Name: _____ Period: _____

Summer Assignment
for AP Statistics
Mrs. Whitt

If you have questions over the summer, email me at: swhitt@lvhs.org
I check my Lenape Valley email over the summer and will answer your questions

Summer 2015

To all students enrolled in AP Statistics and their parents/guardians,

Congratulations on your decision to enroll in AP Statistics! You have decided to take a course that is often a requirement in many college majors in the fields of health science, social science, and business. This course will also have you thinking about how you interpret data to make informed decisions as a consumer.

Although AP Statistics requires only an Algebra II prerequisite, it is not an easy course. As for all AP courses, it does require a great deal of commitment within and outside the classroom. AP Statistics is a special course that combines both mathematical and verbal skills. You will be required to write descriptive paragraphs and concluding sentences to prepare you for the AP Exam. You will have to explain the reasoning behind the method you choose and the conclusions you find.

AP Statistics will expose you to four central themes: exploring data, research design, probability and simulation, and statistical inference. The AP Exam contains questions from each of these 4 clusters. Each chapter test and the AP Exam are comprised of multiple choice and open ended questions.

How can we prepare for this class? Although many of the concepts in the Course are introduced for the first time in this class there are some concepts to review to help you prepare for this course. Complete this summer packet before the first day of class. This packet will be collected for a grade on the first day. I will be starting class assuming you have completed the packet.

****All AP Statistics students are required to have a graphing calculator.** I recommend the TI-84. You may want to watch out for back-to-school sales.

Some sections of the Summer work packet contains notes, while other sections contains problems to work out. Read the entire packet and work out the problems indicated.

I look forward to seeing you next school year!

Sincerely, Mrs.

Whitt

AP Statistics Teacher
swhitt@lvhs.org (Email me over the summer if you have any questions as you work

through the packet.)

Part 1- Algebra Review (Practice Problems):

1.) Evaluate z if $z = \frac{x - \mu}{\sigma}$ where x = 20, $\mu = 10$, and $\sigma = 2$.

2.) Given: $z = \frac{x - \mu}{\sigma}$, solve for σ.

3.) Given: $z = \frac{x - \mu}{\sigma}$, solve for μ.

4.) Given: $0.05 = 1.96\sqrt{\frac{0.5^2}{n}}$, solve for n.

5.) Given: $-1.64 = \frac{60 - \mu}{\sigma}$ and $1.96 = \frac{95 - \mu}{\sigma}$, solve for μ.

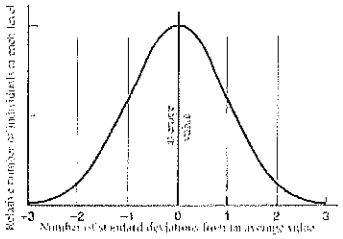
6.) Given: $-1.64 = \frac{60-\mu}{\sigma}$ and $1.96 = \frac{95-\mu}{\sigma}$, solve for σ.

7.) Find the equation of the line in slope intercept form (y=mx + b) that goes the points (-2,4) and (5,7)

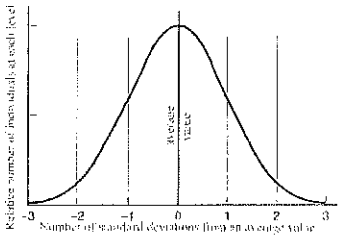
8.) Given: $\log_{10} 100 = x$ evaluate $10^x =$ _____

9.) Given: $\ln 100 = x$ evaluate $10^x =$ _____

10.)Shade the area under curve where $z \geq 2$ (Z is the horizontal axis at the bottom of the curve)



11.)Shade the area under the curve where $z \leq -1$



Part 2- Types of Data (Notes and Practice Problems):

Quantitative *(or numerical) Data*

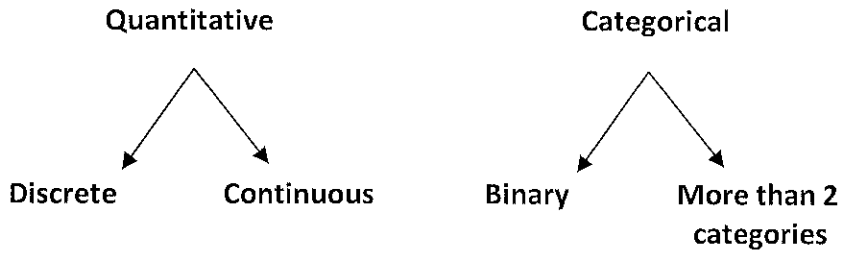
There are data that take on numerical values that actually represent a measurement such as size, weight, how many, how long, score on a test, etc. Some quantitative variables take on **discrete** (counting) values, such as shoe size (6, $6\frac{1}{2}$, 7,...) or the number of soup cans collected by a school.

Other quantitative variables take on **continuous** (measurement) values, such as height (60 inches, 72,999923 inches, 64.039 inches, etc) or how much water it takes to fill up your bathtub (73.293 gallons, or 185.3 gallons or 99 gallons, etc). For these data, it makes sense to find things like "average" or "range" (largest value – smallest value).

Categorical (or qualitative) data

There are data that take on values that describe a characteristic such as color of shirts or gender. Some qualitative variables take on **binary** values , such as gender (M or F) or an answer to the question, "Are you going to the prom?" (yes or no). For these data, it makes no sense to find things like "average" or "range". To analyze these data, you count the number of Males or Females, the number of Yes responses and No responses. Categorical data can sometimes have more than 2 answer choices. An example might include "World language studied in high school". These data could take on the values, Spanish, French, Chinese, Russian, etc.

Two types of variables:



Exercises: Answer the following questions and then decide if the data is quantitative or categorical. (Q or C) If the variable is quantitative, then **also state** whether it is discrete or continuous. (D or C)

	Answer	**Type**
1. In what grade did you take Algebra 1?	_____	_____
2. Do you own more than 20 pairs of shoes?	_____	_____
3. How old was your father when you were born?	_____	_____
4. How many pets do you have?	_____	_____
5. What is your hair color?	_____	_____
6. How many siblings do you have?	_____	_____
7. In what state were you born?	_____	_____
8. How tall are you measured in inches?	_____	_____
9. How many AP classes will you take this year?	_____	_____
10. What is your gender?	_____	_____
11. Where did you eat your last meal?	_____	_____
(1=home, 2=restaurant, 3=other)		
12. How long have you lived in this area?	_____	_____
12. Are you enrolled in Honors English this year	_____	_____
13. What is your GPA?	_____	_____
14. Is your car a standard or manual?	_____	_____

Part 3- Graphical Displays of Univariate (one variable) Data –

Directions for how to make graphs (Notes):

Quantitative Data: 1. Dotplot

2. Stemplot (Stem and Leaf Plot)

3. Boxplot (Box and Whisker Plot)

4. Histogram

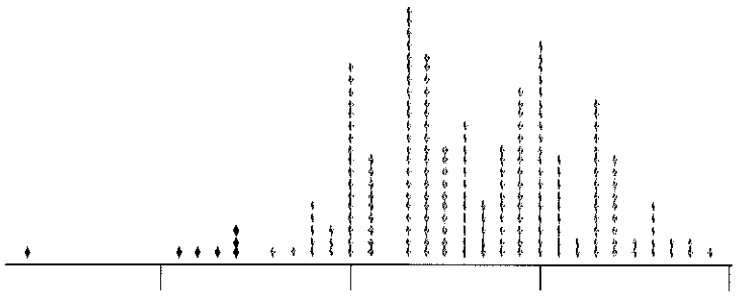
1. The Dotplot:

To make a Dotplot:

1. Draw and label a number line so that all the values in your dataset will fit.

2. Graph each of the data values with a dot. Be sure to line the dots up vertically as well as horizontally so that you can really see the shape of the graph.

Dot plot of Student GPAs



GPA

2. The Stemplot

To make a Stemplot:

1. Put the data in ascending order.

2. Use only the last digit of the number as a leaf (see the numbers to the right of the line –
   each digit is the last digit of a larger number).

3. Use one, two, or more digits as the stem. (Sometimes, you can truncate data when there are too many digits in each data value – i.e. the number 20, 578 would become 20 | 5, where the "20" is in thousands. Note that this is **different** from rounding.)

4. Place the "stem" digit(s) to the left of the line and the leaf digit to the right of the line.
   Do this for each data value. You should then arrange the "leaves" in ascending order.

5. Sometimes, there are many numbers with the same "stem." In this situation it might be useful to break the numbers with the same stem into either two distinct groups (each on a separate line; say, "leaves" from 0 – 4 on the first line and 5 – 9 on the second.) or into five distinct groups as is shown in the graph to the right. Here, the first line for each stem contains all the 0 - 1 leaves, the next line contains the 2 – 3 leaves and so on. This technique is called "splitting the stems." It is useful in some cases in order to show the shape of the data more clearly.

Stemplot of Student GPAs

1	23
1	444
1	67
1	88888999
2	00000000000000000111111111
2	333333333333333333333
2	44444444444444444445555555555
2	66666666666677777
2	88888888899999999999999
3	000000000000000000000111111111
3	2233333333333333
3	44444444455
3	6666677
3	889

3. The Boxplot:

To make a Boxplot:

1. **Draw and label a number line** that includes the minimum and the maximum values for the set of data.

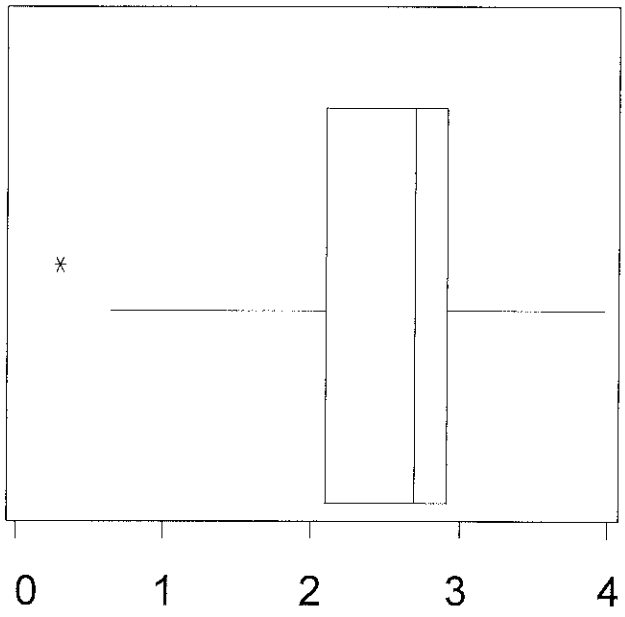
2. Calculate the five-number summary and make a dot for each of these summary numbers above the number line.

3. Draw a line between the 1st and 2nd dot, showing the "lower quartile"; and then draw a line from the 4th
to the 5th dot to show the "upper quartile." These are commonly called the "whiskers."

4. Draw a rectangular box from the 2nd to the 4th dot and draw a line through the box on the middle dot –
the median.

NOTE: ***In AP Statistics, a "modified boxplot" is used. This shows any "outliers." An outlier is a data point that does not fit the pattern of the rest of the data. When your calculator or computer software graphs a modified boxplot, 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 and below the box). These outliers are shown with dots or stars, or any other small symbol.***

Boxplot of Student GPAs



4. The Histogram

To make a histogram:

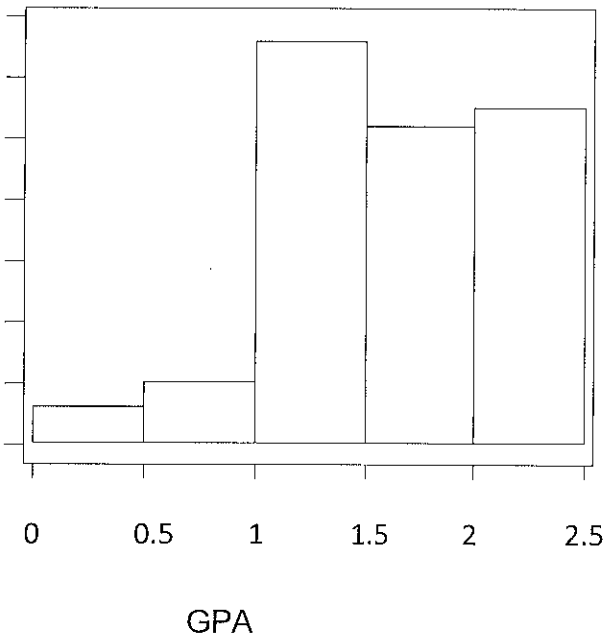
1. Put the data into ascending order.

2. Decide upon evenly spaced intervals into which to divide the set of data (such as 0, 10, 20, 30, etc.) and then count the number of values that fall within each interval. This number is called the "frequency." If you divide each of these frequencies by the size of the data set, *n*, making percents, then you have what are called "relative frequencies."

3. Draw and **label** a 1st quadrant graph using scales appropriate for the data. Be sure to include a title for the x- and for the y-axes.

4. Graph the frequencies that you calculated in step 2.

Histogram of Student GPAs



Part 4- Numerical Descriptions of Univariate (one variable) Quantitative Data – (Notes and Practice Problems):

Measures of Center

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

Example

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

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

Examples

Data: 4, 36, 10, 22, 9 = 4 9 **10** 22 36 ⟶ Median = 10

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

Mode: The most frequently element in a set of data.

Examples

Data: 4, 4, 5, 6, 7 = 4

Data: 4, 5, 6, 7, 8 = no mode (Note: The mode is not 0. Zero is a number.)

Data: 5, 6, 7, 7, 8, 9, 10, 10, = 7 and 10

Measures of Spread:

Range: Maximum value – Minimum value

Example

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

Range = Max. – Min. = 36 – 4 = **32**

Interquartile Range (IQR): The difference between the 75th percentile (Q_3) and the 25th percentile (Q_1). This is $\mathbf{Q_3 - Q_1}$. Q_1 is the median of the lower half of the data and Q_3 is the median of the upper half. In neither case is the median of the data included in these calculations.

The IQR contains 50% of the data. Each quartile contains 25% of the data.

Examples

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

$Q_1 = 6.5$ $Q_3 = 29$

So, the IQR = 29 - 6.5 = 22.5

2. Data: 4 **9** 10 | 22 **36** 43

Q_1 Q_3

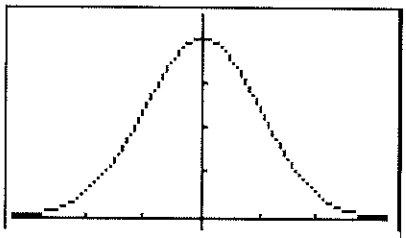
So, the IQR = 36 - 9 = 27

Part 5- Graphically Assessing Shape of a Graph:

Assessing the Shape of a Graph

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

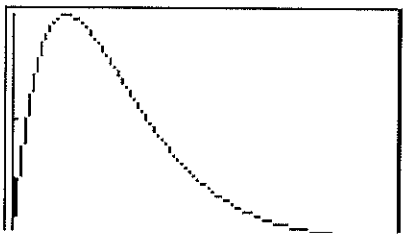
Symmetric: One can tell if a graph is symmetric if a vertical line in the "center" divides the graph into two fairly congruent shapes. (A graph does not have to be "bell-shaped" to be considered symmetric.)



Symmetric

Mean ~ 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.



Skewed Right

Relationship between Mean and Median in a skewed distribution:

Skewed Left, the mean is Less.

Skewed Right, the mean is Might.

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 Shape, Center, Spread, Clusters and gaps, Outliers: SCSCO. 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

Part 6 - Numerical Descriptions - Practice Problems:

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										

Now type the data into your calculator using the list function: **STAT** → **ENTER** → type the Dad into L_1. If you make a mistake, you can go to the error and **DELETE**. If you forget an item, you can go to the line below where it is supposed to be and press **2nd DEL** to insert it. Do the same for the Mom data, but put into L_2.

> NOTE: If the lists you are using already have numbers in them before you start, you can clear them this way: Arrow up (↑) to the line where L_1 is shown. Press CLEAR, then the down arrow (↓).

1. Find the mean and the median for the Dad data. To find the mean using your calculator, go to 2nd STAT → MATH → 5 and then type in L_1 by typing 2nd → 1. This will add all the values in the list. Then divide by 26 to get the mean. If you do not have a calculator, please calculate by hand.

To find the median, sort the data in the lists: **STAT→ 2 → L_1** The median is exactly in the middle between the 13th and the 14th value. If you do not have a TI-84 graphing calculator yet, you may of course find these values by hand.

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 about the same? _____ If no, what are some reasons that might cause this difference?

6. Find Q_1 and Q_3 for the Dad data. Q_1_____ Q_3_____

7. Find Q_1 and Q_3 for the Mom data. Q_1_____ Q_3_____

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	Q_1	Median	Q_3	Maximum

Write the five number summary for the Dad data: _____

Write the five number summary for the Mom data: _____

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

   Dad_____

   Mom_____

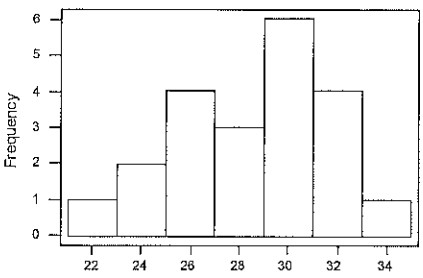
Part 7- Graph Analysis:

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:

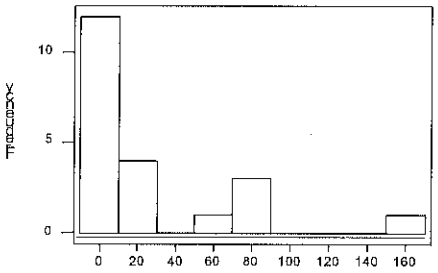
To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. Add these up and divide by two. This tells you where the median is located. Which bar is this value in? That's the median. For graph A, n = 21, so the middle value is 10.5. Starting with the first 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. 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.

A



Shape _____
Center _____
Spread _____

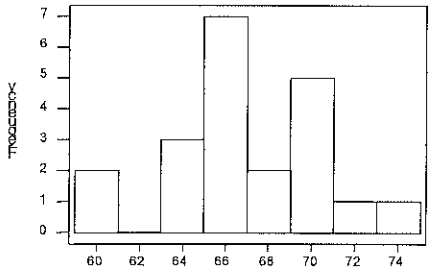
B



Shape_____
Center_____
Spread _____
Clusters, Gaps? ____ Where?

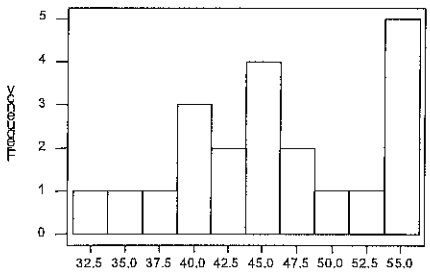
Outliers? _____ Where?

C



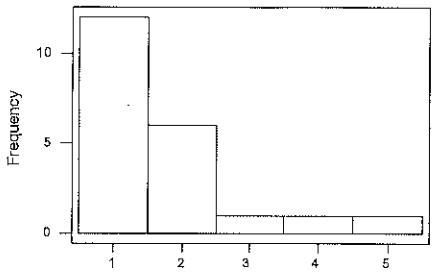
Shape _____
Center _____
Spread _____
Clusters, Gaps? ____ Where?

D



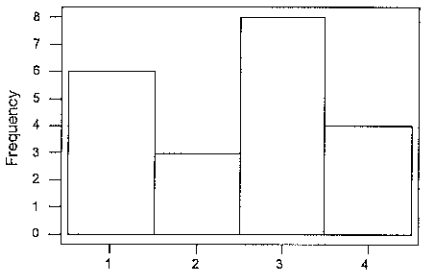
Shape _____
Center _____
Spread _____

E



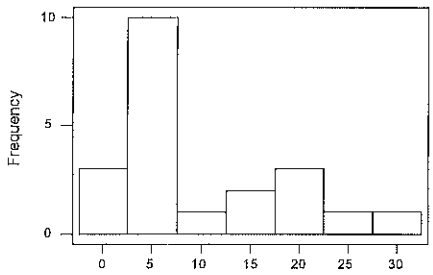
Shape _____
Center _____
Spread _____
Clusters? ____ Where?

F



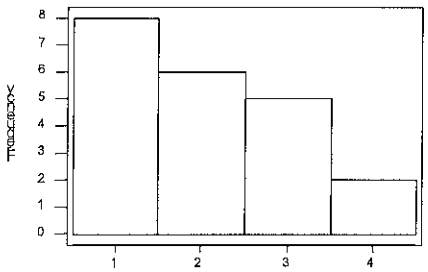
Shape _____
Center _____
Spread _____

G



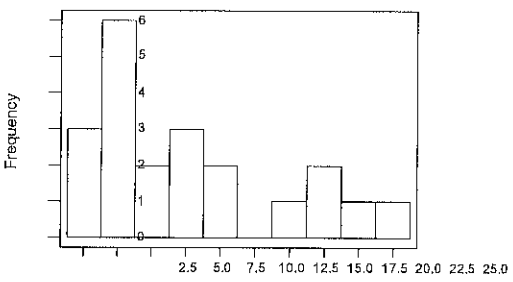
Shape _____
Center _____
Spread _____

H



Shape _____
Center _____
Spread _____

I



Shape _____
Center _____
Spread _____
Clusters, Gaps? ____ Where?

Part 8 - Additional Practice Problems.:

1. This table shows the median US family income (in 2003 dollars) for selected years. ***Let x be the number of years since 1940.***

Year	Median Family Income ($)
1947	21,201
1973	43,219
1979	45,989
1989	49,014
1995	48,679
2000	54,191
2003	52,680

a. Enter the data into your graphing calculator and create the scatter plot for this data. Draw the scatter plot on 8 ½ by 11 inch graph paper drawn to scale.

b. Find the linear regression line using your graphing calculator. (See calculator user's guide if needed.)

c. Superimpose the regression line on the scatter plot.

d. Use your regression line to predict the median US family income in 2013.

e. Calculate the mean, median and mode of the family income.

2. Using 20th century US census data, the population of New York can be modeled by:

$$P(t) = \frac{19.875}{1 + 57.993e^{-0.035005t}}$$

where P is the population in millions and t is the number of years since 1800. Based on the model,

a. What is the population of New York in 1852?

b. What will the population be in 2015?

c. What is New York's maximum sustainable population (limit to growth)?

3. A projectile is launched straight up from ground level with an initial velocity of 288 feet/second.

Projectile motion: $s = -16t^2 + v_0 t + s_0$

Where :
S_0 = initial height
V_0 =initial velocity
s = vertical position t seconds after launch

a. When will the height of the projectile be 1152 feet above the ground?

b. When will the height of the projectile be **at least** 1152 feet above the ground?

4. When a ball is bouncing up and down on a flat surface, its height with respect to time can be modeled using a quadratic function. One form of a quadratic function is the vertex form: y= $a(x-h)^2+k$. In this equation, y represents the height of the ball and x represents the elapsed time. The following is sample data of a bouncing ball over time.

Elapsed Time (seconds)	Height of the Ball (meters)
0.688	0
0.731	0.155
0.774	0.309
0.817	0.441
0.860	0.553
0.903	0.643
0.946	0.716
0.989	0.773
1.032	0.809
1.075	0.828
1.118	0.828
1.161	0.811
1.204	0.766
1.247	0.721
1.290	0.650
1.333	0.563
1.376	0.452
1.419	0.322
1.462	0.169

a. Enter this data into your graphing calculator and create a scatter plot. Draw the scatter plot on 8 ½ by 11 inch graph paper drawn to scale.

b. Find the value of a, h and k so the equation $a(x-h)^2+k$ fits one of the bounces in the data. Do this by approximating the vertex (h,k) from the scatter plot and graph on your graphing calculator and solving algebraically for a. Give the equation.

5. Enter the following data into your graphing calculator. Create a scatter plot of the data. Draw the scatter plot on graph paper drawn to scale.

(0,15) (1,40) (1,35) (2,45) (2,55) (2,60) (3,45) (3,40) (4,56) (5,60) (6,72) (7,80) (7,85) (8,96) (8,92) (9,85)

a. This (x,y) data represents the final exam scores and the number of hours studied for the Honors Chemistry Final Exam. Which variable represents the number of hours studied? Which variable represents the Final Exam grade?

b. (It's time to do some writing.) Describe the scatter plot in as much detail as possible. Does the data suggest a linear, quadratic, exponential or another relationship? Describe the slope. Draw a conclusion about the number of hours studied and the Final Exam grade. Estimate the number of hours necessary needed to study to get a 100% on the test.