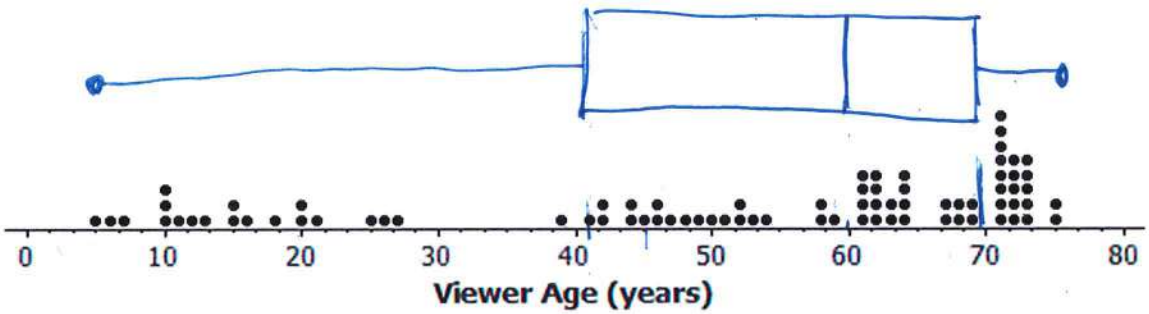
Box Plots, Interquartile Range, Outliers **Name:** Answers

Statistics – Algebra 1

Consider the following scenario. A television game show, "Fact or Fiction", was canceled after nine shows. Many people watched the nine shows and were rather upset when it was taken off the air. A random sample of eighty viewers of the show was selected. Viewers in the sample responded to several questions. The dot plot below shows the distribution of ages of these eighty viewers:

Dot Plot of Viewer Age



1. Approximately where would you locate the mean (balance point) in the above distribution?

between 50 & 60 or so?

2. How does the direction of the tail affect the location of the mean age compared to the median age?

The tail means the data distribution is skewed left so the mean < median.

3. The mean age of the above sample is approximately 50. Do you think this age describes the typical viewer of this show? Explain your answer.

50 doesn't seem typical. Older people seemed to watch the show

4. Using the above dot plot, construct a box plot over the dot plot by completing the following steps:
   i. Locate the middle 40 observations and draw a box around these values. (Why are we putting the 40 middle values in the box?)
   ii. Calculate the median and then draw a line in the box at the location of the median. median = 60
   iii. Draw a line that extends from the upper end of the box to the largest observation in the data set.
   iv. Draw a line that extends from the lower edge of the box to the minimum value in the data set.

5. Recall that the 5 values used to construct the dot plot makes up the 5–number summary. What is the 5–number summary for this data set of ages?

Minimum age: 5

Lower quartile or Q1: 41 not 40

Median Age: 60

Upper quartile or Q3: 70

Maximum age: 75

6. What percent of the data does the box part of the box plot capture?

50% of data, by definition

7. What percent of the data falls between the minimum value and Q1?

25%, by definition

8. What percent of the data falls between Q3 and the maximum value?

25%, by definition.

9. What percent of the viewers have ages between Q1 and Q3? Remember: The difference between Q3 and Q1, or Q3 – Q1, is called the **interquartile range** or **IQR**. What is the interquartile range (IQR) for this data distribution?

50% of data lie between Q_1 + Q_3

$IQR = Q_3 - Q_1 = 70 - 41 = 29$ or ~~70-40=30~~

10. An outlier is defined to be any data value that is more than 1.5×(IQR) away from the nearest quartile. According to this measure, does this set of data have any outliers?

$1.5 \times 29 = 43.5$

No outliers

$Q_3 + 1.5\,IQR = 70 + 43.5 = 113.5$

$Q_1 - 1.5\,IQR = 41 - 43.5 = -3.5$

No data pts are greater than 113.5 or less than -3.5

An advertising agency researched the ages of viewers most interested in various types of television ads. Consider the following summaries:

Ages	Target Products or Services
30–45	Electronics, home goods, cars
46–55	Financial services, appliances, furniture
56–72	Retirement planning, cruises, health care services

11. The mean age of the people surveyed is approximately 50 years old. As a result, the producers of the show decided to obtain advertisers for a typical viewer of 50 years old. According to the table, what products or services do you think the producers will target? Based on the sample, what percent of the people surveyed would have been interested in these commercials if the advertising table were accurate?

- Financial services, appliances, furniture
- About 12 people are between 46-55 yrs old

$\frac{12}{80} = .15 \times 100\% = 15\%$

12. The show failed to generate the interest the advertisers hoped. As a result, they stopped advertising on the show and the show was cancelled. Kristin made the argument that a better age to describe the typical viewer is the median age. What is the median age of the sample? What products or services does the advertising table suggest for viewers if the median age is considered as a description of the typical viewer?

Median = 60 years

Retirement planning, cruises, health care

13. What percentage of the people surveyed would be interested in the products or services suggested by the advertising table if the median age were used to describe a typical viewer?

About 32 people are between 56-72 yrs.

$\frac{32}{80} = .4 \times 100\% = 40\%$

14. What percent of the viewers have ages between Q1 and Q3 (i.e. what is the IQR) for this data distribution?

50% (always)

15. The IQR provides a summary of the variability for a skewed data distribution. The IQR is a number that specifies the length of the interval that contains the middle half of the ages of viewers. Do you think producers of the show would prefer a show that has a small or large interquartile range? Explain your answer. A smaller IQR means less variability. A small interquartile range would allow advertisers to choose an age range to target. A larger IQR means more variability, which means the show is popular across ages.

16. Do you agree with Kristin's argument that the median age provides a better description of a typical viewer? Explain your answer.

Yes. Since the data is skewed left, the mean is pulled left and is therefore less representative of the data. The median gives a more typical age.