

Welcome to AP Statistics! I am looking forward to having you in class in the fall. In order to be ready to hit the ground running in August, you have some preliminary work (Chapters 1-3 in the textbook). The summer packet will be collected the first day of class (**8/28/13**) and graded (as a test grade) for both completion and accuracy. Please take this assignment seriously, as we will use the material presented in the first three chapters throughout the school year.

AP Statistics is as much a writing course as it is a math course. Explaining in complete thoughts (sentences) is required on this assignment and throughout the course. Often, questions will require you to comment on what your graph tells you (so again, write clearly and in complete sentences when applicable).

AP Statistics is not an easy class. No AP class is easy, of course—so expect to spend lots of time reading and studying outside of class, as well as in class. But AP Statistics is special, in that it is a course that combines mathematics, critical thinking and language. There is reading and writing involved. In fact, if you think math is all formulas, equations, and calculations, you'll soon find that this isn't really a math course. It's a course in reading, analyzing, thinking, and writing clearly. On the AP test itself, you'll be asked to write descriptive paragraphs and concluding sentences. You'll have to explain the reasoning behind the method you use and the conclusions you find.

You will be expected to work individually or in small groups to plan and perform data collection and analyses. You have ample opportunity to think through problems, collect and analyze data, make conjectures and decisions, reach conclusions, and share outcomes with other students as well as the teacher using statistical software, MS Excel, PowerPoint, etc.

Please work through the problems in this packet, showing all appropriate work on a separate sheet of paper. All work and answers not on a separate sheet of paper will not be graded and result in a zero. Please keep your work labeled and organized, in numerical order. Work must be shown to support each answer and should be done legibly and neatly. If you are unfamiliar with a term or type of problem, refer back to your textbook or go online to various help sites for Statistics (see below). The skills covered are part of the foundation for your course. Mastery of certain mathematical skills is assumed. Please read each set of directions carefully.

- 1) The purpose of this packet is threefold:
 - I. To brief you on the information that must be understood from your other mathematics classes in order for you to be successful in this course.
 - II. To introduce you to some of the statistical capabilities of the TI-83+/84+ calculator.
 - III. To teach you some of the basic concepts that we will see in the first five chapters of the textbook.
- 2) **Do not wait until the last minute to complete this packet**, instead, you must pace yourself throughout the summer and complete this over an extended period of time.
- 3) Place your answers, where applicable, in the space provided on this handout.
- 4) Staple your work to the back of this handout.
- 5) This packet is worth **100 points!** This will count as a **test grade!**

Helpful websites:

<http://www.stattrek.com/>

<http://statistics-help-for-students.com/>

<http://www.freemathhelp.com/statistics.html>

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.).

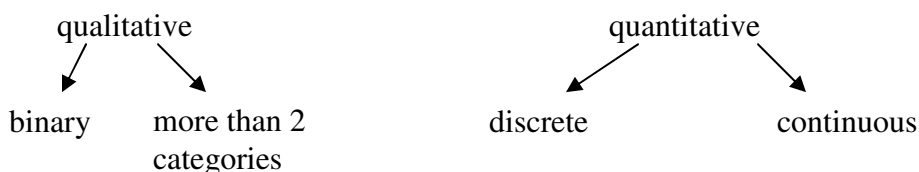
A specific type of qualitative variable is a **binary variable**. A **binary variable** is a qualitative variable that has only two outcomes. Examples include gender, approve or disapprove of the president's handling of the war in Iraq, outcome of a coin toss, outcome of a die roll (when restricted to 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 ½ , 7, 7 ½ , ...), 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/ Kelvin/ 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 not 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**.

Question	Answer	Type
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, D, 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 L_1 through L_6 (Figure 1.1). Alternately, use STAT 1:Edit to go to the list editor and enter the data into columns (Figure 1.2)

$\{1, 6, 7, 8, 9\} \rightarrow L_2$

Figure 1.1

L1	L2	L3	2
-----	1 6 7 8 9 -----	-----	-----
L2(2) = 6			

Figure 1.2

In either case, new lists can be created from existing lists, such as L_2+5 (Figure 1.3). Make sure when you enter the new list that you are on the L_3 icon and not within the list of numbers.

L1	L2	L3	3
-----	1 6 7 8 9 -----	-----	-----
L3 = L2 + 5			

Figure 1.3

L1	L2	L3	3
-----	1 6 7 8 9 -----	6 11 12 13 14 -----	-----
L3 = {6, 11, 12, 13, ...}			

Figure 1.4

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

Source: 1996 Information Please Almanac.

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)

L5	L6	?
-----	-----	

Name=SPEED

Figure 1.6

L5	L6	SPEED ?
-----	-----	70
		50
		43
		40
		39.35
		35
		30

SPEED(?) = 30

Figure 1.7

Example 2:

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.

```
LSPEED*5280/3600
→LFTSEC
```

Figure 1.10

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)

```
SetUpEditor LSPE
ED, LFTSEC
```

Figure 1.11

SPEED	FTSEC	----- 1
70	102.67	
50	73.333	
43	63.067	
40	58.667	
39.35	57.713	
35	51.333	
30	44	

SPEED(1)=70

Figure 1.12

Exercises –

17. Create a list L_1 using {4, 7, 9, 11, 14, 17, 20}.

Create new lists

A. $L_1 - 7$: _____

B. $2 * L_1$: _____

C. L_1^2 : _____

D. $\sqrt{L_1}$: _____

E. $\ln L_1$: _____

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.

PLANET	MEAN DISTANCE IN MILLIONS OF MILES
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. _____

B. PLANK: _____

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

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

Ex. {4, 36, 10, 22, 9, 43}

$$s = \sqrt{\frac{(4-16.2)^2 + (26-16.2)^2 + \dots + (9-16.2)^2}{5-1}} = 12.89$$

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)

```

EDIT [2nd][DEL] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
    
```

Figure 3.1

```

1-Var Stats L1
    
```

Figure 3.2

```

1-Var Stats
x̄=16.2
Σx=81
Σx²=1977
Sx=12.8918579
σx=11.53082824
↓n=5
    
```

Figure 3.3

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

```

1-Var Stats
↑n=5
minX=4
Q1=6.5
Med=10
Q3=29
maxX=36
    
```

Figure 3.4

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 Q_1 and Q_3 and the IQR for the Dad data. Q_1 _____ Q_3 _____ IQR _____

23. Find Q_1 and Q_3 and the IQR for the Mom data. Q_1 _____ Q_3 _____ 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.

$\bar{x}_A =$ _____ $s_A =$ _____ $\bar{x}_B =$ _____ $s_B =$ _____

25. Using you 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 · 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, 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 2nd 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.

L5	L6	SCORE 7
-----	-----	100
		100
		98
		95
		92
		88
		85
		84
		84
		81
		80
		77
		75
		72
		70
		70
		66
		65
		63
		60
		59
		55
		50

SCORE(5) =92

Figure 4.1

```

PLOT1 PLOT2 PLOT3
Off Off
Type: [Bar] [Line] [Pie]
      [Box] [Dot] [Other]
Xlist: SCORE
Freq: 1
    
```

Figure 4.2

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

```

WINDOW
Xmin=50
Xmax=110
Xscl=10
Ymin=-2
Ymax=6
Yscl=1
Xres=█
    
```

Figure 4.3

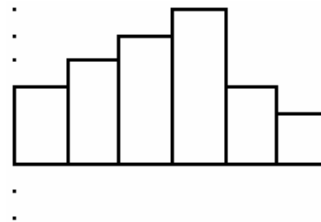


Figure 4.4

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

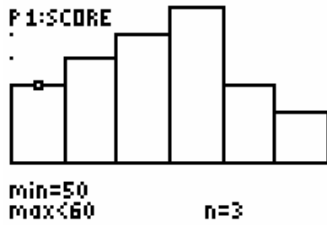


Figure 4.5

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 L₂ will be used for the frequencies and should be indicated in the set up for the plot. (Figure 4.7)

L1	L2	L3	Z
105	2		
95	2		
85	3		
75	4		
65	3		
55	2		
5	1		
L2(?) =			

Figure 2.6

```

2nd [2] Plot2 Plot3
On Off
Type: L1 [ ] [ ] [ ]
Xlist: L1
Freq: L2 [ ]
    
```

Figure 2.7

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)

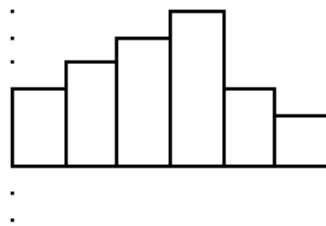


Figure 4.8

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)

```

PLOT1 Plot2 Plot3
Off Off
Type: [1] [2] [3] [4]
Xlist: SCORE
Freq: 1

```

Figure 4.9

```

WINDOW
Xmin=00
Xmax=110
Xscl=10
Ymin=-2
Ymax=6
Yscl=1
Xres=1

```

Figure 4.10

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

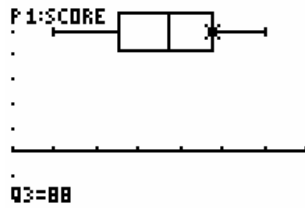


Figure 4.11

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).

```

STAT PLOTS
1:Plot1...On
   [1] SCORE 1
2:Plot2...On
   [1] L1 1
3:Plot3...Off
   [1] L1 L2
4:PlotsOff

```

Figure 4.12

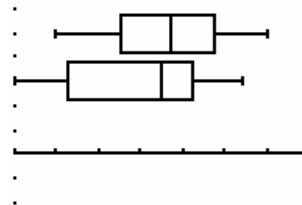


Figure 4.13

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·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.

Example 5:

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.

L5	L6	SCORE 7
		63
		60
		59
		55
		50
		22

SCORE(25) =

Figure 4.14

```

2nd 2nd Plot2 Plot3
Off Off
Type: L1 L2 L3 L4 L5 L6 L7 L8 L9
Xlist: SCORE
Freq: 1
Mark: [ ] + .
    
```

Figure 4.15

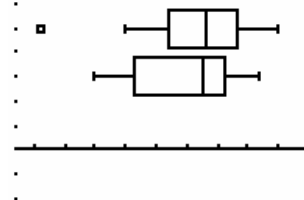


Figure 4.16

Exercises-

4.1 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?

4.2 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?

4.3 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? _____

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

Are there any outliers? _____ If so, name them. _____

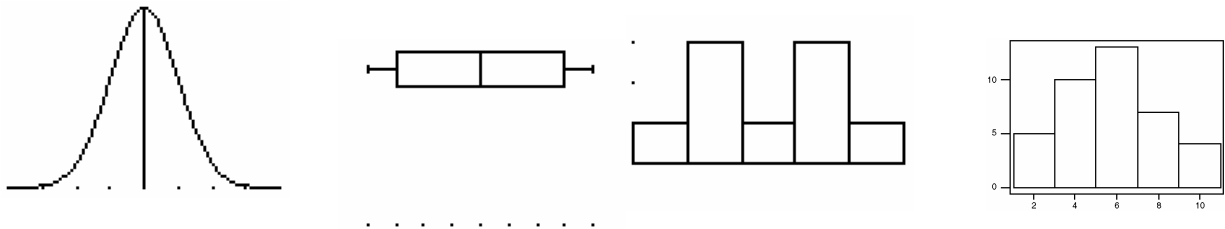
ACTIVITY 5: Assessing the Shape of a Graph

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*.

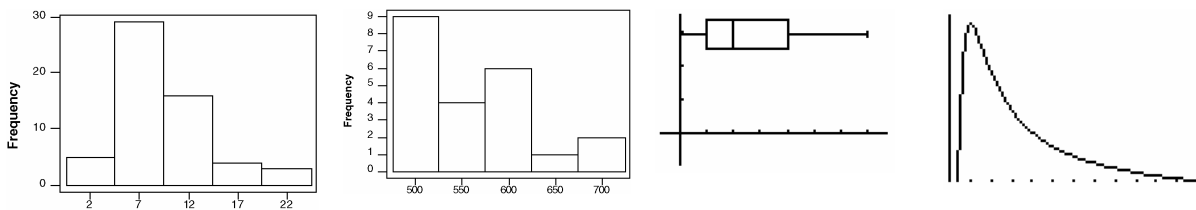
One can tell if a graph is *symmetric* if a vertical line in the “center” divides the graph into two fairly congruent shapes. The following sets of data can be described as symmetric



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

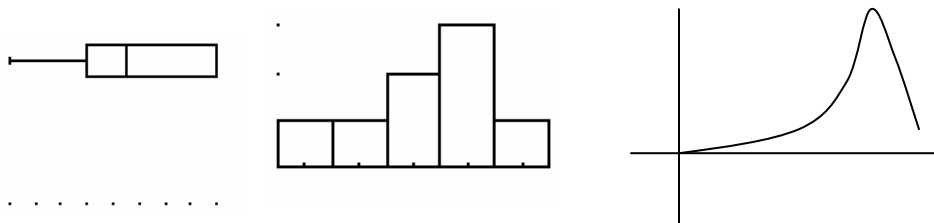
One can tell if a graph is *skewed* if the graph has a big clump of data on either the left (skewed right) or the right (skewed left) with a tendency to get flatter and flatter as the values of the data increase (skewed right) or decrease (skewed left). A common misconception is that the “skewness” occurs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



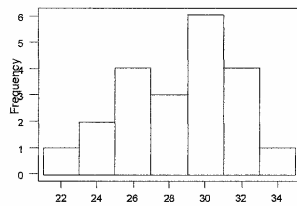
The mean is smaller than the median in a skewed right set of data. The mean is always further along the “tail”.

Exercises-

5.1 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 $1 + 2 + 4 + 3 + 6...$ 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.



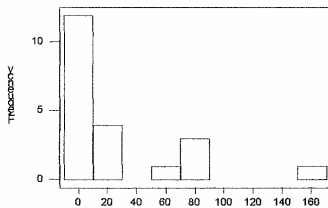
Shape _____

Center _____

Spread _____

Clusters? _____

B.



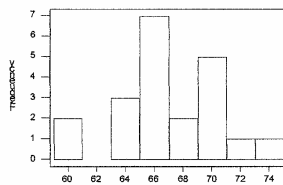
Shape _____

Center _____

Spread _____

Clusters? _____

C.



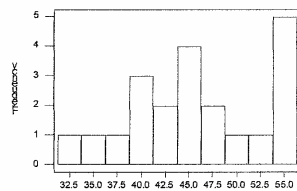
Shape _____

Center _____

Spread _____

Clusters? _____

D.



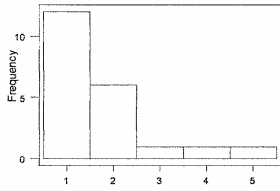
Shape _____

Center _____

Spread _____

Clusters? _____

E.



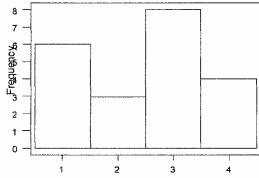
Shape _____

Center _____

Spread _____

Clusters? _____

F.



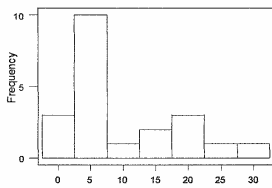
Shape _____

Center _____

Spread _____

Clusters? _____

G.



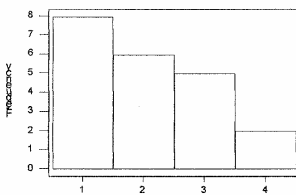
Shape _____

Center _____

Spread _____

Clusters? _____

H.



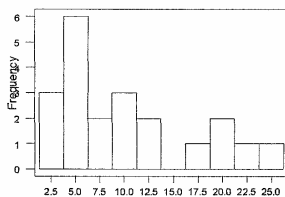
Shape _____

Center _____

Spread _____

Clusters? _____

I.



Shape _____

Center _____

Spread _____

Clusters? _____

5.2 Use the following list of variables to identify which of the graphs in Question 5.1 *could* be a graphical display of the answers for a typical class of students. (Look at things like scale, range, min, max). Write the letter of the correct graph in the blank provided. There are more variables than there are graphs, so don't worry if you have extras.

Variables

Grade when a student takes Algebra _____

Average # of CD's you own _____

Age of your father when you were born _____

Age of your mother when you were born _____

Age of your stat teacher this next year _____ (guess, even if you don't know me!!)

of siblings you have _____

of cousins you have _____

Your height (in inches) _____

of AP classes you will be taking this next year _____

How long you have lived in this area _____

How far away from school you live (in miles) _____

Amount of change in your pocket on the first day of school _____