

AP Stats

Name _____

Unit 01 – Univariate Data

Day 3 Notes

**Measures of Center, Measures of Spread, Outliers,
Linear Transformations, and Comparing Distributions**

The following data are travel times for fifteen people to get to work:

20 30 10 40 25 20 10 60 15 40 5 30 12 10 10

Rewrite the numbers in order from least to greatest:

Make a dotplot of the data for a visual representation:



Measures of Center:

Median:

Mean:

Measures of Spread:

5 # summary with IQR:

Standard Deviation:

Comparing the Mean and Median:

The mean and median of a roughly symmetric distribution are close together. If the distribution is exactly symmetric, the mean and median are exactly the same. In a skewed distribution the mean is usually farther out in the long tail than its median. If the outliers were to increase, it would increase the mean but the median would stay the same.

Standard Deviation by hand:

Given the data set of: 1, 3, 4, 4, 8 find the standard deviation by hand.

Standard deviation/1 variable statistics in calculator:

I.2 Describing Distributions with Numbers

When you first encounter a dataset, it is a good habit to study a graphical display and estimate the SOCS. However, for a more detailed understanding of data, we must calculate numeric summaries of the center and spread. *Note: Be sure you understand how the following measures are calculated before relying on the TI to do the mechanics for you.*

The most common measures of center for a dataset are **mean** (\bar{x}) and **median** (Q2). The most common measures of spread/variability for a dataset are **range** (max-min), **interquartile range "IQR"** (Q3-Q1), and **standard deviation** (s_x).

Calculating Numeric Summaries

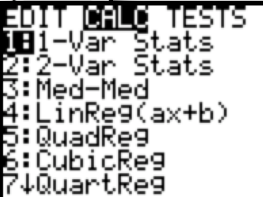
The calculation of each of these measures, especially the standard deviation, can be quite tedious. Thankfully, the TI can automate those calculations for us. Like plotting data, the calculator requires that you enter the dataset before it can report a numeric summary. If you haven't done so already, enter the Bonds and Aaron data into **[STAT] Edit... L1** and **L2**, respectively.

1. Enter data in to into **[STAT] 1:Edit...**

2. Press **[STAT] CALC 1:1-Var Stats [ENTER]**

| L1 | L2 | L3 |
|----|----|----|
| 16 | 13 | |
| 25 | 27 | |
| 24 | 26 | |
| 19 | 44 | |
| 33 | 30 | |
| 25 | 39 | |
| 34 | 40 | |

L3(1)=

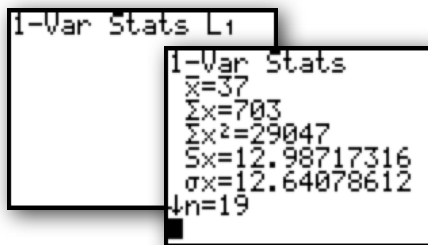


3. Your homescreen should read "1-Var Stats"

4. Press **2ND [1] (L1) [ENTER]**

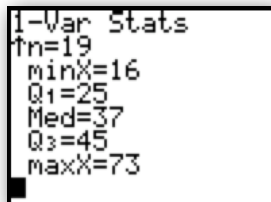
5. A numeric summary of the Bonds data should appear.

6. Repeat Steps 2 through 4 for **L2** to get a numeric summary of the Aaron data.



| 1-Var Stats L1 |
|-----------------------------|
| x̄=37 |
| Σx=703 |
| Σx ² =29047 |
| S _x =12.98717316 |
| σ _x =12.64078612 |
| n=19 |

7. Scroll down on each numeric summary to see the 5-number summary.



| 1-Var Stats |
|-------------|
| n=19 |
| minX=16 |
| Q1=25 |
| Med=37 |
| Q3=45 |
| maxX=73 |

Remember to interpret the numeric summary in the context of the problem!

WHICH MEASURE OF CENTER/SPREAD DO YOU USE?

| | Reasonably Symmetric/No Outliers | Strongly Skewed or has Outliers |
|----------|----------------------------------|---------------------------------|
| Center | | |
| Spread | | |
| Outliers | | |

Determine if there are outliers from the problem on the previous page.

Outliers:

Ex. The follow data compares the amount of text messages that males and females send in two days:

| | | | | | | | | | | | | | | | |
|---------|-----|-----|-----|----|-----|-----|-----|----|-----|-----|----|----|-----|----|-----|
| Males | 127 | 44 | 28 | 83 | 0 | 6 | 78 | 6 | 5 | 213 | 73 | 20 | 214 | 28 | 11 |
| Females | 112 | 203 | 102 | 54 | 379 | 305 | 179 | 24 | 127 | 65 | 41 | 27 | 298 | 0 | 130 |

Question:

Does the data give convincing evidence that the females text more than males?

1. Create a side by side box plot comparing the two data sets



2. Calculate the one-variable stats

| | Mean | SD | Min | Q1 | Median | Q3 | Max | IQR |
|--------|------|----|-----|----|--------|----|-----|-----|
| Male | | | | | | | | |
| Female | | | | | | | | |

3. Which "center of spread" test should best be used to describe this data and why?

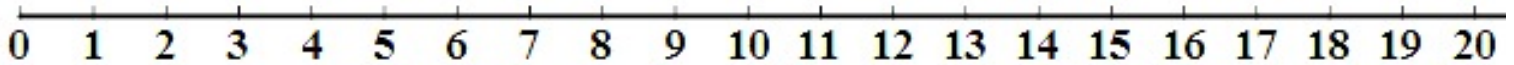
4. Finally compare SOCS to determine if the data is proof that females text more.

Changing the Unit of Measurement:

The following data represents the number of people who ran for senior class president in the past 10 years:

1, 5, 3, 2, 4, 4, 2, 2, 1, 3

1. Create a box plot and calculate mean and standard deviation for this data.



2. Now let's say that we add 10 to each number. What would this do to the box plot? Compare this box plot, mean, and standard deviation with the one above (graph on same number line above)
3. Go back to the original data and multiply each number by 3. What would this do to the box plot? Compare this box plot, mean, and standard deviation with the ones above (graph on the same number line above).

WHAT DO YOU NOTICE?

Resistance:

Resistant measures:

Non-resistant measures: