Lesson 58: Statistics

By the end of the lesson, we will be able to:

~ Use our calculators to find:
- Mean, Median, Min, Max
- Quartiles 1 and 3
- Standard Deviaion
- Interquartile Range (IQR)

~ Make box plots

~ Make histograms

~ Understand and calculate percentiles.

BOXPLOTS, IQR, & STANDARD DEVIATION:

When we use the phrase "measures of central tendency", we are talking about interpreting data based on the central data values. Some of these measures you have seen before - mean, median, and mode.

Lesson 58: Statistics

Mean ($\bar{X}$) is the arithmetic average.

Add all the values together, then divide by the number of values. (It is easily influenced by outliers.)

Median (_med_) is the middle value. (Sometimes called Quartile 2).

Make sure the data are in order before finding the middle value. If there is an even number of values, find the mean of the two center values. (The median is resistant to outliers.)

Mode is the most frequent value (the one that appears most often). A data set may have more than one mode or no mode.

Lesson 58: Statistics

Example A:

Find the mean, median, and mode of the following data set:

Number of days from mailing a letter to the delivery: 6, 4, 3, 4, 2, 5, 3, 4, 5, 2, 3, 4

List the data in ascending order to find median.

____ ____ ____ ____ ____ ____ ____ ____ ____ ____ ____ ____

Mean: _____   Median: _____   Mode: _____

Example A CONT with Calculators:

Find the mean, median, and mode of the following data set:

Number of days from mailing a letter to the delivery: 6, 4, 3, 4, 2, 5, 3, 4, 5, 2, 3, 4

Calculator: Input the list of values in list 1 (L1) under STAT, EDIT, Edit (enter numbers). Then use STAT, CALC, 1-Var Stats, then hit L1 (normally "2nd" "1") and "enter".

Mean: _____   Median: _____   Mode: _____

Screen 1: EDIT CALC TESTS
1: Edit…
2: SortA(
3: SortD(
4: ClrList
5: SetUpEditor

Screen 2:
L1	L2	L3	1
6	------	------	
4			
4			
2			
5			
3			

L1(1)=6

Screen 3: EDIT CALC TESTS
1: 1-Var Stats
2: 2-Var Stats
3: Med-Med
4: LinReg(ax+b)
5: QuadReg
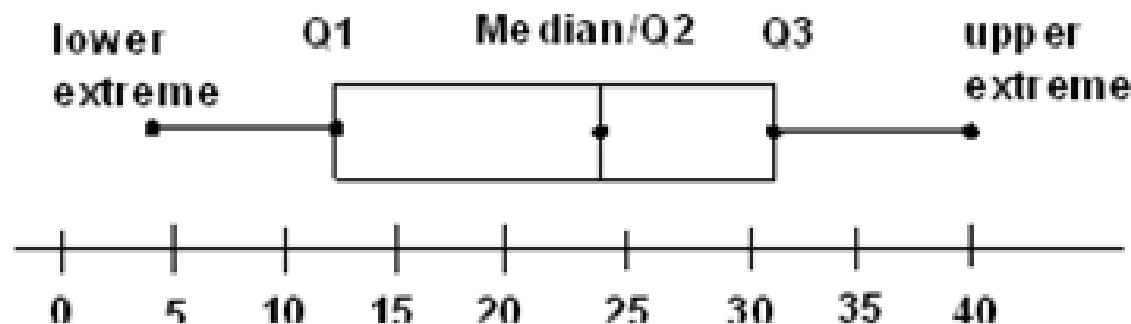6: CubicReg
7↓QuartReg

Screen 4: 1-Var Stats L1

Screen 5: 1-Var Stats
$\bar{x}=3.75$
$\Sigma x=45$
$\Sigma x^2=185$
$Sx=1.215431087$
$\sigma x=1.16368667$
↓n=12

Screen 6: 1-Var Stats
↑n=12
minX=2
$Q_1=3$
Med=4
$Q_3=4.5$
maxX=6

Range, Interquartile Range, and Standard Deviation

--- are measures of **variability** or spread. They measure the way that the data is <u>spread out</u> (clustered together, evenly distributed, skewed, etc.)

Lesson 58: Statistics

Range is the difference between the max and the min values. The range is a simple measure of spread. Extreme low or high values (outliers) will have a big influence on the value of the range.

$$\text{Range} = \text{max} - \text{min}$$

Quartiles divide the data in to 4 parts, just as medians divide data into 2 parts. Quartiles are the medians of the lower and upper halves of the data set. If there are an odd number of data values, do not include the median in either half. The five-number summary of data includes:

Min - Q1 - Med - Q3 - Max

** Each Quartile contains 25% of the data **

Lesson 58: Statistics

Interquartile Range (IQR) is the difference between the first and third quartiles ($Q_3 - Q_1$).

$$IQR = Q_3 - Q_1$$

The IQR is a useful measure of spread when trying to avoid the strong influence of low or high outliers. It measures the spread of the middle 50% of the data.

Outliers in a data set are any values that are farther than 1.5 times the IQR beyond the upper or lower quartile.

Remember Outliers have a big affect on the mean, standard deviation, and the range of the data. Outliers do not affect   median or the IQR.

Lesson 58: Statistics

Standard Deviation (σ_x or s_x) is a measure of the spread often used in "normally distributed" data sets.
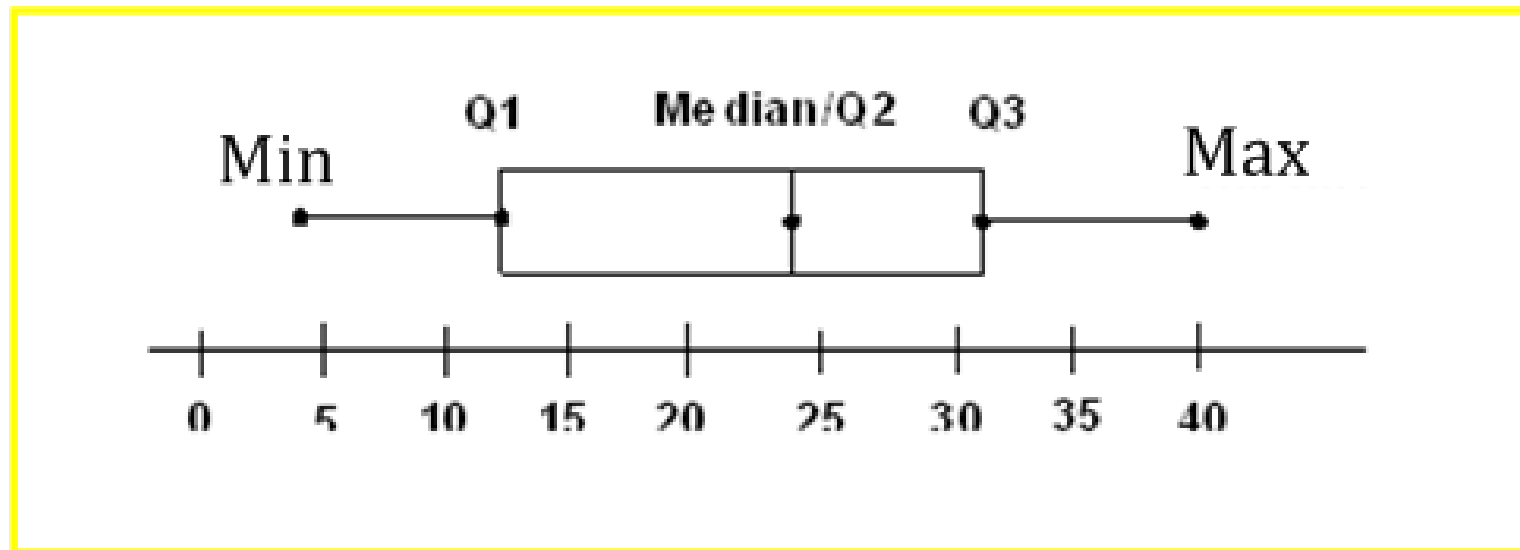In this lesson, we will use σ_x.

For many data sets, about **68%** of the data points fall within 1 standard deviation of the mean, and about 95% of the data points fall within 2 standard deviations of the mean.

Lesson 58: Statistics

A **box-and-whisker plot**, or boxplot, is a graph showing the distribution of data using the five-number-summary.

Look at the general boxplot below:

Lesson 58: Statistics

Example B: Find the mean, range, IQR, standard deviation, and sketch the box plot.

1. The data values represent the number of milligrams of a substance in a patient's blood, found on consecutive doctor visits.   14, 13, 16, 9, 3, 7, 11, 12, 11, 4

2. Put data into L1 (list 1) in your calculator.

3. Run the 1-Var Stats on L1. Find the following:

    Min:_____   Q1:_____   Med:_____   Q3:_____

    Max:_____   SD:_____   Mean:_____

4. Sketch the box plot. (make sure you show the scale below)

<u>Lesson 58: Statistics</u>

Calculator: Input the data in list 1 (L1) under STAT, EDIT, Edit. Then use STAT, CALC, 1-Var Stats L_1. Search down the list for Q1, Q3, and σ .

5. Calculate the **range**. Range= _____

6. Calculate the **IQR**. IQR= _____

7. About **50%** of the data points fall between: _____

8. About **68%** of the data points fall between: _____

HISTOGRAMS & PERCENTILES:

Histograms are similar to bar graphs because both graphs show frequencies of data. However, there are some important differences:

- Histograms display *numerical* data (like average height), while bar graphs display categorical data (like eye color).

- Histogram bars represent numerical *intervals*, while bar graph bars represent different categories.

- Histogram bars are directly *adjacent*, while bar graph bars are spaced apart.

Lesson 58: Statistics

Percentile is used to rank data values in comparison with the entire data set.

National standardized test scores are frequently reported in terms of percentiles, rather than raw scores.

Percentile *tells what percent of the data is equal to or below that value.*

Note that there is a difference between a percentage and a percentile:

~ a *percentage* compares the number of occurrences to the **total** possible occurrences

~ a *percentile* compares the number of occurrences at that value or below to the total possible occurrences.

Example C:

Assume your score report for the ACT test stated that you were in the 48th percentile for the science section and 71st percentile for the reading section. Explain what this means.

Lesson 58: Statistics

Example D: The 1995 World Almanac reported the following data for record low temperatures for the 50 state in the US. Use the histogram of the data to answer the questions.

Temp °F	# of states
-80 to -66	2
-65 to -51	10
-50 to -36	17
-35 to -19	16
-18 to -2	4
-1 to 15	1

Calculate the percentages!
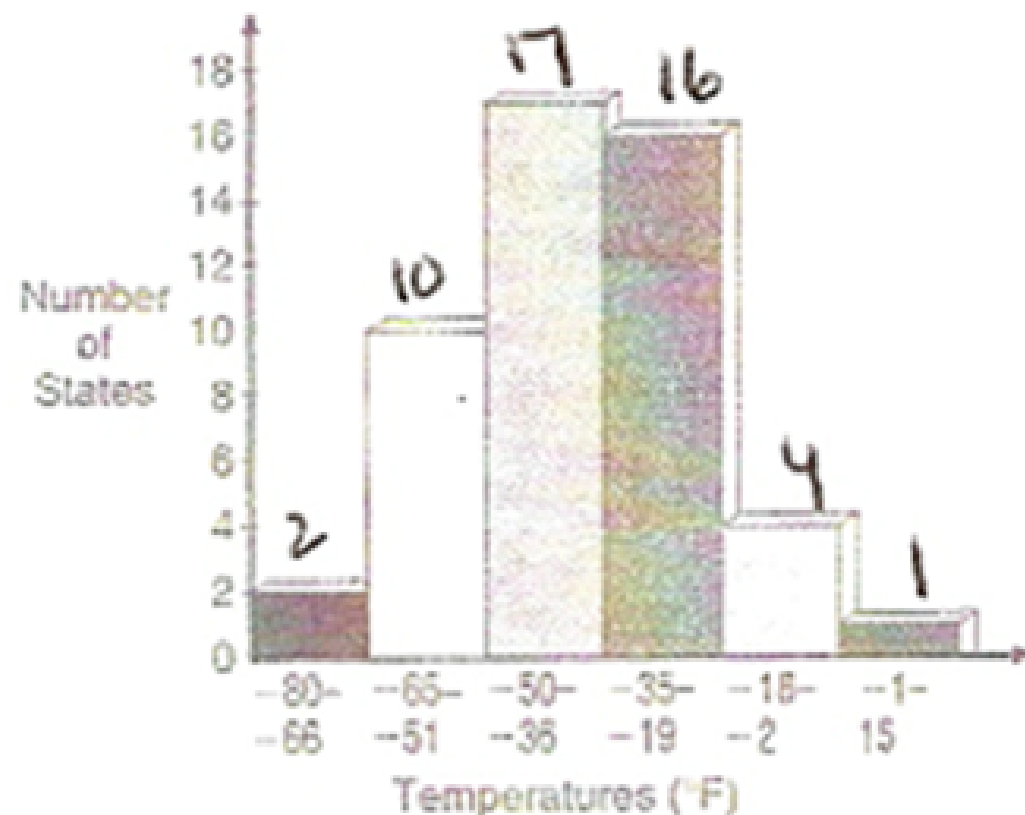
Lesson 58: Statistics

1. What percent of the states have record low temperatures between -65 degrees and -36 degrees?

2. What percent of the states have record low temperatures of −36 degrees or colder?

3. What percent of the states have record low temperatures of −18 degrees of warmer?

Lesson 58: Statistics

Percentiles: Calculate percent for EACH column and then add the percents together...

4. Calculate the **percentile** for a state which has a record low temperature of -2 degrees?

5. Calculate the **percentile** for a state which has a record low temperature of -36 degrees?

Example E:

a) Find the Mean and St. Deviation to the nearest hundredth.

b) The number 5.4 is at what percent?

Stem	Leaf
4	3 5 6 8
5	2 4 5 6
6	1 2 4 5 5 6 7 7 7

Key: 4|5 = 4.5

Example F:

a) Find the Mean and St. Deviation to the nearest hundredth.

b) The number 59 is at what percent? _____

Stem	Leaf
5	3 4 5 5 6
•	7 7 7 7 8 9
6	2 3 4 5 6

Key: 6|5 = 65

Lesson 58: Statistics

By the end of the lesson, we will be able to:

~ Use our calculators to find:
- Mean, Median, Min, Max
- Quartiles 1 and 3
- Standard Deviaion
- Interquartile Range (IQR)

~ Make box plots

~ Make histograms

~ Understand and calculate percentiles.

Homework:

Assignment 58

- Pg 240 #9, 17: don't find outliers.