PDF File

PROCESS DATA

In this unit, you will examine ways to record data in an efficient and organized manner. You will examine the standard stem-and-leaf plots and also back-to-back stem-and-leaf plots. You will analyze data using the measures of central tendency and measures of variation.

Stem-and-Leaf Plots

A way to organize data is by using a stem-and-leaf plot. The leaf is the last digit of a number. The stem is remaining digits of a number. Stems are listed in numerical order to create categories.

In the example shown below, the leaves is the ones digit of the numbers and the stems are first digits (tens digits).

Example: Arrange the following numbers in a stem-and-leaf plot: 58, 72, 53, 95, 62, 67, 58, 84, 77, 52, 89, 96, 64, 81.

First we'll make a rough draft of our stem-and-leaf plot.

first digits (stem) / ones digits (leaf)

stem	leaf			
5	8	3	8	2
6	2	7	4	
7	2	7		
8	4	9	1	
9	5	6		

Key: 5 | 8 represents 58.

*The 5 in the stem column together with the 8 in the leaf column represent the number 58 from the data points.

Now, we'll rewrite the stem-and-leaf plot placing the leaves in order from left to right.

first digits (stem) / ones digits (leaf)

stem	leaf			
5	2	3	8	8
6	2	4	7	
7	2	7		
8	1	4	9	
9	5	6		

Key: 5 | 8 represents 58.

QuickTime Stem-and-Leaf Plots--Dog Weights (02:09)

Back-to-Back Stem-and-Leaf Plots

A back-to-back stem-and-leaf plot is used when comparing two sets of data. The stems are in the center and the left leaves are read in reverse.

Example: Refer to the table to create a back-to-back stem-and-leaf plot of a basketball team´s winning and losing scores for the season.

Measures of Central Tendency

To analyze sets of data, researchers often try to find a number that can represent the whole set. These numbers or pieces of data are called **measures of central tendency**. The three common measures we are going to study are: **mean, mode,** and **median**.

Mean: The mean of a set of data is the sum of all the data divided by the number of pieces of data (average).

Mode: The mode of a set of data is the number that occurs most often.

Median: The median of a set of data is the number in the middle when the data are arranged in order. When there are two middle numbers, the median is the average (mean) of the two numbers.

Let's take a look at Amanda's data from her class contest and determine the mean, mode, and median.

6	**4**	**4**	**10**	**6**
5	**2**	**4**	**1**	**5**
3	**3**	**7**	**4**	**2**
0	**9**	**5**	**7**	**10**

Mean: Add all the numbers, and then divide the sum by the number of numbers in the set.

6 + 4 + 4 + 10 + 6 + 5 + 2 + 4+ 1 + 5 + 3 + 3 + 7 + 4 + 2 + 0 + 9 + 5 + 7 + 10 = 97

$$97 \div 20 = 4.85$$

4.85 is the mean of the set of data.

Mode: Arrange the numbers in order from the smallest to the largest and determine which number occurs most often.

0, 1, 2, 2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7, 7, 9, 10, 10

4 is the mode of the set of data because it occurs most often.

Median: Use the arranged numbers from the mode and determine what the middle number is.

0 1 2 2 3 3 4 4 4 4 5 5 5 6 6 7 7 9 10 10

Introduction: Center and Spread (02:30)

Since there are two middle numbers, we must add them together and find the mean.

Measures of Variation

The study of **quartiles** helps us learn about the nature and tendency of data. Quartiles divide data that is arranged in order from least to greatest into four equal parts. The ***median***, sometimes referred to as the Second Quartile, separates the data in half. The ***Lower Quartile (LQ),*** sometimes referred to as the First Quartile, is the median of the first half of the data. The ***Upper Quartile (UQ)***, sometimes referred to as the Third Quartile, is the median of the second half of the data. The ***range*** of the data is the difference between the highest and lowest data values and is found by subtracting these values. The ***Interquartile Range*** is the range of the middle half of the data and is found by subtracting the lower quartile from the upper quartile **(UQ – LQ).** Many companies analyze data to determine the promotion and implementation of their product.

Below are several examples of data arranged in order from least to greatest. The median, LQ, UQ, and interquartile range, and range have been determined for each set of data.

If there is an even number of numbers in the data, there will be two middle numbers. To find the median, calculate the average of the two middle numbers.

*Note: The first step in determining quartiles is to put the **data in order** from least to greatest.

Example 1:

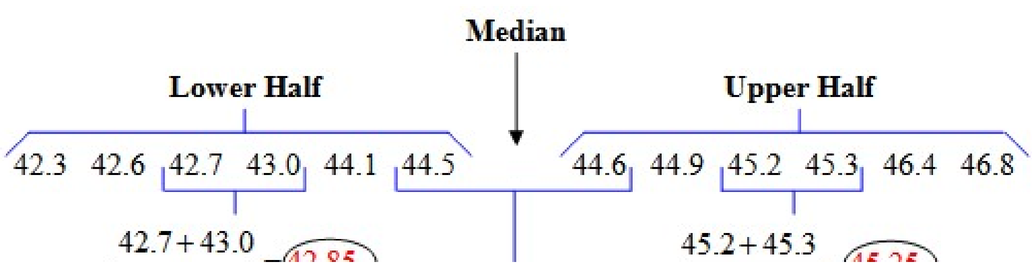
In this example there are an even number of data points in the data; thus, the **Second Quartile (median)** is the average of the two middle numbers. (44.55)

There are an even number of data points in the lower half of the data; thus, the **First Quartile** is the average of the two middle numbers in the lower half. (42.85)

There are also an even number of data points in the upper half of the data; thus, the **Third Quartile** is the average of the two middle numbers in the upper half. (45.25)

The **range** is the difference between the highest and lowest data points. (4.5)

The **interquartile range** is the difference between the Upper Quartile (Third Quartile) and the Lower Quartile (First Quartile). (2.4)



QuickTime Box-and-Whisker Plots--Popular Dog Size (02:58)

Calculate Outliers

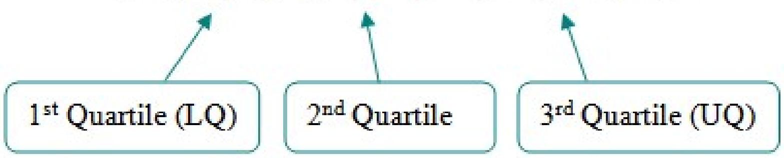
Outlier – An outlier is a data point widely separated from the main cluster of points in a sample of data.

To check for outliers follow these steps:

1.) Subtract the first quartile from the third quartile. (UQ – LQ)
2.) Multiply the difference by 1.5
3.) Add the product found in Step 2 to the third quartile (UQ). Compare to see if any numbers in the data are greater than this sum. If so, the data point is an outlier.
4.) Subtract the product found in Step 2 from the first quartile (LQ). Compare to see if any numbers in the data are less than this sum. If so, the data point is an outlier.

Example: Check for outliers.

2 6 7 8 9 12 15 15 18 24 45



Step 1: $18 - 7 = 11$

Step 2: $11 \times 1.5 = 16.5$

Step 3: $18 + 16.5 = 34.5$ *Any data points above 34.5?* Yes! Data point 45 is above 34.5

Step 4: $7 - 16.5 = -9.5$ *Any data points below −9.6?* No!

There is one outlier, **45**.

QuickTime Outliers (09:56)

Review Worksheets

The review worksheets are provided to give extra practice in skill areas presented in the unit. The worksheets are *optional* unless otherwise specified by the instructor. The worksheets are Adobe Acrobat files. Click on the pencil icon to open the document. Save the document to a folder on the computer, and then enter answers for the problems in the textboxes. Once the document is completed, make sure to save it again, and then send the document to the instructor via email. The answer key provided is for the instructors only and is password protected.

Range = Highest Value – Lowest Value

Range = 40– 3 = 37
Interquartile Range = UQ – LQ
Interquartile Range = 32.5 – 13 = 19.5

Practice Worksheet: **Mean, Median, Mode, and Range**

Answer Key (Password Protected)

Example 3

Practice Worksheet: **Analyze Data**

Answer Key (Password Protected)

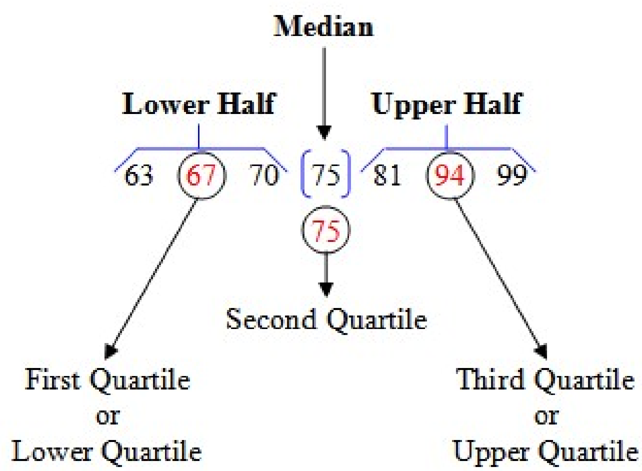
...is example there are an odd number of data points in the data; thus, the **Second Quartile (median)** is the middle number. (75)

There are an odd number of data points in the lower half of the data; thus, the **First Quartile** is the middle number in the lower half. (67)

There are also an odd number of data points in the upper half of the data; thus, the **Third Quartile** is the middle number in the upper half. (94)

The **range** is the difference between the highest and lowest data points. (36)

The **interquartile range** is the difference between the Upper Quartile (Third Quartile) and the Lower Quartile (First Quartile). (27)



Range = Highest Value – Lowest Value
Range = 99 – 63 = 36
Interquartile Range = UQ – LQ
Interquartile Range = 94 – 67 = 27