Day 1 – Introduction to Statistics

Statistics is the field of math that involves collecting, analyzing, interpreting, and presenting data.

When is statistics actually used in real life?

All of the time! It is the one field of math that relates to almost every other topic in life. *When used correctly, statistics tell us any trends in what happened in the past and can be useful in predicting what may happen in the future.*

Here are some examples of how statistics shape your life when you don't even know it.

1. Weather Forecasts

Do you watch the weather forecast sometime during the day? How do you use that information? Have you ever heard the forecaster talk about weather models? These computer models are built using statistics that compare prior weather conditions with current weather to predict future weather.

2. Emergency Preparedness

What happens if the forecast indicates that a hurricane is imminent or that tornadoes are likely to occur? Emergency management agencies move into high gear to be ready to rescue people. Emergency teams rely on statistics to tell them when danger may occur.

3. Predicting Disease

Lots of times on the news reports, statistics about a disease are reported. If the reporter simply reports the number of people who either have the disease or who have died from it, it's an interesting fact but it might not mean much to your life. But when statistics become involved, you have a better idea of how that disease may affect you.

For example, studies have shown that 85 to 95 percent of lung cancers are smoking related. The statistic should tell you that almost all lung cancers are related to smoking and that if you want to have a good chance of avoiding lung cancer, you shouldn't smoke.

4. Medical Studies

Scientists must show a statistically valid rate of effectiveness before any drug can be prescribed. Statistics are behind every medical study you hear about.

5. Genetics

Many people are afflicted with diseases that come from their genetic make-up and these diseases can potentially be passed on to their children. Statistics are critical in determining the chances of a new baby being affected by the disease.

6. Political Campaigns

Whenever there's an election, the news organizations consult their models when they try to predict who the winner is. Candidates consult voter polls to determine where and how they campaign. Statistics play a part in who your elected government officials will be

7. Insurance

You know that in order to drive your car you are required by law to have car insurance. If you have a mortgage on your house, you must have it insured as well. The rate that an insurance company charges you is based upon statistics from all drivers or homeowners in your area.

8. Consumer Goods

Wal-Mart, a worldwide leading retailer, keeps track of everything they sell and use statistics to calculate what to ship to each store and when. From analyzing their vast store of information, for example, Wal-Mart decided that people buy strawberry Pop Tarts when a hurricane is predicted in Florida! So they ship this product to Florida stores based upon the weather forecast.

9. Quality Testing

Companies make thousands of products every day and each company must make sure that a good quality item is sold. But a company can't test each and every item that they ship to you, the consumer. So the company uses statistics to test just a few, called a sample, of what they make. If the sample passes quality tests, then the company assumes that all the items made in the group, called a batch, are good.

10. Stock Market

Another topic that you hear a lot about in the news is the stock market. Stock analysts also use statistical computer models to forecast what is happening in the economy.

Day 2 - Statistical Measurements

(round to the nearest *hundredth* when necessary)

Data: 16, 27, 32, 20, 20, 37, 123, 36, 40, 29, 43, 35, 41, 20, 36, 34	
In Order:	
Measures of Central Tendency:	**Calculations:**
Mean: The average of all of the data values; $\frac{\text{sum of scores}}{\text{number of scores}}$	
Median: The middle number or the average of the two middle numbers when the data is written in order	
Mode: The data item that occurs the most often; there can be zero, one, or multiple modes	
Measures of Dispersion:	**Calculations:**
Range: The difference between the highest and lowest scores	
Quartiles: The 3 numbers that split the data into 4 equal parts: 1^{st} quartile = Q1 = 25^{th} percentile 2^{nd} quartile = Q2 = 50^{th} percentile = median 3^{rd} quartile = Q3 = 75^{th} percentile	
Interquartile Range: The difference between the 1^{st} quartile and the 3^{rd} quartile: IQR = Q3-Q1	
Standard Deviation: Shows, on average, how all the scores differ from the mean	

Why you probably have an above-average number of feet

Rebecca Goldin Ph.D , March 10, 2008
Looking at three kinds of averages

The average kid on the block might have a lot of trouble understanding what an average is. Every time Garrison Keillor signs off his "News from Lake Wobegon," as a place "where all the women are strong, all the men are good-looking, and all the children are above average," he gets a laugh. But the mathematical meaning of average is not always the same as the colloquial meaning – and even within math, there are three different kinds of "averaging" that are commonly referred to.

The average of a bunch of numbers is the one that we all learn in school – if only because our grade might have been determined by the average of our scores on the tests. This average is formally called the *mean* of the numbers. It's computed by adding up *n* numbers and then dividing by that *n.* But the mean can be really misleading; for example, most people earn below average salaries, but have an above-average number of feet, as an excellent BBC article on this topic points out.

The main point for salaries is that the average is easily affected by a few people making a ton of money. For example, suppose three people make $47,000 each, three people make $50,000 each, three people make $53,000 each, and one person makes $500,000 (all annually). Their combined average salary is $95,000, even though almost everyone makes a lot less than that. For salary, a more appropriate number might be the *median*, which is a number such that half of everyone makes more, and half of everyone makes less. In our mock situation with only ten people, the median would be $50,000 – which is a good estimate for most people in the sample. According to the Census Bureau, the median household income in 2007 was $50,233.00. The mean is over $60,000 because those top earners have a higher weight in the average.

A different problem happens when you look at the number of feet people have. In this case, almost everyone has two feet, but there are a few people who have just one or no feet. Suppose we have five people, say four have two feet and one has one foot . If you take the average number of feet, you will find that the average is 1.8 feet. It follows that almost everyone has an above-average number of feet. In this case, the mean is again an inappropriate number to look at – we would do better to think about the *mode*, which is the number that occurs most frequently in the data set. Since four people have two feet, "two" will occur most often in the data. Why is mode a better choice than median? Mode is a good choice to use for a data set with a small set of values – in this case, the possibilities are zero, one or two, while median is a better choice for data that can vary continuously, like income.

Questions:

1. Even though the mean or average is the most commonly used measure of central tendency, it is not always the best measure. Why can the mean be misleading?

2. When is the median the best measure of central tendency?

3. When is the mode the best measure of central tendency?

Let's collect our own set of data and we will learn how to use the graphing calculator to find most of these statistical measurements. (Use common sense for rounding in these problems.)

Most you ever spent on a gift
1)
2)
3)
4)
5)
6)
7)
8)
9)
10)
11)
12)
13)
14)
15)
16)
17)
18)
19)
20)

Number of Scores:

In Order:

Mean:

Median:

Mode:

Which one would you use? Why?

Range:

Interquartile Range:

Standard Deviation:

Calculator Directions:

1. Press STAT.
2. Press 1 to choose Edit.
3. Enter the data in List 1 (L_1). Be sure to press ENTER or the down arrow after each entry.
4. Go back to STAT, choose #2: SortA, press 2^{nd} L1, and ENTER. If you go back to your list, it will now be in order. This will help you find the mode.
5. To find basic statistical information on your list, press STAT, go over to CALC, choose #1 (1-Var Stats).

Statistical Calculation	Calculator Symbol	Answer and meaning for this set of data
Mean		
Median		
Mode		
Maximum		
Minimum		
Range		
1st Quartile		
2nd Quartile		
3rd Quartile		
Interquartile Range		
Standard deviation		

Number of Siblings

Number of Scores:

In Order:

Mean:

Median:

Mode:

Which one would you use? Why?

Range:

Interquartile Range:

Standard Deviation:

Height in Inches

Number of Scores:

In Order:

Mean:

Median:

Mode:

Which one would you use? Why?

Range:

Interquartile Range:

Standard Deviation:

Number of Hours of Work Per Week

Number of Scores:

In Order:

Mean:

Median:

Mode:

Which one would you use? Why?

Range:

Interquartile Range:

Standard Deviation:

?	?

Number of Scores:
In Order:

Mean:

Median:

Mode:

Which one would you use? Why?

Range:

Interquartile Range:

Standard Deviation:

Day 3 - Percentiles

(Round to the nearest *percent*)

The percentile of a value is a measurement comparing one value to the others in that group. Remember, all percents are found by using the formula, $\frac{part}{whole} * 100$. The challenge is figuring out what is the whole and what part of the whole are you comparing. Make sure the numbers are arranged from smallest to largest!

Percentile Rank: *The percent of values that fall at or below a certain value.*

Whole:

Part:

Formula:

$$\frac{\text{number of lower values} + .5(\text{number of equal values})}{\text{total number of values}} \bullet 100$$

Formula in your own words:

1. Let's say the grades on the last math test in your class were, 65, 72, 77, 78, 80, 80, 83, 85, 90, 92, 95, and 97.
You received the 78. What was your percentile rank?

2. Let's say the grades on the last math test in another Math 3 class were 52, 65, 69, 72, 74, 77, 78, 78, 80, 83, 85, 92, and 97.
What was the percentile rank of a student with a 78 in this class?

Conclusion:

3. The following are the point totals midseason for my modified volleyball team:
17, 6, 0 ,32, 38, 26, 19, 38, 17, 9,13, 24, 24, 24, 42, 52.

a. What would be the percentile rank of the girl who has 13 points?

b. What would be the percentile rank of one of the girls who has 24 points?

4. Find YOUR percentile rank for the most spent on a gift, number of siblings, height, number of hours of work, and _____ surveys we did the other day.

a. Most Spent on a Gift:

Total number of scores:

Number of scores at or below your score:

b. Number of Siblings:

c. Height:

d. Number of Hours of Work:

e. ?

5. What would it mean if you ranked in the 50^{th} percentile on a test?

6. Could someone rank in the 90^{th} percentile on a test and still have failed it? Why or why not?

7. Could someone rank in the 20^{th} percentile on another test and still have passed it? Why or why not?

Day 4 - Movie Preferences: Gathering Data

To choose the ending for a movie, the producer often makes two or more endings and uses them on a test audience. Each member of the audience votes on which ending he or she likes the best, and the results are used to choose the ending. The following is a plot summary for the movie, *I Am Legend,* the original ending, and the alternate ending. There has been controversy over which ending should be used. Our class is going to act as a new test audience and choose the ending we like the best.

I am Legend: Plot Summary

After the spread of a lethal virus, U.S. Army virologist Lieutenant Colonel Robert Neville (Will Smith) is left as the last healthy human in New York City, and possibly the entire world.

The story, set in 2012, opens with a series of flashbacks and recorded news programs which reveal that three years earlier a genetically re-engineered measles virus, originally created as a cure for cancer, mutated into a lethal strain which rapidly infected humans and some animal species. By the end of the first year following the infection, more than 90% of the planet's human population died. Over 9% were infected, but did not die, instead degenerating into a physically deformed, primal, aggressive state, with a painful sensitivity to UV radiation that forces them to hide in buildings and other dark places during the day. The remaining 1% were immune to the virus, but were hunted and killed by the infected, killed by fellow survivors, or committed suicide due to isolation. Three years after the outbreak, Robert Neville fears he may be the last healthy human in the world.

In a Manhattan devoid of humanity, Neville's daily routine includes experimentation to find a cure for the virus on both infected rats and humans, and trips through the city to hunt for food and forage for supplies. He waits each day for a response to his automated, continuous radio broadcasts, which instruct any uninfected survivors to meet him at midday at the South Street Seaport. Flashbacks reveal that his wife and daughter died during the chaotic evacuation of Manhattan, just prior to the militarily-enforced quarantine of the island on Christmas Eve 2009. Neville's isolation is broken only by the companionship of his dog Samantha ("Sam"), interaction with mannequins he has set up as patrons of a video store, and recordings of old news and entertainment broadcasts.

Neville finds a promising treatment derived from his own blood, so he sets a snare trap and captures an infected woman while an infected male watches from within a building. At his laboratory, located in the basement of his heavily-fortified home, Neville treats the infected woman without success. Later, while driving through New York, he notices one of the mannequins has been moved. When he inspects it, he is ensnared in a trap similar to the one he used to capture the woman. By the time Neville escapes, it is near dark and he is attacked by infected dogs (released by the infected male), one of which bites Sam.

Neville takes Sam home and injects her with a serum in the hope of saving her, but when she shows signs of infection and tries to attack him, Neville kills her. The next night, he goes out and attacks a large group of infected with his SUV. He is almost killed, but is rescued by survivor Anna Montez (Alice Braga), who has traveled with a young boy named Ethan (Charlie Tahan) from Maryland after hearing one of Neville's broadcasts. They take the injured Neville back to his home where Anna explains that they survived the outbreak aboard a Red Cross evacuation ship from São Paulo and are making their way to a rumored survivors' camp in Bethel, Vermont.

The next night a group of infected—who have followed Anna and Neville and are led by the Alpha Male—attack the house and overrun its defenses.

Ending 1

Neville, Anna, and Ethan retreat into the basement laboratory, sealing themselves in with the infected woman Neville has been treating. Discovering that the last treatment is starting to work and the woman's signs of infection are diminishing, Neville draws a vial of her blood and gives it to Anna. He pushes Anna and Ethan into an old coal chute, and then sacrifices himself to save their lives, using a hand grenade to kill the attacking infected.

Anna and Ethan escape to Vermont and locate the survivors' colony, where Anna hands over the cure. In the closing voice-over, she states that Neville's cure enabled humanity to survive and rebuild, establishing his legend.

Ending 2

At that point, Neville's assumptions about the nature of these creatures are shown to be incorrect. They have actually retained some of their humanity. There is a very important moment between the Alpha Male and Neville. Neville turns the infected female he has been treating over to the Alpha Male, and there is this little love moment between the two of them. Once the infected have retrieved the captured female, they spare Neville's life. He then looks at the photos of the infected he has experimented on and killed and realizes that he is the monster; the infected think of him as someone who hunts down their people and kills them.

The final shot follows Neville, Anna, and Ethan as they cross the remnants of the George Washington Bridge in hopes of finding other survivors, accompanied by a recording from Anna telling possible survivors that there is hope.

Fill out the following questionnaire about the movie. In the last column, give a reason why the producer might have added each additional question.

Questions	Response	Reason for the Question
1) Which ending did you prefer?	Ending 1 __ Ending 2 ___	**********
2) Is this the kind of movie you would go to see in a theater?	Yes ___ No ___	
3) What is your age group?	16 or under ____ 17 or older ____	
4) How many movies, on average, do you see a year?	11 or less ____ 12 or more _____	
5) What is your gender?	Male ____ Female ____	

In your opinion, are there any other questions that should have been included on the questionnaire?

Name _____ Date _____

5. Movie Preferences: Analyzing Data

Use the form below to record the responses from the movie questionnaire.

Possible Choices	Male Responses	Female Responses	Totals M	Totals F	Totals All
1. Ending #1					
Ending #2					
2. Yes					
No					
3. Under 13					
13 or older					
4. 11 or fewer					
12 or more					

1. Now that you have collected all the responses on your record sheet, you need to present the data to the producer in a useful manner.
   (a) Count up all the male responses to each choice and enter the totals under the *M* column.
   (b) Count up all the female responses to each choice and enter the totals under the *F* column.
   (c) Count up all the responses for each choice regardless of gender. Put these totals under the *All* column.
2. Which ending did more of the class prefer? Is this result significant? _____
3. If the movie producer decides to target teenage girls with this movie, which ending should she use for the film? Is this result significant? _____
4. Did more males or more females indicate that they would see this kind of movie in a theater? Is this result significant? _____
5. Is the movie producer more likely to use your data if most of the teenagers in your class say they see more than a dozen movies per year? Why or why not? _____

Movie Preferences: Analyzing Data Continued

There are many different ways to organize data. Deciding which method to use is based upon what you want to learn from the information you have collected. Use the following situations to organize the information in other ways to help you solve the problems.

6. How would you organize the data if you were interested in learning more about the people who watch 12 or more movies each year?

Possible Choices			Totals		

7. Which ending should you use if you want to target this group?

8. How would you organize the data is you wanted to make a movie for younger viewers?

Possible Choices			Totals		

9. What ending would you choose for this group?

Day 5 – Patterns in Data

Q: How are two variables related?

Q: How can this relationship help us make predictions?

One of the most practical uses of math is to use what we know to make predictions about what we don't know. For example, suppose you wanted to know how much it will cost to buy the average house in ten years when you are ready to make that purchase. There is no possible way you can go ten years in the future, but you can use the information today to find patterns and predict the future.

To complete this task, we first have to find the patterns and describe them with graphs, lines, and equations. Then we can use these tools to make the predictions we are trying to find. We can use the same tools to make predictions about sports, financial situations, weather, astronomy, medicine, work and school, grades, growth and decay, and really just about anything.

When making predictions, it is also important to measure how sure you are about the patterns you have discovered. In other words, how strong is the correlation between the two variables? If the relationship is strong, you can place more faith in the power of your prediction.

Important Vocabulary:

1) Bivariate Data

2) Correlation Coefficient

3) Extrapolation

4) Interpolation

5) Linear Regression

6) Scatter Plot

7) Slope

8) Slope – Intercept Form

9) Trend Line

10) Univariate Data

11) Variable

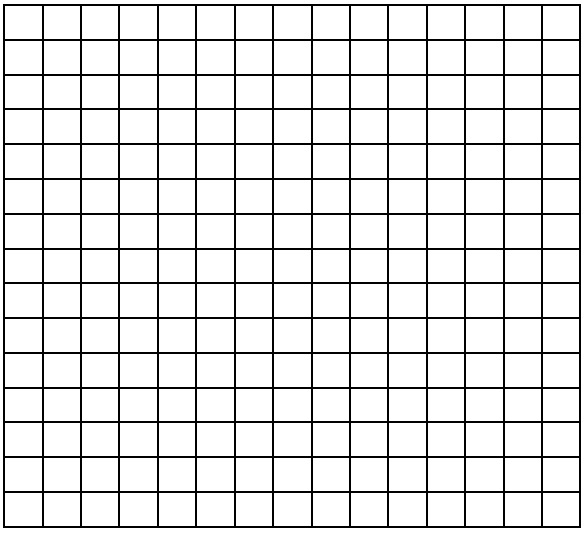
12) Y-intercept

Day 6 - Scatter Plots and Trend Lines

For the remainder of activities we will be doing in this unit, the data is bivariate. This means that there are two different variables or things that can change in value. This is opposed to univariate data that only has one variable. An example of univariate data would be your math average. Bivariate data could be your height AND your math average.

With bivariate data, we often are looking to see if there is a relationship between the two variables. The easiest way to SEE the relationship is with a graph such as a scatter plot. A scatter plot is a graph that displays bivariate data as points on a graph. We will make one to remind you of what they look like and what they can tell us about the data.

The following table shows the relationship between the size of a hot air balloon and the number of candles it would take it to stay afloat. Draw a scatter plot of this information



Ex)

Cubic Feet (Size of Balloon)	Number of Candles
1	7
5	20
10	32
20	50
30	66
40	80
50	93

The first list is always the x-values and should be graphed on the x-axis.
The second list is always the y-values and should be graphed on the y-axis.

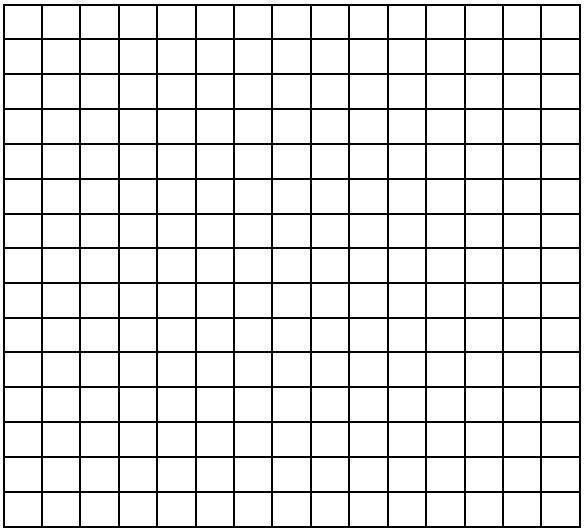
How would you describe the pattern in the scatter plot above?

We can draw a trend line or line of best fit to model this relationship. For now, we will get the trend line by drawing the line that appears to be the line that comes closest to all of the points in the scatter plot.

Draw the trend line on the scatter plot above.

1) Sketch a scatter plot of the following situation

Hours of Work Per Week	School Average
27	72
4	94
0	85
10	86
15	78
5	90



2) Draw the trend line on this scatter plot. How is this line different than the trend line for the hot air balloon size vs. number of candles graph?

3) In which graph do the points come closer to the line of best fit? What does this tell you?

4) Other than describing a relationship, what does the trend line allow us to do?

5) Use your first trend line to predict how many candles are needed to lift a balloon that is 60 cubic feet?

6) Use your second trend line to predict what school average will be obtained if a student works 20 hours per week?

7) Use your first trend line to predict how large a balloon is if it requires 75 candles to lift it?

8) Use your second trend line to predict how many hours a week a student is working if they have an 80 average?

9) Use your first trend line to predict how many candles are needed to lift a balloon that is 250 cubic feet?

10) What makes this last question difficult?

11) Did you have slightly different answers than some of your classmates? Why do you think this happened?

Both of these problems will be solved once we learn how to do regressions.

Day 7 - Correlation Coefficient

In the last lesson, you made some predictions based upon your scatter plots and trend lines. What would make you more likely to put faith in that prediction?

There is actually a mathematical measurement called the correlation coefficient that measures how strong the relationship or correlation is between the two variables. In this section we are going to learn how to read that number, so when the calculator gives us the exact value we will know what it means.

Symbol:

Range:

Sign of r:
+:
0:
-:

Value of r:

IMPORTANT NOTE!

Just because two things are correlated or have a linear relationship, it does not mean we can jump to the conclusion that one thing causes another. We do not know which one starts the reaction or even if there is not a third factor that causes both of them.
Example: Hours of Work (x) per week vs. School Average (y)
A. X impacts Y
B. Y impacts X
C. Another factor (Z) impacts both of them

Draw a scatter plot with a correlation coefficient of:

r=1.0

r=.5

r=0

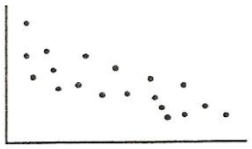
r=-.5

r=-1.0

Which scatter plot represents data with a stronger negative correlation?

Name: _____

1) Based on the scatter plot shown, it would be reasonable to conclude that the correlation coefficient for x and y is

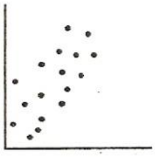


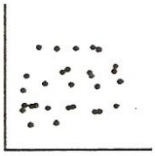
A) less than 0
B) greater than .5
C) between 0 and .5
D) equal to 0

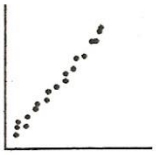
2) Sketch a scatter plot that represents a strong positive correlation.

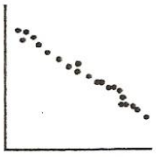
3) Sketch a scatter plot that represents a weak negative correlation.

4) Which scatter plot represents a weak positive correlation?

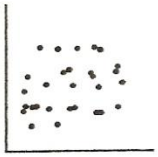
A) 

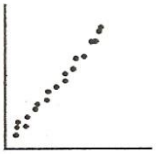
C) 

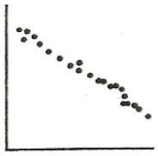
B) 

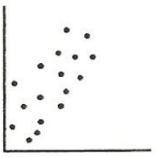
D) 

5) Which scatter plot represents a strong positive correlation?

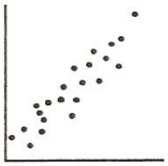
A) 

C) 

B) 

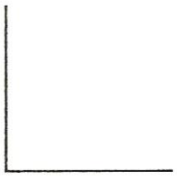
D) 

6) Based on the scatter plot shown, it would be reasonable to conclude that the correlation coefficient for x and y is



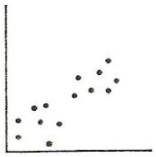
A) greater than 0
B) between -.5 and 0
C) equal to 0
D) less than -.5

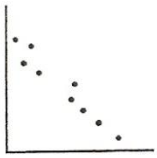
7) Sketch a scatter plot that represents a strong negative correlation.

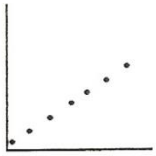


8) Which of the following scatter plots has a correlation coefficient equal to approximately .5?

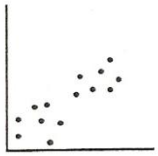
A) 

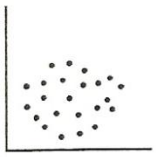
C) 

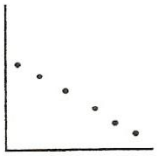
B) 

D) 

9) Which scatter plot most likely has a correlation coefficient equal to 1?

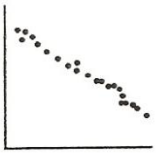
A) 

C) 

B) 

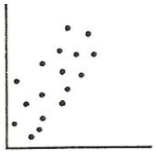
D) 

10) Which scatter plot represents a weak negative correlation?

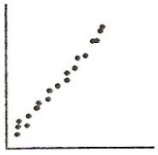
A) 

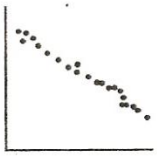
C) 

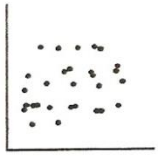
B) 

D) 

11) Which scatter plot represents a strong negative correlation?

A) 

C) 

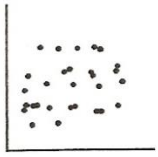
B) 

D) 

12) For which scatter plot is the correlation coefficient approximately 0?

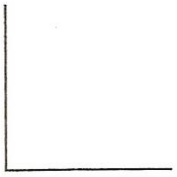
A) 

C) 

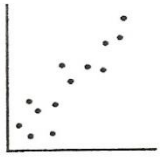
B) 

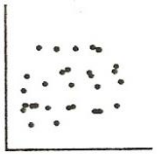
D) 

13) Sketch a scatter plot that represents a weak positive correlation.

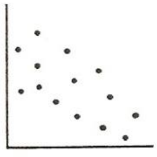


14) For which scatter plot might $r = -.98$?

A) 

C) 

B) 

D) 

Day 8 - Scatter Plots on the Graphing Calculator

You can also use the graphing calculator to make a scatter plot and find the exact equation of the trend line.

Using the information from the table,

1) Go to STAT #1 (Edit).
2) You will see a series of lists, L_1, L_2, L_3, *etc.*, in which you can enter lists of numbers.

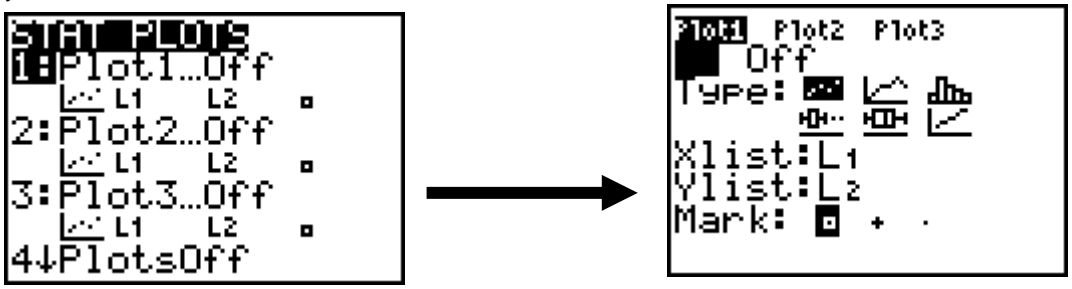
If there is information in a list, you can clear it by going up to the title (L_1, L_2, L_3, *etc.*) and pressing CLEAR ENTER.

3) **Enter the values for the independent variable (x) in L_1 and the values for the dependent variable (y) in L_2.**

This may seem time consuming, but the values need to be in the lists for the line of best fit calculations we will be learning anyway, so it is not an extra step.

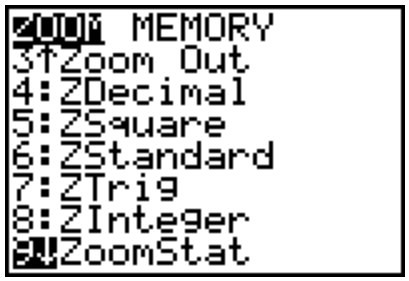
Cubic Feet (Size of Balloon)	Number of Candles
1	7
5	20
10	32
20	50
30	66
40	80
50	93

4) Next, go to **STATPLOT**: 2nd [ ]
5) Choose Plot #1
6) Turn Plot 1 on:



7) If you began with your calculator reset no other changes are necessary. The first picture for "type" is the scatter plot type and your data should be in lists 1 and 2. If you are using lists other than these change them before going on.

8) **Press** Zoom **and choose** 9:ZoomStat.



The calculator will choose the best viewing window based on the type of plot and the data in the lists. It will also take you immediately to the graphing window.

Sketch scatter plot here:

Window:

9) **After sketching the scatter plot, press the WINDOW button and label the Xmin, Xmax, Ymin, and Ymax values on your graph.**

What do you think the correlation coefficient is for the data in the graph?

Summary:

1) STAT 1 (Edit)
2) (If necessary) Clear Lists (Cursor on Title, CLEAR ENTER)
3) Enter your lists into L_1 and L_2.
4) STATPLOT 1 (Plot 1)
5) Turn Plot 1 on.
6) Zoom 9 (Zoom Stat)
7) Sketch scatter plot and label graph (max, min)

For the next two problems, create a scatter plot on your calculator and by hand.

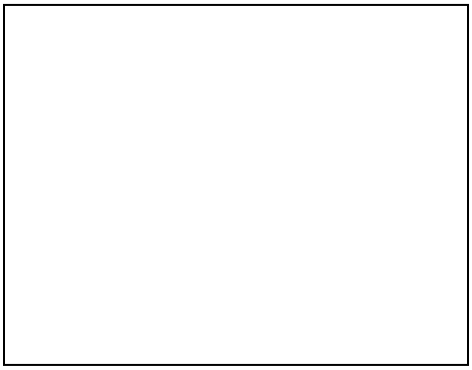
A. Sketch a picture of the scatter plot you see on your calculator. Write the window (X-min, X-max, Y-min, Y-max).

B. Approximate the correlation coefficient between the two lists of information.

C. Draw a trend line on the scatter plot.

1. The chart below shows the amount of active ingredients, *m*, in milligrams, of a blood pressure medication present in a person's bloodstream *h* hours after the medication is consumed.

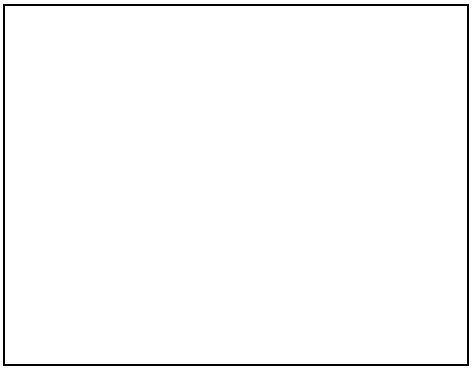
h	*m*
.5	112
1	86
1.5	53
2	41
2.5	27
3	19



r=_____

2. Eight students SAT scores were recorded as shown.

Student	*Verbal (V)*	*Math (M)*
Mary	560	680
Joe	450	520
Sam	520	560
Bonnie	780	750
Tom	700	680
Frank	480	540
Sue	500	550
Alice	640	700



r=_____

Day 9 - Linear Regressions

(round values to the nearest hundredth)

In the last lesson, you learned how to make scatter plots on the graphing calculator. The calculator also finds the equation of the trend line. This equation that best fits the data is called a linear regression.

Why would the linear regression be better than a hand drawn trend line for making predictions?

1.

2.

The calculator will also find the correlation coefficient.

To review, the correlation coefficient tells us _____ so that we know _____.

To set the calculator to display the **correlation coefficient**,

a. Press 2nd 0, (Catalog).
b. Press the button with the green D over it (x^{-1}) and scroll down to Diagnostics On.
c. Press Enter Enter. It should say DONE.
d. You only need to do this the first time you use the calculator for regressions.

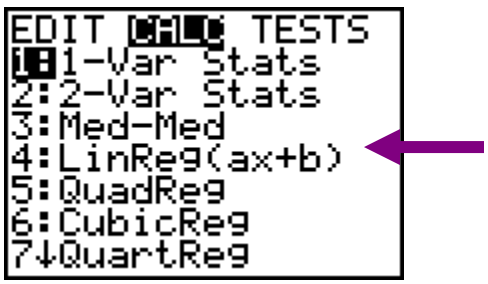
Finding the regression equation

This process begins the same way as making a scatter plot.

1. Go to STAT 1.
2. Clear the lists if necessary (Cursor on title, Clear Enter.
3. Enter the two lists of information into L_1 and L_2.

New Information:

4. Go to STAT and arrow over to CALC. The screen should look like this:



Cubic Feet (Size of Balloon)	Number of Candles
1	7
5	20
10	32
20	50
30	66
40	80
50	93

5. Choose 4, LinReg (ax+b). Press Enter. (Do not choose 8; it is different).

The calculator will give you
y=ax+b
a=_____
b=_____
$r^2 =$_____
r=_____

6. Round you're a and b values to the nearest hundredth, and fill them into the given equation, y=ax+b. This is the **linear regression** or the equation of the **line of best fit**.

7. Write down r, the correlation coefficient, to the nearest hundredth.

Let's see what your regression looks like. This is not a necessary step in making predictions, but it will show us how the linear regression really is the equation of the trend line.

a) Make a scatter plot of the two lists you entered. (See the summary in the previous section for a reminder of how to do this).
b) Go to Y=. Enter the regression equation in Y1.
c) Press Graph.

Summary:
Stat 1
Enter information in lists
Stat → Calc 4
Enter
Write your equation with a and b plugged in.
Write down your correlation coefficient, r.

Compare this with your trend line the first time we used this information. This should be on page 24. Even though these look the same, remember that the linear regression equation is more accurate and allows us to predict values outside what we see on the graph.

1. The following data comes from a survey about number of hours of work each week and school averages. Find the regression equation for this information.

State the correlation coefficient. Why is it negative?

Hours of Work Per Week	School Average
27	72
4	94
0	85
10	86
15	78
5	90

2. The chart below shows the amount of active ingredients, *m*, in milligrams, of a blood pressure medication present in a person's bloodstream *h* hours after the medication is consumed.

h	*m*
.5	112
1	86
1.5	53
2	41
2.5	27
3	19

Find the linear regression for the data.

What is the correlation coefficient? What does it tell you about the data?

3. Eight students SAT scores were recorded as shown.

Student	*Verbal (V)*	*Math (M)*
Mary	560	680
Joe	450	520
Sam	520	560
Bonnie	780	750
Tom	700	680
Frank	480	540
Sue	500	550
Alice	640	700

Find the linear regression for the data.

What is the correlation coefficient? What does it tell you about the data?

4. The shoe sizes and heights of fourteen people are shown in the chart below.

Male Shoe Size	Height in Inches
11	68
8	60
10	64
13	78
12.5	74
14	78
11	74
7	60
10	70
9	64
11	72
13	74
12	72
13.5	78

Find the linear regression for the data.

What is the correlation coefficient? What does it tell you about the data?

Day 10 - What does the linear regression tell us about the data?

(Note: We were rounding the answers for our regressions to the nearest hudredth. To understand the slope and y-intercept in this lesson, it might make more sense to change the rounding so that the numbers are more familiar. We will do so on some of the examples as needed.)

To answer this question, we will look at the hot air balloon example first. Write down the equation you found from the previous activity here. _____

- In Math 1 and Math 2 you learned that the equation of a line was _____. This was called slope-intercept form. The calculator is really using this form except their equation is _____. "a" and "m" are just different letters for the same variable.
- In this equation"b" is called the _____. It tells us where the line crosses the y-axis or what the y-value is when x is zero. We want to know more. What does it tell us about our data?
- The y-intercept of the equation you found is _____. This means that the graph crosses the y-axis at _____.

Equation: x=0 y=_____

In this problem, x is the cubic feet of the balloon and y is the number of candles. We can use this information to write a sentence interpreting the y-intercept.

What does the y-intercept tell you about the data?

The _____(y-variable) is _____(y-intercept) when _____ (x-variable) is zero.

The grammar in this sentence might need to change depending on the variables, but the **basic setup will always be the same**. Also, this may not make sense practically, but remember the regression is written to fit the points in your table.

- "m" in y=mx+b or "a" in y=ax+b is the _____ .

The **slope** is the rate of change between the two variables. It can be calculated by $\frac{\text{change in Y}}{\text{change in X}}$. The slope in the hot air balloon example is _____. Written as a fraction this would be _____. (If the slope is negative, always put the negative on the top.)

Remember, x is the cubic feet of the balloon and y is the number of candles. We can use this information to write a sentence interpreting the slope.

What does the slope tell you about the data?

As the _____(x-variable) increases by 1, the _____(y-variable) increases/decreases by _____ (slope).

1. For the hours of work per week vs. grade in school data,

   a. Copy the equation from the previous activity. _____

   x is _____

   y is _____

   b. What is the y-intercept? What does it tell you about the data?

   c. What is the slope? What does it tell you about the data?

2. For the hours after medicating vs. medication in the body data,

   a. Copy the equation from the previous activity. _____

   x is _____

   y is _____

   b. What is the y-intercept? What does it tell you about the data?

   c. What is the slope? What does it tell you about the data?

3. For the verbal SAT scores vs. Math SAT scores data,

   a. Copy the equation from the previous activity. _____

   x is _____

   y is _____

   b. What is the y-intercept? What does it tell you about the data?

   c. What is the slope? What does it tell you about the data?

4. For the shoe size vs. height data,

   a. Copy the equation from the previous activity. _____

   x is _____

   y is _____

   b. What is the y-intercept? What does it tell you about the data?

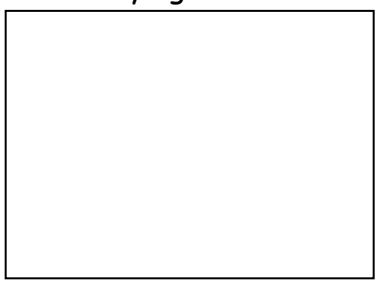
   c. What is the slope? What does it tell you about the data?

Practice with Understanding Regressions

(round to the nearest hundredth)

Hours Spent Studying (x)	.5	1	1.5	2	2.5
Average Test Score (y)	58	78	85	89	92

1. The table gives the hours spent studying and the resulting grade on the Math 3 final.



a. Create a scatter plot of the situation on your calculator. Label your window.

b. What is the linear regression equation?

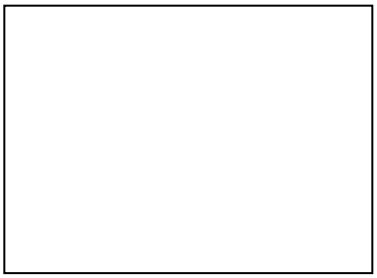
c. What is the correlation coefficient. What does it tell you about the data?

d. What is the y-intercept of the regression line? Write a sentence describing what this tells you about the data.

e. What is the slope of the regression line? Write sentence describing what this tells you about the data.

Distance in feet	Weight in Pounds
2	116
3	58
4	36
5	24
6	18

2. On a certain planet the weight, w, in pounds, of an object and the distance, d, in feet, this object is from the surface of the planet were recorded and are shown in the accompanying table.



a. Create a scatter plot of the situation on your calculator. Label your window.

b. What is the linear regression equation for the data?

c. What is the correlation coefficient. What does it tell you about the data?

d. What is the y-intercept of the regression line? Write a sentence describing what this tells you about the data.

e. What is the slope of the regression line. Write a sentence describing what this tells you about the data.

Day 11 - Making Predictions

(Round regression values to the nearest hundredth. Round predictions to the nearest whole number.)

Now that you have learned how to find regression equations, you can use your skills to solve problems you might encounter in business, psychology, history, sports, and a number of other areas.

The two reasons we find regression equations is to describe the pattern between two variables (what we did in the last lesson) and to make predictions by using the collected data to find unknown values.

There are two types of prediction. Interpolation predicts values for the two variables that are between the maximum and minimum data points. Extrapolation predicts values outside the data set. Extrapolation is hard to do just with a graph and a trend line because the information you are trying to find is usually outside the graph.

Example: Consider the following table of information about the chirps per second for crickets and the corresponding temperature in Fahrenheit. First we will begin with what we have already learned.

CHIRPS PER SECOND	20	16	20	18	17	16	15	17	15	16
TEMPERATURE	89	72	93	84	81	75	70	82	69	83

1) Sketch a scatter plot for the information using your calculator.

2) Sketch the trend line on the same graph as your scatter plot.

3) Find the linear regression for this information.

4) What is the correlation coefficient? What does this tell you about the data?

5) What is the slope? What does it tell you about the data?

Using your regression equation, you can predict the temperature based on the number of cricket chirps. You can also predict the number of cricket chirps if you know the temperature.

6) Use your linear regression to make a prediction: If a cricket is chirping 19 times per second, what is the temperature?

Equation: x= y=

7) Use your linear regression to make a prediction: If the temperature is 91 degrees, how many times do you think a cricket would chirp?

Equation: x= y=

Both of the past two answers you could have estimated with the trend line because they found values within the data. Remember this is called _____. The answers to the next couple of questions will probably not appear on your graph which is why the linear regression is so necessary.

8) Use your linear regression to make a prediction: If a cricket is only chirping 5 times per second, what is the predicted temperature?

Equation: x= y=

9) Use your linear regression to make a prediction: If the temperature is 55 degrees, how many times do you think the cricket will be chirping?

Equation: x= y=

Remember, these are predictions and not hard facts. Still, you can figure out how much faith you should put in the prediction. The _____ _____ can help you know how much to trust it. The closer to ____ or ____ this value is, the safer the prediction.

Here is another example comparing the weight of an object hanging from a spring and the length of the spring or how much the spring gets stretched out.

Weight (grams)	Length (cm)
10.36	8
12.13	12
14.35	16
16.21	20
18.52	24

10) Find the linear regression for this information.

11) What is the correlation coefficient? What does this tell you about the data?

12) What is the slope? What does it tell you about the data?

13) Use your linear regression to make a prediction: If a weight of 30 grams is attached, how long would the spring be, to the nearest tenth of a centimeter?

14) Use your linear regression to make a prediction: If a weight of 100 g was attached, what would be the spring length to the nearest tenth of a centimeter?

15) Use your linear regression to make a prediction: If the spring is 18 cm long, what is the weight on the spring to the nearest gram?

16) Use your linear regression to make a prediction: If the spring is 70 cm long, what is the weight on the spring to the nearest gram?

17) What real life applications would this data have?