

8.1 Notes: Analyzing Data

People often refer to the mean as the *average*. The mean is only one of the measures considered the average, a measure of the center of a set of data.

Essential Understanding You can describe and compare sets of data using various statistical measures, depending on what characteristics you want to study.

Statistics is the study, analysis, and interpretation of data. One way to analyze data is by finding a *measure of central tendency*. A **measure of central tendency** indicates the “middle” of the data set. The *mean*, *median*, and *mode* are the most common measures of central tendency.

Take note

Key Concepts Measures of Central Tendency

Measure	Definition	Example, using 1, 2, 3, 3, 4, 5, 5, 9
Mean	$\frac{\text{sum of the data values}}{\text{number of data values}}$	$\frac{1 + 2 + 3 + 3 + 4 + 5 + 5 + 9}{8} = 4$
Median	for a data set listed in order: the middle value for an odd number of data values; the mean of the two middle values for an even number of data values	For 1, 2, 3, 3, 4, 5, 5, 9, the middle two values are 3 and 4. The median is their mean $\frac{3+4}{2} = 3.5$.
Mode	the most frequently occurring value(s)	Two modes: In 1, 2, 3, 3, 4, 5, 5, 9, both 3 and 5 occur twice.

A **bimodal** data set has two modes. If a data set has more than two modes, then the modes are probably not statistically useful. If no value occurs more frequently than any other, then there is no mode.



Problem 1 Finding Measures of Central Tendency

Career The frequency table shows the number of job offers received by each student within two months of graduating with a mathematics degree from a small college. What are the mean, median, and mode for the job offers per student?

Job Offers	0	1	2	3	4
Students	2	2	4	5	2

An **outlier** is a value that is substantially different from the rest of the data in a set. If the data is in one variable, outliers can occur at the “ends.” They can be misleading because they can affect measures of central tendency.



Problem 2 Identifying an Outlier

Multiple Choice Which is an outlier for this data set: 56 65 73 59 98 65 59?

(A) 42

(B) 65

(C) 98

(D) 59

The **range of a set of data** is the difference between the greatest and least values. If you order data from least value to greatest value, the median divides the data into two parts. The median of each part divides the data further and you have four parts in all. The values separating the four parts are **quartiles**. The **interquartile range** is the difference between the third and first quartiles.



Problem 3 Comparing Data Sets

Temperature The table shows average monthly water temperatures for four locations on the Gulf of Mexico. How can you compare the 12 water temperatures from St. Petersburg with the 12 water temperatures from Key West?

Gulf of Mexico Eastern Coast Water Temperatures (°F)

Location	J	F	M	A	M	J	J	A	S	O	N	D
St. Petersburg, Florida	62	64	68	74	80	84	86	86	84	78	70	64
Key West, Florida	69	70	75	78	82	85	87	87	86	82	76	72
Dauphin Island, Alabama	51	53	60	70	75	82	84	84	80	72	62	56
Grand Isle, Louisiana	61	61	64	70	77	83	85	85	83	77	70	65

SOURCE: National Oceanographic Data Center

A *box-and-whisker plot* uses minimum and maximum values, the median, and the first and third quartiles to display the spread, or variability, in a data set.

Take note

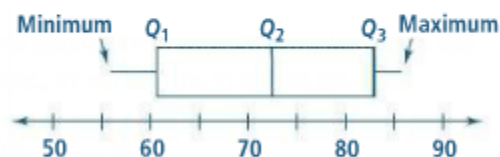
Key Concepts Box-and-Whisker Plot

Definition

A box-and-whisker plot is a way to display data that uses

- quartiles to bound the center box and
- the minimum and maximum values to form the whiskers.

Graph

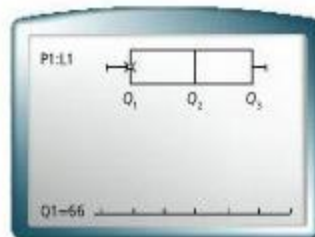
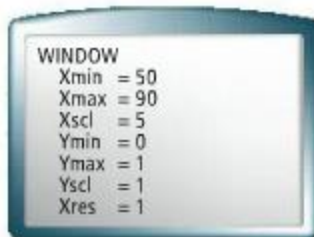




Problem 4 Using a Box-and-Whisker Plot

How can you use a graphing calculator box-and-whisker plot to find quartiles for the water temperature data of St. Petersburg from Problem 3?

- Step 1** For St. Petersburg, use **STAT EDIT** to enter the temperature data in **L1**.
- Step 2** In **STAT PLOT**, select a box-and-whisker plot. Enter **L1** for the St. Petersburg data. Enter the window values. Draw the box-and-whisker plot.



- Step 3** Use **TRACE** to find the quartiles: $Q_1 = 66$, $Q_2 = 76$, and $Q_3 = 84$.



Practice Make a box-and-whisker plot for each set of values

6. 12 11 15 12 19 20 19 14 18 15 16

A **percentile** is a number from 0 to 100 that you can associate with a value x from a data set. It shows the percent of the data that are less than or equal to x . If x is at the 63rd percentile, then 63% of the data are less than or equal to x .



Problem 5 Finding Percentiles

Testing Here is an ordered list of midterm test scores for a Spanish class. What value is at the 65th percentile?

Of the 20 values, 65% fall at or below the value at the 65th percentile.

41	54	61	65	67	73	74
77	77	77	79	80	82	88
89	93	97	98	98	100	

Got It? Testing Here is an ordered list of midterm test scores for a Spanish class

41	54	61	65	67	73	74
77	77	77	79	80	82	88
89	93	97	98	98	100	

- What is the value at the 55th percentile?
- What is the value at the 95th percentile?

Entering Data:

Data is stored in *Lists* on the calculator. Locate and press the **STAT** button on the calculator. Choose **EDIT**. The calculator will display the first three of six lists (columns) for entering data. Simply type your data and press **ENTER**. Use your arrow keys to move between lists.

L1	L2	L3	1
████████	-----	-----	
L1(1)=			

Data can also be entered from the home screen using set notation -- **{15, 22, 32, 31, 52, 41, 11}** → **L1** (where → is the **STO** key)

Data can be entered in a second list based upon the information in a previous list. In the example below, we will double all of our data values in **L1** and store them in **L2**. If you arrow up ONTO **L2**, you can enter a formula for generating **L2**. The formula will appear at the bottom of the screen. Press **ENTER** and the new list is created.

L1	L2	L3	2
15	████████	-----	
22			
32			
31			
52			
41			
11			
L2(1)=			

L1	2	L3	2
15	-----	-----	
22			
32			
31			
52			
41			
11			
L2 = 2*L1			

L1	L2	L3	2
15	30	-----	
22	44		
32	64		
31	62		
52	104		
41	82		
11	22		
L2(1)=30			

Clearing Data:

To clear all data from a list: Press **STAT**. From the **EDIT** menu, move the cursor up **ONTO** the name of the list (**L1**). Press **CLEAR**. Move the cursor down. **NOTE:** The list entries will not disappear until the cursor is moved down. (**Avoid** pressing **DEL** as it will delete the entire column. If this happens, you can reinstate the column by pressing **STAT #5 SetUpEditor**.)

1	L2	L3	1
15	-----	-----	
22			
32			
31			
52			
41			
11			
L1 = {15, 22, 32, 31...			

You may also clear a list by choosing option **#4** under the **EDIT** menu, **ClrList**. **ClrList** will appear on the home screen waiting for you to enter which list to clear. Enter the name of a list by pressing the **2nd** button and the yellow **L1** (above the **1**).

To clear an individual entry: Select the value and press **DEL**.

Sorting Data: (helpful when finding the mode)

Locate and press the **STAT** button. Choose option **#2, SortA(**. Specify the list you wish to sort by pressing the **2nd** button and the yellow **L1** list name. Press **ENTER** and the list will be put in ascending order (lowest to highest). **SortD** will put the list in descending order.

2nd	CALC TESTS
1:	Edit...
2 :	SortA(
3:	SortD(
4:	ClrList
5:	SetUpEditor

Find the Mean and Median:

Method 1: *(fast and easy)*

Press **2nd MODE (QUIT)** to return to the home screen.

Press **2nd STAT (LIST)**. Arrow to the right to **MATH**.

Choose option **#3: mean(** if you want the mean.

Choose option **#4: median(** if you want the median.

Your choice will appear on the home screen waiting for you to tell it which list to use.

Remember the List names appear on the face of the calculator above the number keys 1-6.

```
NAMES OPS MODE
1:min(
2:max(
3:mean(
4:median(
5:sum(
6:prod(
7:stdDev(
```

```
mean(L1)
8.210526316
```

Find the Mode:

 (While there is no specific calculator function to find the mode, the calculator is helpful in ordering the data so that you can find the mode easily.)

Sort the data into ascending or descending order to help find the mode.

STAT, #2 SortA(, and specify **L1**, or the list you are using.

Look at the list (**STAT, #1 EDIT**).

Examine the data to see which value(s) appear(s) most often.

The mode for this data set is 7 and 9. Each of these values appears 3 times in the data set.

```
2ND CALC TESTS
1:Edit
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
```

```
SortA(L1) Done
```

L1	L2	L3	1
7			
7			
7			
9			
9			
9			
L1(12) = 9			