

AP Statistics Prerequisite Packet

Types of Data

Quantitative (or measurement) Data

These are data that take on numerical values that actually represent a measurement such as size, weight, how many, how long, score on a test, etc. For these data, it makes sense to find things like “average” or “range” (largest value – smallest value). For instance, it doesn’t make sense to find the mean shirt color because shirt color is not an example of a quantitative variable. Some quantitative variables take on **discrete** values, such as shoe size (6, 6 ½, 7, ...) or the number of soup cans collected by a school. Other quantitative variables take on **continuous** values, such as your height (60 inches, 72.99999923 inches, 64.039 inches, etc.) or how much water it takes to fill up your bathtub (73.296 gallons or 185. 4 gallons or 99 gallons, etc.).

Categorical (or qualitative) Data

These are data that take on values that describe some characteristic of something, such as the color of shirts. These values are “categories” of a population, such as *M* or *F* for gender of people or “Don’t Drive” or “Drive” for the method of transportation used by students to get to school. These are examples of **binary** variables. These variables only have two possible values. Some categorical variables have more than two values, such as hair color, brand of jeans, and so on.

Two Types of Variables



Exercises: Answer the following questions and then decide if the data is quantitative or categorical. (*Q* or *C*)

	ANSWER	TYPE
1. In what grade did you take your first algebra class (Math I, Coord. Alg., etc.)?	_____	_____
2. How many pairs of shoes do you own?	_____	_____
3. How old was your father when you were born?	_____	_____
4. How old was your mother when you were born?	_____	_____
5. Choose a random integer from 1 to 20.	_____	_____
6. How many siblings do you have? (all, whether you live with them or not)	_____	_____

AP Statistics Prerequisite Packet

7. How many cousins do you have? _____
8. How tall are you (**in inches**)? _____
9. How many AP classes will you be taking **THIS** year? _____
10. What gender are you? _____
11. Where did eat your last meal?
(1 = home, 2 = restaurant, 3 = other) _____
12. How long have you lived in this area? _____
13. How far away from school do you live? _____

Numerical Descriptions of Quantitative Data

Measures of Center

Mean: The sum of all the data values divided by the number (n) of data values.

Example

$$\text{Data: } 4, 36, 10, 22, 9 \quad \text{Mean} = \bar{x} = \frac{\sum x_i}{n} = \frac{4+36+10+22+9}{5} = \frac{81}{5} = 16.2$$

Median: The middle element of an ordered set of data.

Examples

$$\text{Data: } 4, 36, 10, 22, 9 \quad = 4 \ 9 \ \underline{10} \ 22 \ 36 \quad \text{Median} = 10$$

$$\text{Data: } 4, 36, 10, 22, 9, 43 \quad = 4 \ 9 \ 10 \ | \ 22 \ 36 \ 43 \quad \text{Median} = \frac{10+22}{2} = 16$$

Measures of Spread:

Range: Maximum value – Minimum value

Example

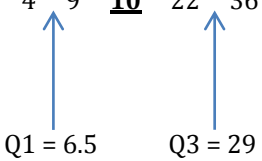
$$\text{Data: } 4, 36, 10, 22, 9 \quad = \quad 4 \ 9 \ 10 \ 22 \ 36$$

$$\text{Range} = \text{Max.} - \text{Min.} = 36 - 4 = 32$$

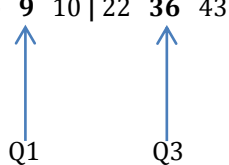
AP Statistics Prerequisite Packet

Interquartile Range (IQR): The difference between the 75th percentile (Q3) and the 25th percentile (Q1). This is $Q3 - Q1$. Q1 is the median of the lower half of the data and Q3 is the median of the upper half. In neither case is the median of the data included in these calculations. The IQR contains the middle 50% of the data. Each quartile contains 25% of the data.

Examples

1. Data: 4, 36, 10, 22, 9 = 4 9 10 22 36

 $Q1 = 6.5$ $Q3 = 29$

So, the IQR = $29 - 6.5 = 22.5$

2. Data: 4 9 10 | 22 36 43

 $Q1$ $Q3$

So, the IQR = $36 - 9 = 27$

Prior to doing all exercises in this packet, please read and reference pages 1-8 of the TI-84 guide so that you can learn how to do basic statistical operations using your calculator.

Exercises

Last year students collected data on the age of their moms and dads when they (the students) were born. The following are their results.

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										

AP Statistics Prerequisite Packet

1. Find the mean and the median for the Dad data.

Mean: _____ Median: _____

Are they the same? If not, which is larger?

2. Find the mean and the median for the Mom data.

Mean: _____ Median: _____

Are they the same? If not, which is larger?

3. Now compare the two means you calculated. Which is larger? _____

Is this result what you expected? _____

Why/why not?

4. Calculate the range for each set of data. Dad _____ Mom _____

5. Are these ranges the same? _____ If not, what could account for the differences?

6. Find the Q1 and Q3 for the Dad data: Q1: _____ Q3: _____

7. Find the Q1 and Q3 for the Mom data: Q1: _____ Q3: _____

8. You have now calculated the "Five-Number Summary." This can also be used as a way to determine the spread of a set of data. The five-number summary consists of:

Minimum Q1 Median Q3 Maximum

Write the five number summary for the Dad data: _____

Write the five number summary for the Mom data: _____

AP Statistics Prerequisite Packet

9. Now calculate the IQR for each of the two sets of data.

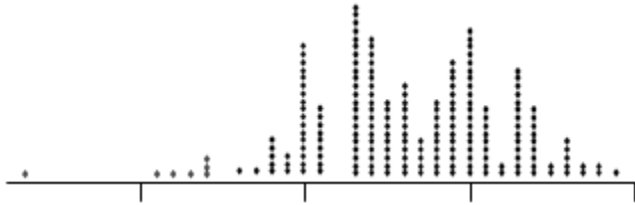
Dad: _____

Mom: _____

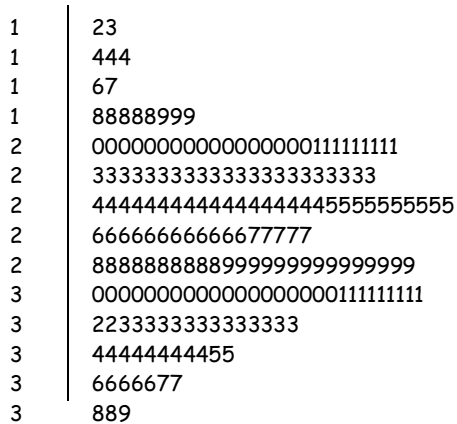
Graphical Displays of Univariate (one variable) Data

Quantitative Data: Dotplot
 Boxplot (Box and Whiskers)
 Stemplot (Stem and Leaf)
 Histogram

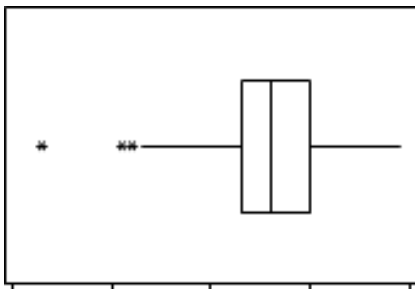
Dotplot of Student GPA's



Stemplot of Student GPA's

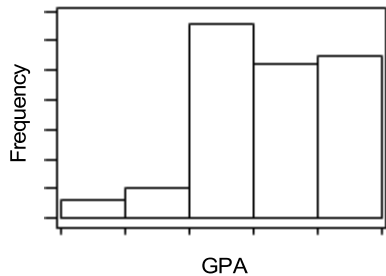


Boxplot of Student GPA's (see TI-84 guide on how to make these)



AP Statistics Prerequisite Packet

Histogram of Student GPA's (see TI-84 guide on how to make these)



Categorical Data: Bar Graph
Circle Graph

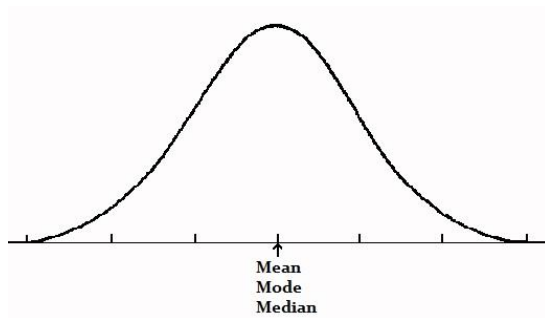
**I'm assuming that you already know how to make these two types of graphs.*

Assessing the Shape of a Graph

There are two basic shapes that we will examine: **Symmetric** and **Skewed**.

Symmetric: One can tell if a graph is symmetric if a vertical line in the “center” divides the graph into two fairly congruent shapes. (A graph does not have to be “bell-shaped” to be considered symmetric.)

Mean is approximately equal to the Median in a symmetric distribution



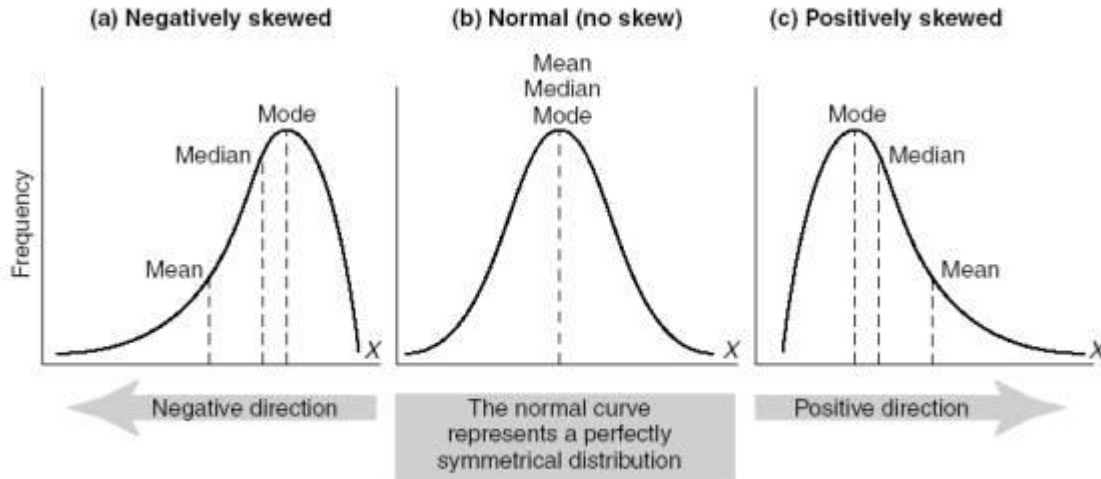
Skewed: One can tell that a graph is skewed if the graph has a big clump of data on either the left (skewed right) or on the right (skewed left) with a tendency to get flatter and flatter as the values of the data increase (skewed right) or decrease (skewed left). A common misconception is that the “skewness” occurs at the big clump – it does not!

Relationship between Mean and Median in a skewed distribution:

“Skewed Left, the mean is Less” or $mean < median$

“Skewed Right, the mean is Might” or $mean > median$

AP Statistics Prerequisite Packet



For the distributions above, graph (a) is more commonly referred to as a left-skewed distribution. Graph (c) is more commonly referred to as a right-skewed distribution.

Gathering Information from a Graphical Display

The first thing that should be done after gathering data is to examine it graphically and numerically to find out as much information about the various features of the data as possible. These will be important when choosing what kind of procedures will be appropriate to use to find out an answer to a question that is being investigated.

The features that are the most important are **S**hape, **O**utliers, **C**enter, **C**lusters and gaps, and **S**pread: **SOCCS**. Most of these can only be seen in a graph. However, sometimes the shape is indistinct – difficult to discern. So, in this instance (usually because of a very small set of data), it's appropriate to label the shape “indistinct.”

Exercises

1. Construct a boxplot for each the following sets of data taken from consumer ratings of different brands of peanut butter in the September, 2013, issue of Consumer Reports. **Use the same number line for both graphs.** (You could do it this way: Draw a number line. Above this line construct the "Crunchy" boxplot. Then, above the "Crunchy" boxplot, construct the "creamy" boxplot.) Please place your boxplots below.

Crunchy:	62	53	75	42	47	40	34	62	52	50
	34	42	36	75	80	47	56	62		
Creamy:	56	44	62	36	39	50	53	45	65	40
	56	68	41	30	40	50	56	30	22	

AP Statistics Prerequisite Packet

(a) Find the range for: Creamy: _____ Crunchy: _____

(b) Find the median for: Creamy: _____ Crunchy: _____

(c) Looking at your boxplots and comparing the medians what type of peanut butter do consumers tend to prefer?

2. The following data is taken from the Statistical Abstract of the United States (112th Edition). These are the ages of drivers arrested for DUI from a random sample of size 50. Make a stemplot to show the distribution of this age data.

45	16	41	26	22	33	30	22	36	34
63	24	26	18	27	24	31	38	26	55
31	47	27	43	35	22	64	40	58	20
49	37	53	25	29	32	23	49	39	40
24	56	30	51	21	45	27	34	47	35

(a) What is the shape of this graph? _____

(b) Using your stemplot, find the median of this data. _____

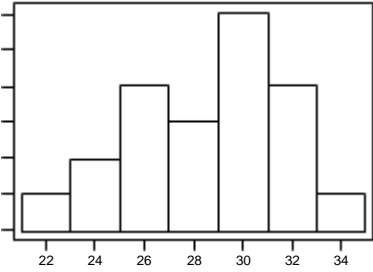
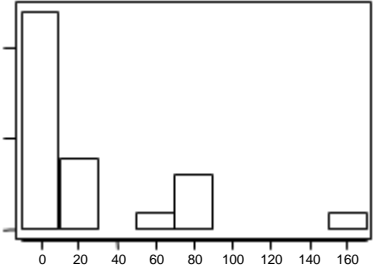
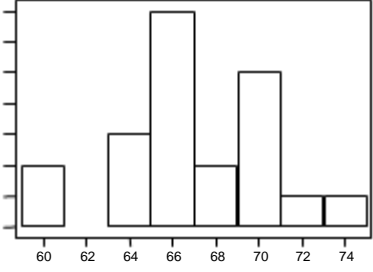
(c) Which data display is better – a boxplot or a stemplot? _____

Why? (Be specific.)

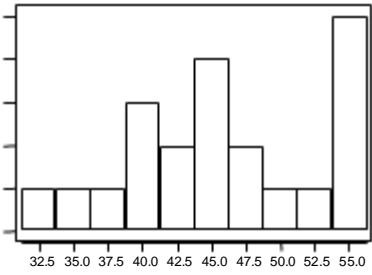
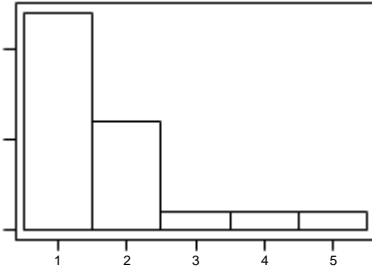
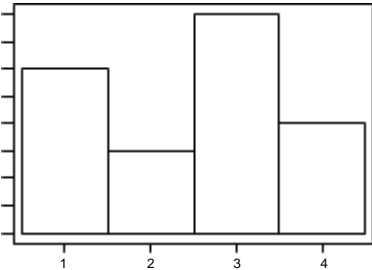
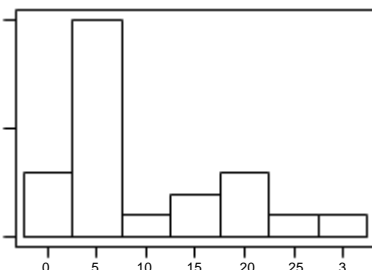
AP Statistics Prerequisite Packet

3. For the following graphs, find the center (**just do the median**), shape, and spread (find only the **range**). If there are 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.

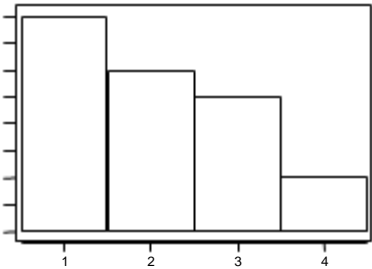
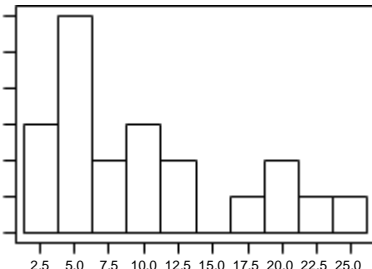
Note: To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. Add these up and divide by two. This tells you where the median is located. Which bar is this value in? That's the median. For graph A, $n = 21$, so the middle value is 10.5. Starting with 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. To find a **VERY** rough estimate of the mean, take the frequency for each bar and multiply it by the value along the x-axis for that bar. Add these up for all the bars and then divide by 21. You get the mean = 28.571.

<p>Graph A</p> 	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p>
<p>Graph B</p> 	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p> <hr/> <p>Clusters or gaps? Where?</p> <hr/> <p>Outliers? Where?</p>
<p>Graph C</p> 	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p> <hr/> <p>Clusters or gaps? Where?</p>

AP Statistics Prerequisite Packet

<p>Graph D</p>  <p style="font-size: small; text-align: center;">32.5 35.0 37.5 40.0 42.5 45.0 47.5 50.0 52.5 55.0</p>	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p>
<p>Graph E</p>  <p style="font-size: small; text-align: center;">1 2 3 4 5</p>	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p> <hr/> <p>Clusters or gaps? Where?</p>
<p>Graph F</p>  <p style="font-size: small; text-align: center;">1 2 3 4</p>	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p>
<p>Graph G</p>  <p style="font-size: small; text-align: center;">0 5 10 15 20 25 3</p>	<p>Center?</p> <hr/> <p>Shape?</p> <hr/> <p>Spread?</p>

AP Statistics Prerequisite Packet

<p>Graph H</p>  <p>A histogram with four bars. The x-axis is labeled 1, 2, 3, 4. The y-axis has tick marks. The bars decrease in height from left to right.</p> <table border="1"><thead><tr><th>Category</th><th>Frequency</th></tr></thead><tbody><tr><td>1</td><td>10</td></tr><tr><td>2</td><td>7</td></tr><tr><td>3</td><td>5</td></tr><tr><td>4</td><td>2</td></tr></tbody></table>	Category	Frequency	1	10	2	7	3	5	4	2	<p>Center?</p> <p>Shape?</p> <p>Spread?</p>												
Category	Frequency																						
1	10																						
2	7																						
3	5																						
4	2																						
<p>Graph I</p>  <p>A histogram with ten bars. The x-axis is labeled 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0. The y-axis has tick marks. There are two peaks: one at 5.0 and another at 20.0.</p> <table border="1"><thead><tr><th>Category</th><th>Frequency</th></tr></thead><tbody><tr><td>2.5</td><td>4</td></tr><tr><td>5.0</td><td>10</td></tr><tr><td>7.5</td><td>3</td></tr><tr><td>10.0</td><td>5</td></tr><tr><td>12.5</td><td>3</td></tr><tr><td>15.0</td><td>1</td></tr><tr><td>17.5</td><td>2</td></tr><tr><td>20.0</td><td>4</td></tr><tr><td>22.5</td><td>2</td></tr><tr><td>25.0</td><td>2</td></tr></tbody></table>	Category	Frequency	2.5	4	5.0	10	7.5	3	10.0	5	12.5	3	15.0	1	17.5	2	20.0	4	22.5	2	25.0	2	<p>Center?</p> <p>Shape?</p> <p>Spread?</p> <p>Clusters or gaps? Where?</p>
Category	Frequency																						
2.5	4																						
5.0	10																						
7.5	3																						
10.0	5																						
12.5	3																						
15.0	1																						
17.5	2																						
20.0	4																						
22.5	2																						
25.0	2																						

AP Statistics Prerequisite Packet

4. Use the following list of variables to identify which of the graphs in Question 3 **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 repeats.

Variable	Your Answer	Correct Graph?
Grade when a student takes their first algebra course		
Number of pairs of shoes you own		
Age of your father when you were born		
Age of your mother when you were born		
Number of siblings you have		
Number of cousins you have		
Your height (in inches)		
Number of AP classes you'll take this year		
How long you have lived in this area		
How far away you live from school (in miles)		
Amount of change in your pocket today		