

Warm Up (12/4/17)

Calculate the deviation from the mean for the following set of data:

Data Set	8	9	4	2	1	6	5
Deviation	3	4	-1	-3	-4	1	0

$$\text{Deviation} = X - \text{mean}$$

$$\text{Mean} = \frac{8+9+4+2+1+6+5}{7} = \frac{35}{7} = 5$$

$$8 - 5 = 3$$

Module 2: Lesson 7

Measuring Variability for Skewed Distributions (Interquartile Range & Box Plots)

Lesson Summary

- Nonsymmetrical data distributions are referred to as skewed.
- Left-skewed or skewed to the left means the data spread out longer (like a tail) on the left side.
- Right-skewed or skewed to the right means the data spread out longer (like a tail) on the right side.
- The center of a skewed data distribution is described by the median.
- Variability of a skewed data distribution is described by the interquartile range (IQR).
- The IQR describes variability by specifying the length of the interval that contains the middle 50% of the data values.
- Outliers in a data set are defined as those values more than $1.5 \times (IQR)$ from the nearest quartile. Outliers are usually identified by an "*" or a "•" in a box plot.

Box Plot (Box-and-Whisker Plot)

A box plot (or a box-and-whisker plot) is a way that we can organize data which highlights the middle data values.

5-Number Summary

Minimum - The smallest value in the data set

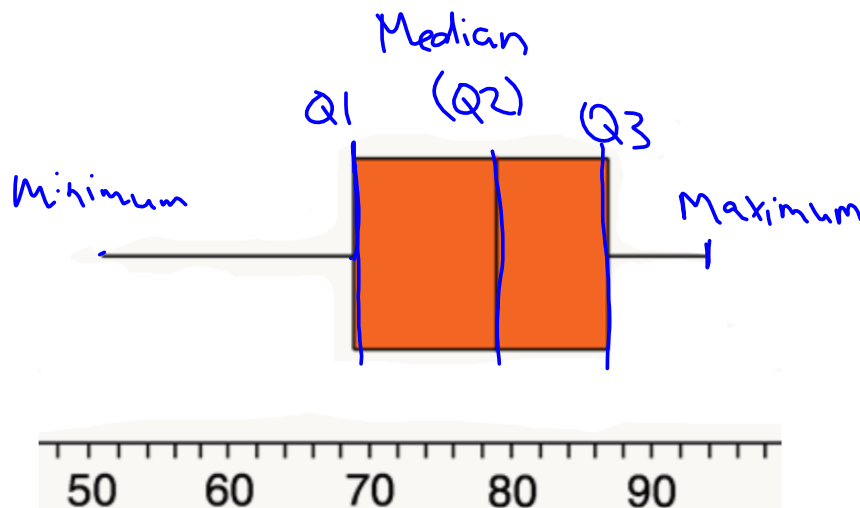
Q1 – Quartile 1, the median of the lower half of the data set

Q2 – Quartile 2, the median of the entire data set

Q3 – Quartile 3, the median of the upper half of the data set

Maximum - The largest value in the data set

IQR – Interquartile range, the difference from Q3 to Q1



Steps for Constructing a Box-and-Whisker Plot

1. Order the values from least to greatest.
2. Find the median of the data (the value in the middle).
3. Find the median of the data values that are less than the median. (This will be the value for Q1).
4. Find the median of the data values that are greater than the median. (This will be Q3).
5. Find the minimum and maximum values.
6. Create a number line that will contain all of the data values.
7. Draw a box from Q1 to Q3 and a vertical line at the median.
8. Extend "whiskers" from the box to the minimum and maximum values.

Interquartile Range and Outliers

Interquartile Range (IQR) is a measure of variability. It is calculated by taking the difference of Q3 and Q1.

Outliers are values that are distant from the current set of data and are sometimes excluded from the set of data.

A point is considered an outlier if:

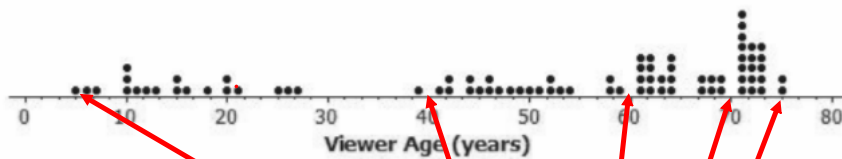
$$\text{Data Value} > Q3 + (1.5)IQR$$

Or

$$\text{Data Value} < Q1 - (1.5)IQR$$

Consider the following scenario. A television game show, *Fact or Fiction*, was cancelled 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



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

- Minimum age: 6
- Lower quartile or Q1: 40
- Median age: 60
- Upper quartile or Q3: 70
- Maximum age: 75

Create a box plot from this 5-Number summary and calculate the Interquartile Range (IQR):

Minimum age: 5

Lower quartile or Q1: 40

Median Age: 60

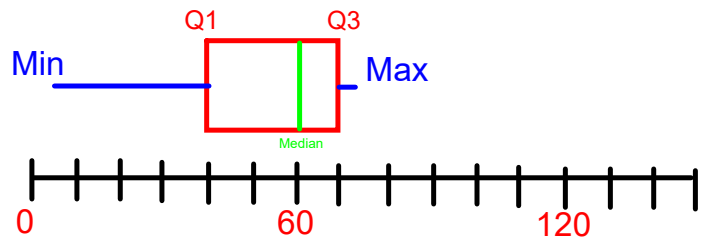
Upper quartile or Q2: 70

Maximum age: 75

IQR: $IQR = Q3 - Q1$

$$IQR = 70 - 40$$

$$IQR = 30$$



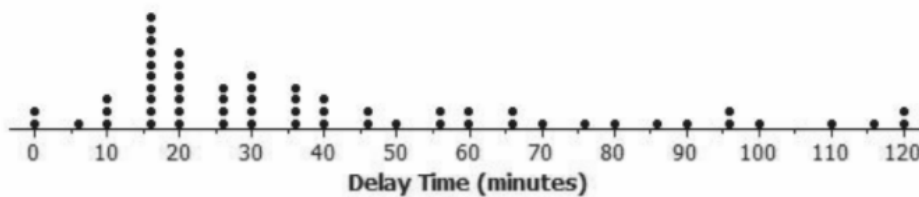
Consider the following scenario. Transportation officials collect data on flight delays (the number of minutes a flight takes off after its scheduled time).

Consider the dot plot of the delay times in minutes for 60 BigAir flights during December 2012:

Dot Plot of December Delay Times

Module 2
Pg 52

Warm up for 12/5/17



1. Find the 5-number summary and calculate the Interquartile Range.
2. Create a box plot based on the data above.