

# Cambridge Christian School

# AP Statistics



2020-2021

Mrs. Cain



1. The purpose of this packet is to create as effective a partnership between student and teacher as possible. To that purpose your summer work will:

I. brief and review you on the information that must be understood from your other mathematics classes in order for you to be successful in this course. To be effective, we must have mastered the concepts through Algebra 2

II. introduce you to some of the statistical capabilities of the graphing calculator. To be effective we must be comfortable with the tools we will use throughout the year.

III. teach you the basic vocabulary and concepts that we will see in the first several units of the class. To be effective, we must speak the same language.

2. The use of a TI-83 or TI-84 calculator is required to complete much of the work here; it is in your best interest to acquire a suitable calculator of your own, both for this summer work, as well as the work during the school year. Instruction will focus on TI-84 calculator techniques, and much of your homework will require its use.

2. **Do not wait until the last minute to complete this packet**, instead, pace yourself throughout the summer and complete this over an extended period of time.

3. **All work should be shown in a neat and orderly fashion on a separate sheet of paper.**

4. Place your answers, where applicable, in the space provided on this handout.

5. Staple your work to the back of this handout.

6. **Turn this packet in on the first full day of class.** Completion is mandatory, and the result will be your first grade for the course. Start with an A!

## Activity 1: Data Types

There are two types of data: *qualitative* (or categorical) and *quantitative*.

*Qualitative variables* or *categorical variables* are variables that categorize individuals (place them in groups). These variables may take on values that are labels for categories. Examples are eye color (blue, hazel, etc.), gender (male or female), method of transportation to school (bike, car, bus, etc.), class rank (senior, junior, etc.). A specific type of qualitative variable is a *binary variable*. A binary variable is a qualitative variable that has only two outcomes. Examples include gender, approve or disapprove of the president's handling of the war in Iraq, outcome of a coin toss, outcome of a die roll (when restricted to a four or not a four), the response to the question "Do you play basketball?"

*Quantitative variables* are numerical variables that represent an amount or quantity. There are two kinds of these: *discrete* and *continuous*.

*Discrete variables* are quantitative variables that assume only a countable number of values. Examples of these include shoe size (... , 6, 6 1/2 , 7, 7 1/2 , ...), score on a test, class size, number of cans collected for MIFA.

*Continuous variables* are quantitative variables that can assume an infinite number of values. In the case of continuous variables, the values can generally assume any decimal quantity within a small range of values (even though we may round the answer like when we measure our height). *These are typically values that result from some kind of measurement.* The units of measurement are pounds/ inches/ Keivin/ degrees/ feet! etc. Examples are height, weight, surface area of oranges, era in baseball (3.23, 2.78, etc.), GPA.

*Just because your variable's values are numbers, don't assume that it's quantitative.* For example 9, 10, 11, and 12 are labels for different class rankings at MUS. Class rank is a qualitative variable (even though it may be answered with a 9, 10, 11, or 12). Social security number is another example of a numerical output that is not a quantitative variable. SSN doesn't stand for any type of numerical quantity (you are riot the 412,327,642 person born in the US!). Phone number is not a quantitative variable either: The 901 area code is a designation for a geographic region; it is not a numerical quantity.



Answer the following questions and then decide if the data is qualitative or quantitative. Then decide if it is also binary, discrete, or continuous.

<b>Question</b>	<b>Answer</b>	<b>Type</b>
1. In which year did you take Algebra I?	_____	_____
2. How many CD's do you own?	_____	_____
3. What is your zip code?	_____	_____
4. Choose a random integer from 1 to 20.	_____	_____
5. How many siblings do you have?	_____	_____
6. Do you like chocolate?	_____	_____
7. Who is your favorite math teacher?	_____	_____
8. What gender are you?	_____	_____
9. How tall are you (in inches)?	_____	_____
10. Where did you eat your last meal? (1:=home, 2=restaurant, 3=other)	_____	_____
11. How many AP classes will you be taking this year?	_____	_____
12. How far away from school do you live?	_____	_____
13. How many miles per gallon does your vehicle get while driving in the city?	_____	_____
14. What grade did you earn in your last math class: A, B, C, <b>D</b> , or F?	_____	_____
15. What time is it?	_____	_____
16. How fast can you run the 40'?	_____	_____

## ACTIVITY 2: DATA AND LISTS

Qualitative data can be stored on the TI-83 in **lists**. There are several ways to create a list. From the home screen braces can be used to define a data set, which then can be stored in one of the list names L1 through L6 (Figure 1.1). Alternately, use STAT 1:Edit to go to the list editor and enter the data into columns (Figure 1.2)

Figure 1.1  
`<1 6;-7;-8 9)-+L;::I`

Figure 1.2

L1	L2:	L3
-----	1 7 8 9	-----

L2:(2:) =6

In either case, new lists can be created from existing lists, such as L2+5 (Figure 1.3). Make sure when you enter the new list that you are on the L3 icon and not within the list of numbers.

Figure 1.3

L1	U:	
.....	7 8 9	

L3 = L2 + 5

Figure 1.4

L1	L2	
-----	1 6 11 12 13	1 6 11 12 13

L3 = {6, 11, 12, 13, ...}

On the TI-83, lists may also be given their own names and will be retained in memory until deleted. This is particularly useful for data that will be used repeatedly.

### Example 1:

Create a named list for the following set of running speeds in mph for various animals:

Cheetah	70	Warthog	30
Lion	50	Cat	30
Coyote	43	Man	27.89
Hyena	40	Pig	11
Greyhound	39.35	Tortoise	0.17
Rabbit	35	Snail	0.03

**Solution:**

To create a named list go to the list editor and move to the right past L6. A "Name" prompt will appear and the list name can be typed (figure 1.6). The values can be entered in the usual way (Figure 1.7)

Figure 1.6

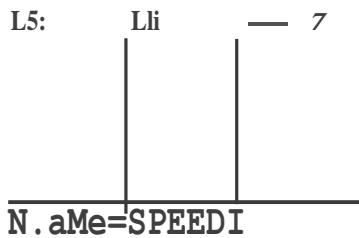
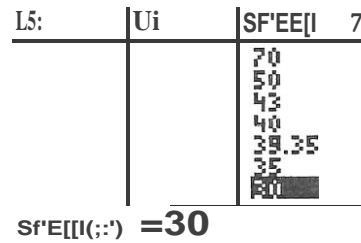


Figure 1.7



**Example2:**

Create a new list showing the speeds in feet per second.

**Solution:**

New lists can be created from named lists on either the home screen or in the list editor. On the home screen, the speeds in mph can all be converted to ft/sec and stored in a list named FTSEC by a single command (Figure 1.10). In order for the TI-83 to distinguish a user defined list name from other symbols it is necessary to preface a list name with a special character L that is located in the OPS sub-menu. The L character may also be found in the CATALOG.

Figure 1.10

LSPEED\*5280..<3600  
7LFTSECI

To organize named lists in the list editor use STAT5:SetUpEditor followed by the names in the order they are to appear (Figure 1.11). The lists will appear in columns as requested (Figure 1.12)

Figure 1.11

SetUpEditor LSPE  
ED LFTSECI

Figure 1.12

S::FEED	F::TS::E::C
70	102.67
50	72.333
43	63.067
40	58.667
39.35	57.713
35	51.333
30	44

SPEED(1)=70

Exercises-

17. Create a list L1 using {4, 7, 9, 11, 14, 17, 20}.

Create new lists

- A. L1-7: \_\_\_\_\_
- B. 2\* L1: \_\_\_\_\_
- C.  $L_1^2$ : \_\_\_\_\_
- D. L1: \_\_\_\_\_
- E. ln Lr \_\_\_\_\_

18. Create a list PLANM showing the mean distance from the sun in millions of miles for each planet. Create a new list PLANK showing the mean distance in millions of kilometers . (use the internet to find the miles:kilometers conversion)

PLANET	<u>MEAN DISTANCE IN MILLIONS OF MILES</u>	<u>MEAN DISTANCE IN MILLIONS OF KM</u>
Mercury	36.0	
Venus	67.24	
Earth	92.9	
Mars	141.71	
Jupiter	483.88	
Saturn	887.14	
Uranus	1783.98	
Neptune	2796.46	
Pluto	3660.0	

- A. Explain the procedure used to enter the lists. \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_



## ACTIVITY 3: Numerical Descriptions of Quantitative Data

There are two categories of numbers that are used to describe a set of data: measures of center and measures of spread.

Measures of Center:

1. The *mean* is the average number. It is the sum of all the data values divided by the number (n) of values.

$$\text{Ex. } \{4, 36, 10, 22, 9\} \quad \text{mean} = X = \frac{\sum i}{n} = \frac{4+36+10+22+9}{5} = 16.2$$

2. The *median* is the value that separates the bottom 50% of data from the top 50% of data. It is the middle element of an ordered set of data that is odd in number. It is the average of the two middle elements of an ordered set of data that is even in number.

$$\text{Ex. } \{4, 36, 10, 22, 9\} = \{4, 9, 10, 22, 36\} \quad \text{median} = 10$$

$$\{4, 36, 10, 22, 9, 43\} = \{4, 9, 10, 22, 36, 43\} \quad \text{median} = \frac{10+22}{2} = 16$$

3. The *mode* is the value that occurs most often in a set of data. If the data occurs with the same frequency, then there is no mode. If two (or more) values occur the most then they are both the mode.

Measures of Spread:

1. The *range* is a measure of the spread of the entire data. It is calculated by subtracting the minimum value from the maximum value.

$$\text{Ex. } \{4, 36, 10, 22, 9\} = \{4, 9, 10, 22, 36\} \quad \text{range} = 36 - 4 = 32$$

2. The *interquartile range* (IQR) is a measure of the spread of the middle 50% of the data. It is calculated by subtracting the 25th percentile (Q1) from the 75th percentile (Q3). Q1 is the median of the lower half of the data. It separates the bottom 25% of values from the top 75% of values. Q3 is the median of the upper half of the data. It separates the top 75% of values from the bottom 25% of values. In neither of these cases is the median considered in the top half or the bottom half of the data.

$$\text{Ex. } \{4, 36, 10, 22, 9\} = \{4, 9, 10, 22, 36\} \quad \text{IQR} = 29 - 6.5 = 22.5$$

$$\begin{array}{cc} \mathbf{i} & \mathbf{t} \\ \text{Q1}=6.5 & \text{Q3}=29 \end{array}$$

$$\text{Ex. } \{4, 36, 10, 22, 9, 43\} = \{4, 9, 10, 22, 36, 43\} \quad \text{IQR} = 36 - 9 = 27$$

$$\begin{array}{cc} \mathbf{t} & \mathbf{t} \\ \text{Q1}=9 & \text{Q3}=36 \end{array}$$

3. The *standard deviation* is the measure of the spread around the mean. It is

calculated using the following formula:  $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

$$\text{Ex. } \{4, 36, 10, 22, 9, 43\} \quad s = \sqrt{\frac{(4-16.2)^2 + (26-16.2)^2 + \dots + (9-16.2)^2}{5-1}} = 12.89$$

This means that the average number differs from the mean by about 12.89 units. The smaller the standard deviation the closer the data should be clustered around the mean.

To see statistical results including the quartiles and standard deviation, use STAT CALC 1:1-Var Stats (Figure 3.1), Followed by the list name (Figure 3.2)

Figure 3.1

```
EDIT M:1.D TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4=LinRe-9(ax+b)
5:Qu.adRe9
6:Cubic.Re9
7.J.QuartRe9
```

Figure 3.2

```
1-1Var Stat:;. L1
```

Figure 3.3

```
1-1Var Stat:;.
X:;.
x=81
x2:=1977
sx=12.8918579
CRX- 1.1,1,0,8,4
J,n=5
```

If you push the down arrow key, then you can see the rest of the statistics (Figure 3.4).

Figure 3.4

```
1-Var Stats
tn=5
MinX=4
Q1=6.5
Med=10
Q3:=29
MaxX=36
```

Exercises

Here is a list of parents' ages at the time their sons were born

Dad: 41 27 23 31 30 33 26 32 43 25 34  
27 25 34 27 26 28 32 32 35 27 33  
34 34 34 35

Mom: 39 26 23 30 28 33 23 32 38 23 35  
24 24 33 24 23 24 32 23 30 24 29  
34 35 26 31

Enter these as two lists in your calculator and use the 1-Var Stat option to calculate the following:

19. Find the mean and median for the Dad data:

Mean \_\_\_\_\_ Median \_\_\_\_\_ Which is larger? \_\_\_\_\_

20. Find the mean and median for the Mom data:

Mean \_\_\_\_\_ Median \_\_\_\_\_ Which is larger? \_\_\_\_\_

21. Now compare the two means you calculated. Which is larger? \_\_\_\_\_

Is this what you expected? \_\_\_\_\_. Explain why or why not.

---

---

22. Calculate the standard deviations for both sets of data: Dad \_\_\_\_\_ Mom \_\_\_\_\_

Why might these values be different? Explain.

---

---

23. Find Q1 and Q3 and the IQR for the Dad data. Q1 \_\_\_\_\_ Q3 \_\_\_\_\_ IQR \_\_\_\_\_

Find Q1 and Q3 and the IQR for the Mom data. Q1 \_\_\_\_\_ Q3 \_\_\_\_\_ IQR \_\_\_\_\_

24. A company has two machines that fills cans of soft drinks. Samples from each machine show the following number of ounces per can:

Machine A: 11.1, 12.0, 11.4, 12.1, 11.7, 11.5, 12.2, 11.4, 11.3, 11.9

Machine B: 10.9, 12.4, 12.7, 11.8, 12.3, 11.9, 12.0, 12.5, 12.7, 11.6

Find the mean and standard deviation for both machines.

$\bar{A}X =$  \_\_\_\_\_  $\bar{A}S =$  \_\_\_\_\_  $\bar{B}X =$  \_\_\_\_\_  $\bar{B}S =$  \_\_\_\_\_

25. Using your answer to #24, explain which machine is "better" at filling soft drink cans.

---

---

There are two basic graphical techniques used for qualitative data as described below:

1. *bar chart* - one axis is the values of your variable (the order of their listing doesn't matter) the other axis is the frequency or relative frequency. *The bars do not touch.*

2. *pie chart* - the whole group of values is shown as a circle. Each "piece of the pie" corresponds to the relative frequency of the values. To determine the central angle that forms each "piece" use the following:  $central\ angle = relative\ frequency \times 360$

You should have extensive experience with these techniques so we won't cover them in this packet.

## ACTIVITY 4: Graphical Displays of Univariate (one variable) Data

A set of quantitative data is often separated into groups or intervals. For example, test scores may be separated into 90-99, 80-89, and so on. The frequency refers to the count for each interval. A histogram is a visual representation of the frequencies plotted against the interval.

### Example 1:

Enter the following test scores into a list named SCORE and create a histogram using intervals of width 10: 100, 98, 95, 92, 88, 85, 84, 84, 81, 80, 77, 75, 72, 70, 70, 66, 65, 63, 60, 59, 55, 50

### Solution:

First enter the scores into a list using one of the methods learned in the previous activity. (Figure 4.1) To create a histogram, go to STAT PLOT, the 2<sup>nd</sup> function of the Y= button. Then choose one of the Plots by pressing 1, 2, or 3. Now activate the plot by selecting ON in the second row. At the TYPE prompt choose the third icon for histogram. At the X LIST prompt, we want the name of the list (Figure 4.2). The name can be typed, but it is easier to use the LIST NAMES menu and type the number that corresponds to the desired list, in this case SCORE.

Figure 4.1

LS:	L1	SCORE	7
		1M	
		10(1	
		9B	
		81	
		BB	
		65	

:>CIIRE(S:) =92

Figure 4.2

```

Plot1 1 t 1 t.
Off-
Type: !...
lit. !D-1 L:....
XI ist.. SCOREI
Fre-.:t: 1
    
```

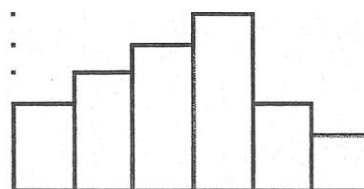
Set an appropriate WINDOW (Figure 4.3) X scl determines the width of the interval, thus the number of bars in the histogram. Then GRAPH (Figure 4.4)

Figure 4.3

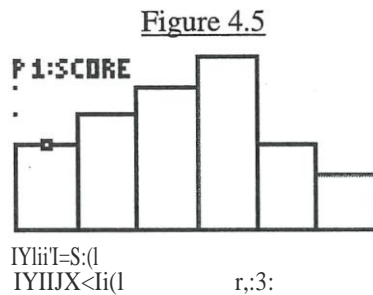
```

WINDOW
XMin=50
XMax=110
Xscl=10
YMin=-2
YMax=6
Yscl=1
Xres=1
    
```

Figure 4.4



You may now TRACE the graph to see the intervals and their frequencies (Figure 4.5).



Example 2:

Create six intervals of width 10 for the test scores and display a histogram.

Solution

Enter the midpoint of each interval in one list and the frequencies in a second list. (Figure 4.6). In this instance will be used for the frequencies and should be indicated in the set up for the plot. (Figure 4.7)

Figure 4.6

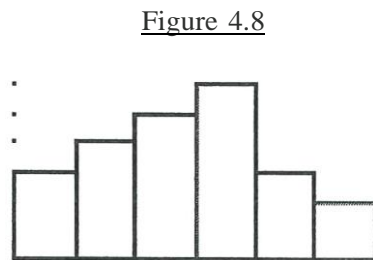
L1	L2	L3	...
1 (15	...	-----	...
9S:	3		
BS:	li		
75	5		
liS	...		
SS:	3		
-----	A		

**L2(7) =**

Figure 4.7

f)l)t,;! F'le:t.3  
f.:p fL !:....  
tOt.. 10)t l.....  
Xlist:L1  
FreqCa:L;:fll

It is not necessary, but practical to set the Xscl = 10, to match the intervals of the grouped data. If so, the histogram will be the same as the previous one (Figure 4.8)



A *box plot* is a graphical display for a set of quantitative data that only displays the *five number summary*: minimum      Q1      median      Q3      maximum

**Example 3:**

Create a box plot of the test scores.

**Solution:**

To create this plot use STAT PLOT, and select a plot 1, 2, or 3. For Type select the fourth icon. Use the LIST NAMES menu to enter list name SCORE (Figure 4.9). Set an appropriate WINDOW. The Y min and Y max values are not important; however, it is convenient to move the x-axis near bottom of the screen or off the screen so that it doesn't pass through the box plot. (Figure 4.10)

Figure 4.9

```
Plot1 PLOT 1 )
Off
Type:1|... 1... d|h|
Q:1|/
Xlist: SCORE
Freq: 1
```

Figure 4.10

```
WINDOW
XMin=10
XMax=110
Xscl=10
Vrmin=-2
VMax=6
Yscl=1
Xres=1
```

TRACE to read the extremes and quartiles (Figure 4.11).



Figure 4.11

Box plots are particularly effective for comparison of two or more sets of data.

**Example 4:**

Suppose another class receives these test scores: 94, 93, 90, 83, 82, 82, 77, 76, 75, 70, 64, 60, 53, 52, 48, 44, and 40. Create two box and whiskers plots to compare their performance with the original group.

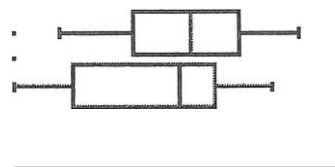
**Solution:**

Enter the new scores into another list, such as L1. Each class is set up in STAT PLOT (Figure 4.12) to obtain two box plots (Figure 4.13).

Figure 4.12

```
Plot1 PLOT 1 )
Off
Type:1|... 1... d|h|
Q:1|/
Xlist: SCORE
Freq: 1
Plot2...Off
ID-4 L1 1
3: Plot3...Off
k L1 L2: a
4.J..PlotsOff
```

Figure 4.13



It appears as though the first set of scores has a higher median and a higher maximum score. It also appears that the two sets of scores have approximately the same range (about 50) but the second set has a larger IQR (30 compared to 23).

In AP Statistics a "modified box plot" is used. This shows any *outliers*. An outlier is defined to be a data point that does not fit the pattern of the rest of the data. When your calculator graphs a modified box plot an algorithm is used to determine what it takes to "not fit the pattern of the rest of the data". This algorithm is  $1.5 \cdot \text{IQR}$  away from the box part of the graph (above or below). These outliers are shown with dots or stars or any other small symbol.

**ExampleS:**

Suppose one student from the original class takes the test late and scores 22. Create a box plot for the original class.

**Solution:**

Add a score of 22 to the SCORE list. For Type of graph select the fourth icon. Change the X min to 14 then graph.

Figure 4.14

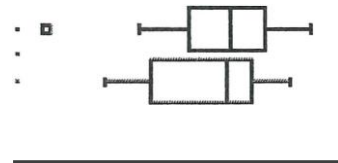
LS:	l.:i	S:(DF;E 7
		63:
		li(
		S9
		SS:
		o
		iJJIII

SCDFIE(2:S)=

Figure 4.15

PLYt P1 J  
 Off  
 e:k::: Lc:: Jib.  
 Otk:::,  
 Xlist: SCORE  
 Fre-Qj: 1  
 M.ar·k: lii · ·

Figure 4.16





Exercises-

26. Go back to the WINDOW in Example 1 of the first histogram of SCORE.

A. Change the Xscl to 5 then GRAPH. Describe how this histogram differs from the first histogram. Which histogram would best describe the data?

---

---

B. Then go back and change the Xscl to 15 then GRAPH. Describe how this histogram differs from the first two. Again state which histogram would best describe the data.

---

---

C. From these two problems what conclusions can you draw about how the Xscl affects the histogram?

---

---

27. Create a histogram of the list PLANM (from Activity 1), which shows the average distance from the sun for the planets of our solar system.

A. Describe how you graphed the data in the calculator.

---

---

B. What would be the best setting for the Xscl with this data and why?

---

---

28. A company has two machines that fills cans of soft drinks. Samples from each machine show the following number of ounces per can:

Machine A: 11.1, 12.0, 11.4, 12.1, 11.7, 11.5, 12.2, 11.4, 11.3, 11.9

Machine B: 10.9, 12.4, 12.7, 11.8, 12.3, 11.9, 12.0, 12.5, 12.7, 11.6

Create a box plot for each machine. Sketch one above the other.

Find the range for: Machine A: \_\_\_\_\_ Machine B: \_\_\_\_\_

Find the median for: Machine A: \_\_\_\_\_ Machine B: \_\_\_\_\_

Which machine is more "consistent"? Why? \_\_\_\_\_

\_\_\_\_\_

Say that Machine A gives one more sample with a fill of 10 oz. Is this an outlier? \_\_\_\_\_

Say that Machine B gives one more sample with a fill of 10 oz. Is this an outlier? \_\_\_\_\_

29. Create a boxplot for the running speeds example first used in Example 1 of Activity #2.

Are there any outliers? \_\_\_\_\_ If so, name them. \_\_\_\_\_

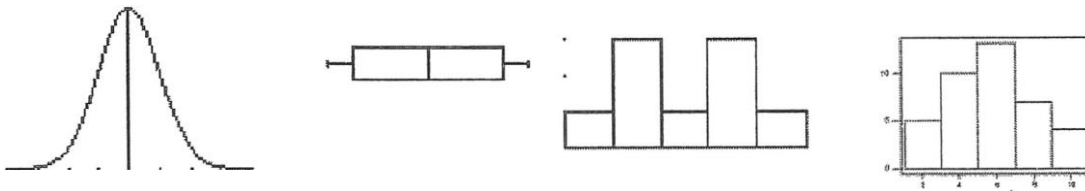
## ACTIVITY 5: Assessing the Shape of a Graph

When describing a set of data we look at the following features:

Shape          center          spread          clusters and gaps          outliers

We have several terms that we use to describe the shape but this packet will concentrate on only two: *symmetric* and *clustered*.

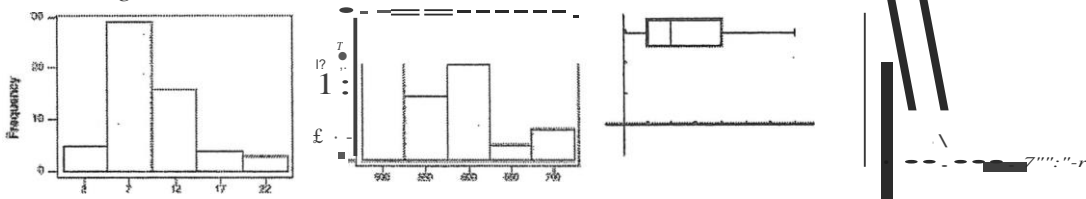
One can tell if a graph is *symmetric* if a vertical line in the "center" divides the graph into two fairly congruent shapes. The following sets of data can be described as symmetric



*The mean and the median are approximately the same in a symmetric set of data.*

One can tell if a graph is *skewed* if the graph has a big clump of data on either the left (skewed right) or the right (skewed left) with a tendency to get flatter and flatter as the values of the data increase (skewed right) or decrease (skewed left). A common misconception is that the "skewness" occurs at the big clump. The following sets of data can be described as skewed:

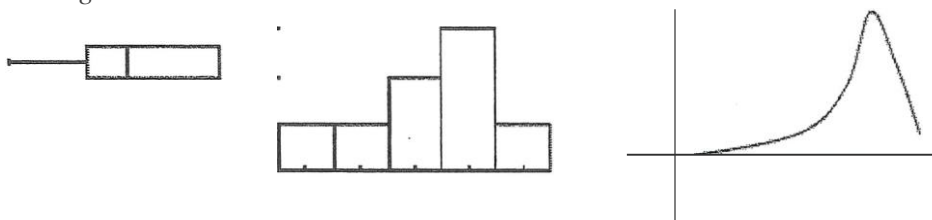
*skewed right:*



*The mean is larger than the median in a skewed right set of data. The mean is always further along the "tail".*

*skewed left:*

*The mean is smaller than the median in a skewed right set of data. The mean is always further along the "tail".*

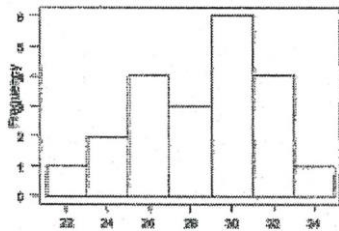


**Exercises-**

30. For the following graphs, find the shape, center (**just do the median**), and spread (**find only the range**). If there any other notable features evident in the graph (clusters, gaps, or outliers), then say where they are. Otherwise do not comment on clusters, gaps or outliers.

(Note: To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. Add these up and divide by two. This tells you where the median is located. Which bar is this value in? That's the median. For graph A,  $n = 21$ , so the middle value is 10.5. Starting with the first bar count  $1 + 2 + 4 + 3 + 6 \dots$  So the median is in the bar that contains the 10.5 value (bigger than 10 anyway). That's 30. So, the median is 30.

A.



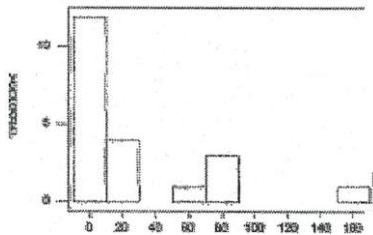
Shape: \_\_\_\_\_

Center: \_\_\_\_\_

Spread: \_\_\_\_\_

Clusters? \_\_\_\_\_

B.



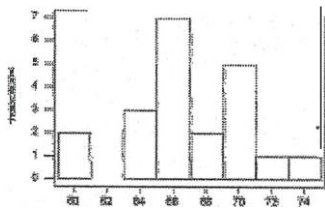
Shape: \_\_\_\_\_

Center: \_\_\_\_\_

Spread: \_\_\_\_\_

Clusters? \_\_\_\_\_

C.



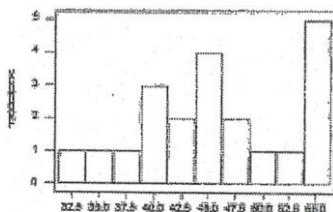
Shape: \_\_\_\_\_

Center: \_\_\_\_\_

Spread: \_\_\_\_\_

Clusters? \_\_\_\_\_

D.



Shape: \_\_\_\_\_

Center: \_\_\_\_\_

Spread: \_\_\_\_\_

Cluster: \_\_\_\_\_

