. Question 9 satisfies Chebyshev’s Theorem but does not satisfy the Empirical Rule. Chebyshev’s Theorem k deviations are within the percent ranges 114, 138, and 143 out of 146. Empirical Rule is not satisfied because 143 (3 standard deviations) is not 99.7% of the values.