Statistics Lab 5 **Name:** _____

The sample represents final exam scores from a geography class.

78 72 83 79 85 64 69 87 72 63 74

77 52 38 91 66 97 90 74 63 94 68 42

Use the sample given above to find the value of each statistic below.

1. (2 points) mean 1. _____

2. (2 points) median 2. _____

3. (2 points) mode 3. _____

4. (2 points) standard deviation 4. _____

5. (2 points) variance 5. _____

6. (2 points) range 6. _____

7. (2 points) minimum 7. _____

8. (2 points) quartile 1 8. _____

9. (2 points) quartile 2 9. _____

10. (2 points) quartile 3 10. _____

11. (2 points) maximum 11. _____

The data below is a sample of retirement ages of 12 randomly selected doctors in San Diego. (Section 2.1 Exercise 39)

70 54 55 71 57 58 63 65 60 66 57 62

Use the sample given above to find the value of each statistic below.Include the correct units in your answers!.

12. (2 points) mean 12. _____

13. (2 points) median 13. _____

14. (2 points) mode 14. _____

15. (2 points) standard deviation 15. _____

16. (2 points) variance 16. _____

17. (2 points) range 17. _____

18. (2 points) minimum 18. _____

19. (2 points) quartile 1 19. _____

20. (2 points) quartile 2 20. _____

21. (2 points) quartile 3 21. _____

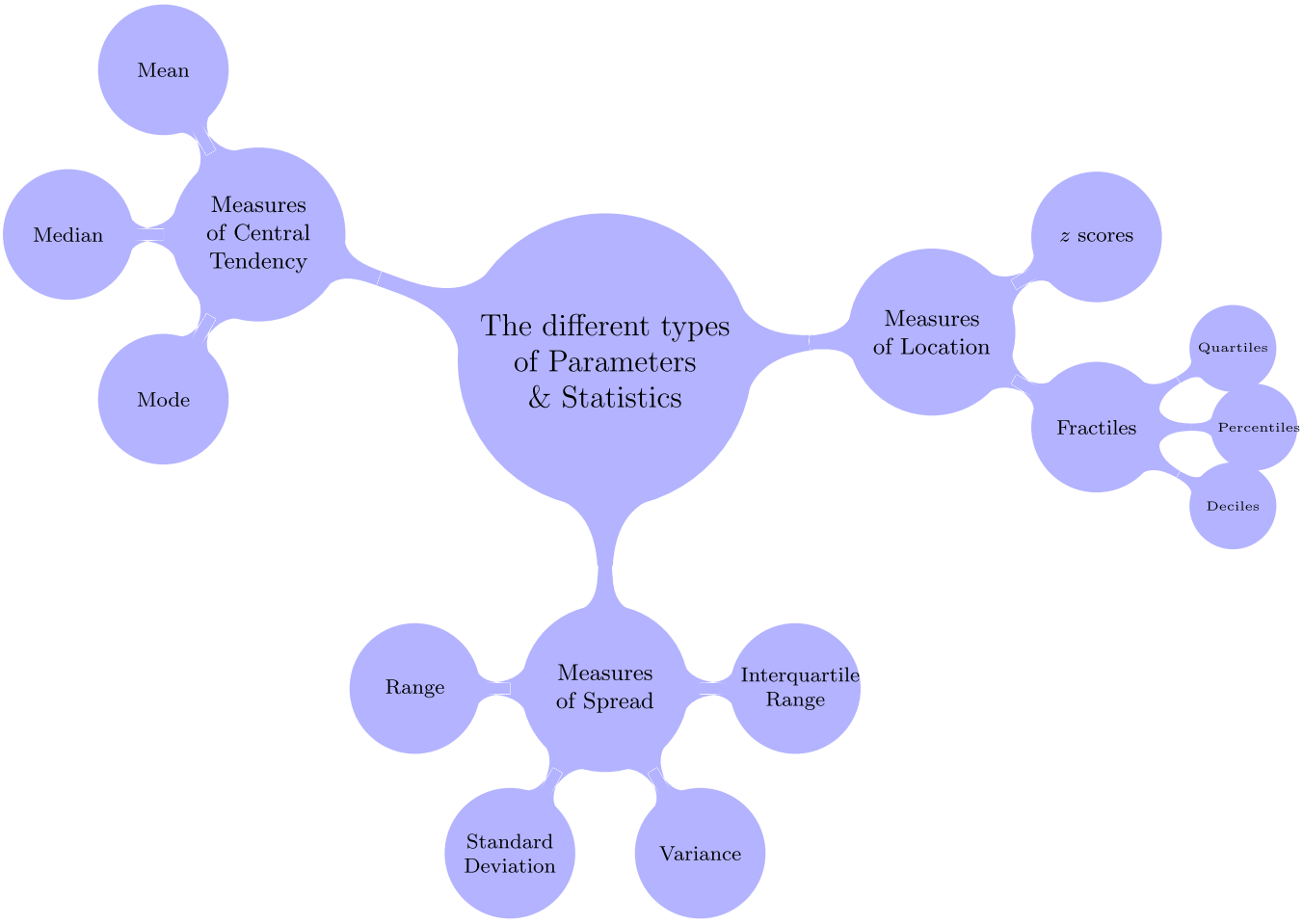
22. (2 points) maximum 22. _____

TABLE 3.1 Choosing Appropriate Measures for Describing Center and Spread

If the Shape of the Data Distribution Is...	Describe Center and Spread Using...
Approximately symmetric	Mean and standard deviation
Skewed or has outliers	Median and interquartile range

Figure 1: Table 3.1, pg 112

The different types of parameters and statistics for quantitative data can be grouped into three different categories: Measures of Central Tendency, Measures of Spread and Measures of Location. The next three sections in the workbook and textbook are devoted to those three topics.



Definition **Measures of Central Tendency** are numbers that represent the typical value in a data set. Measures of Central Tendency measure the location of the center of the data.

Definition **Measures of Spread** (also called Measures of Variations) are numbers that measure how spread out a data set is along the x axis. The four common measures of spread we can find for a quantitative data set are: the range, the variance, the standard deviation and the interquartile range.

Definition **Measures of Location** (also called Measures of Relative Standing) are numbers that measure the location of data value relative to the center of the data (z-scores); and relative to the other data values (fractiles).

23. The accompanying data on number of cell phone minutes used in one month are consistent with summary statistics published in a report of a marketing study of San Diego residents (Tele-Truth, March, 2009):

189	0	189	177	106	201	0	212	0	306
0	0	59	224	0	189	142	83	71	165
236	0	142	236	130					

What two measures best describe the center and spread of the data?

24. For each brand of car sold in the United States, data on a customer satisfaction rating (called the APEAL rating) are given (USA Today, July, 2010). The APEAL rating is a score between 0 and 1,000, with higher values indicating greater satisfaction.

822	832	845	802	818	789	748	751	794
792	766	760	805	854	727	761	836	822
820	774	842	769	815	767	763	877	780
764	755	750	745	797	795			

What two measures best describe the center and spread of the data?

Using the Interquartile Range to Identify Outliers

1. Find Q_1 and Q_3
2. Find the interquartile range, IQR $= Q_3 - Q_1$
3. Multiply the IQR by 1.5
4. Subtract (1.5)×(IQR) from Q_1 to find the lower fence.

$$lowerfence = Q_1 - 1.5 \cdot (IQR)$$

   Any data entry less than the lower fence is an outlier.

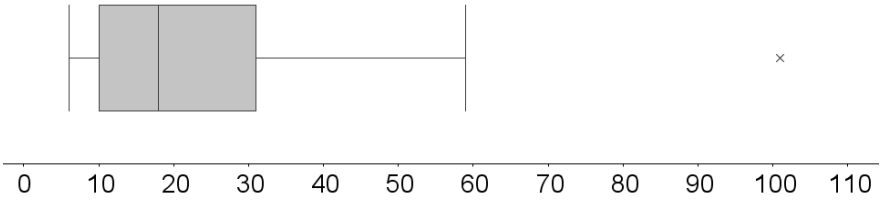
5. Add 1.5×IQR to Q_3 to find the upper fence.

$$upperfence = Q_3 + 1.5 \cdot (IQR)$$

   Any data entry greater than the upper fence is an outlier.

Example When carrying out this list of steps for the power plant data we find:

1. $Q_1 = 10$ and $Q_3 = 31$
2. Find the interquartile range, IQR $= Q_3 - Q_1 = 21$
3. Multiply the IQR by 1.5: $1.5 \times 21 = 31.5$
4. lower fence $= 10 - 31.5 =$ $\boxed{\text{-21.5}}$. There is no data entry less than -21.5 value.
5. upper fence $= 31 + 31.5 =$ $\boxed{62.5}$. So, 104 is an outlier since it is greater than 62.5.



Modified Boxplot The boxplot is redrawn with the whisker on *the right extended out to the largest value in the data set that is not larger than the upper fence*, namely data value 59. Had there been an outlier on the lower end, then our graph would have *the left whisker extended to the lowest value in the data set that was not an outlier.* Also, notice that outliers are marked on the graph with a cross marker.

Your Turn! Use the sample data below answer the questions on the next page. The sample represents the number of paid vacation days used by 20 employees in a recent year.

16 25 1 33 15 5 18 8 20 14 17 19 16 10 21 28 14 37 18 15

1. List the data in ascending fashion.

2. Find the minimum 2. _____

3. Find Q_1 3. _____

4. Find Q_2 4. _____

5. Find Q_3 5. _____

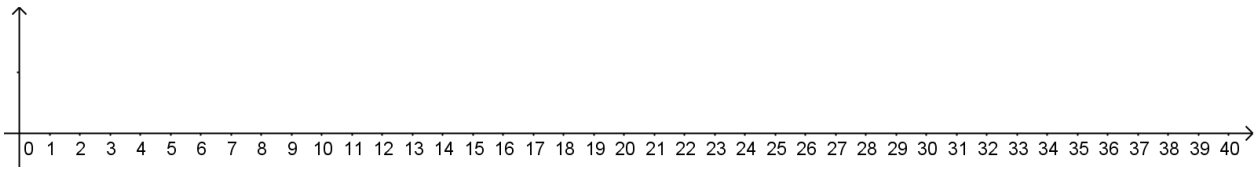
6. Find the maximum 6. _____

7. Find the lower fence. 7. _____

8. Find the upper fence 8. _____

9. What numbers are outliers? 9. _____

10. Sketch the graph of a modified boxplot. Label the x axis.



11. About 75% of the employees in the sample took at least how many days off?

11. _____

12. What percentage of employees took more than 16 and a half days off?

12. _____

13. You randomly select one employee from the sample. What is the likelihood that the person took less than 20.5 days off?

13. _____

VIDEO LINK: https://www.youtube.com/watch?v=_Np8mJGeQ3g

How to use the calculator to graph a Boxplot

or just google 'tim busken how to graph a modified boxplot'

The TI-83, TI-83 Plus, and TI-84 Plus calculators will take a list of data and automatically draw a box-and-whisker plot for that data. Since there are a number of different types of plots available on the calculator, it is important to make sure that all other plots are turned off before you begin or your graph will be cluttered with several unrelated plots being graphed at the same time. Even worse, it is possible for previous plots to become invalid or the data sets that were used before are changed or deleted, causing an error whenever you try to graph anything new. Before you begin any plotting:

- Press the **Y=** key at the top left of the keyboard, and delete or deselect any equations being plotted there.
- Press STAT PLOT (above the **Y=** key) and choose 4 to turn off any other statistical plots.

There are three stages to creating a box-and-whisker plot.

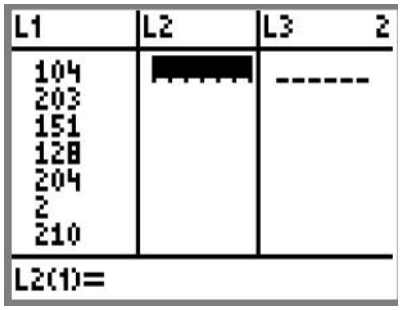
1. Enter the data into a list as before.
2. Tell the calculator what kind of plot you want.
3. Tell the calculator what size to draw the window for the plot.

Example: Doctor Incomes

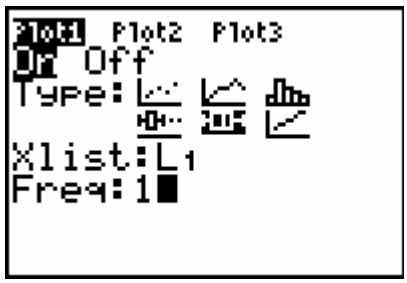
The following numbers represent 2016 incomes in thousands of dollars for seventeen doctors:

104 203 151 128 204 2 210 162 185 169 135 178 350 213 122 135 5

We'll use this data to construct a box-and-whisker plot. First store the data in the list L1.



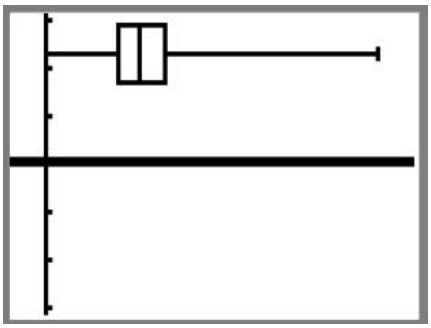
Press the **2nd** key and then press the **Y=** key to get to STATPLOT.
Press the number **1** key.
Move the cursor to On and press the **ENTER** key.
Use the arrow keys and highlight the box-and-whisker plot picture.
Type L1 in for Xlist:



Set the Freq: to 1.

Press the GRAPH key.
Press the **ZOOM** key.
Press the number **9** key.

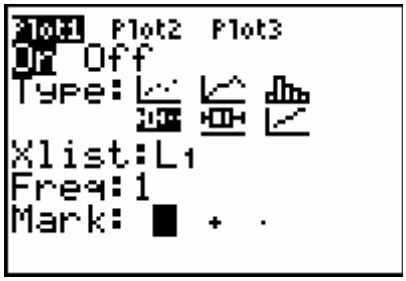
(to select "zoomstat")



We have two different types of box-and-whisker plots to select from. One will separate outliers from the maximum or minimum value (the modified boxplot) and the other will include the outliers in to whiskers (the boxplot).

If you select the picture that shows two dots after the maximum on the plot, the outliers will be shown outside of the whiskers.

This is how you graph the "modified boxplot."



Press the "trace" button, and use your arrow keys to find the five-number summary and the outliers.

min = 2 Q1 = 125 Q2 = 162 Q3 = 203.5 max = 350

outliers = 2, 5, 350