Connecticut Common Core Algebra 1 Curriculum

Professional Development Materials

Unit 5 Scatter Plots

Contents

*** These items appear only on the password-protected web site.**

Unit 5: Scatter Plots & Trend Lines
(4 weeks)

UNIT OVERVIEW

Students will begin the unit by exploring measures of central tendency and spread and displays of one-variable data including, dot plots, histograms, and box-and-whisker plots. They will use the five number summary to create box-and-whisker plots and identify outliers with the 1.5 X IQR rule. They will be introduced to using the STAT menu on the graphing calculator.

In investigation two, students will be introduced to scatter plots and trend lines. They will fit a trend line to a scatter plot by hand and find its equation.. They will use the equation of the trend line to make predictions by interpolating or extrapolating. The students will develop a deeper understanding about the meaning of the slope and intercepts in context. These ideas are revisited in subsequent inestigations.

In Investigation three, students will continue to explore trend lines and predictions. They will use technology (either a graphing calculator or a spreadsheet) to calculate the linear regression equation and to find the correlation coefficient. The students will be able to interpret the meaning of the correlation coefficient and explain the difference between correlation and causation.

During investigation four, students will perform experiments in which they collect and analyze data using linear models. In this investigation, students will apply their knowledge from the previous two investigations. In this investigation the teacher will get the class prepared and organized, and then will walk around to observe and ask questions.

Investigation five, students will work with data sets that contain outliers to identify the influence that outliers have on the calculation and interpretation of the slope, y-intercept, linear regression equation, and correlation coefficient.

In the last investigation students will explore situations in which the data represents more than one trend, will fit a line to each section of the data set, and will use the lines to make predictions. In this way they will be introduced to piecewise linear functions.

Essential Questions

- How do we make predictions and informed decisions based on current numerical information?
- What are the advantages and disadvantages of analyzing data by hand versus by using technology?
- What is the potential impact of making a decision from data that contains one or more outliers?

Enduring Understandings

- Although scatter plots and trend lines may reveal a pattern, the relationship of the variables may indicate a correlation, but not causation**.**

Unit Content

Investigation 1: One Variable Data (three days)
Investigation 2: Introduction to Scatterplots and Trend Lines (two day)
Investigation 3: Technology and Linear Regression (two days)
Investigation 4: Explorations of Data Sets (four days)
Investigation 5: Exploring the Influence of Outliers on Trend Lines (two days)
Investigation 6: Piecewise Functions (two days)
Performance Task: Linearity is in the Air — Can You Find It? (three days)
(Note: The performance task should begin early in the unit to give students time to collect data.)
Suggested Time line: Day 1- Following investigation 3 - brainstorming and research
Day 2- Following investigation 5 - collecting and analyzing data
Day 3- Following investigation 6 – work day
Review (one day)
End-of-Unit Test (one day)

Common Core Standards

Mathematical Practices #1 and #3describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. Practices in bold are to be emphasized in the unit.

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. **Model with mathematics.**
5. **Use appropriate tools strategically.**
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

Standards Overview

- Analyze functions using different representations
- Summarize, represent, and interpret data on a single count or measurement variable
- Summarize, represent, and interpret data on two categorical and quantitative variables
- Interpret linear models

Standards with Priority Standards in Bold

8-SP 1. Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

8-SP 2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

8-SP 3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*

S-ID 2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S-ID 3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

S-ID 6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data.

c. Fit a linear function for a scatter plot that suggests a linear association.

S-ID 7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

S-ID 8. Compute (using technology) and interpret the correlation coefficient of a linear fit.

S-ID 9. Distinguish between correlation and causation.

Vocabulary

Boxplot
causation
correlation
correlation coefficient
data
data set
dependent variable
distribution
domain
extrapolation
graphical representation
histogram
independent variable
interpolation
inter quartile range (IQR)
line of best fit
linear regression
linear relationship/model
mean (average)
median
measures of central tendency
mode
nonlinear relationship/model
ordered pair
outlier
piecewise function
prediction
regression equation
scale
scatter plot
skewed distribution
slope
trend line
variable
x-intercept
y-intercept

Assessment Strategies

Performance Task: Linearity is in the Air — Can You Find It?
During the unit, have students develop a hypothesis about a real-world ‘nearly’ linear situation interesting to them, find relevant data, model the data, analyze the mathematical features of the model, and make and justify a conclusion. By the end of the unit, all students will present their findings to the class. NOTE: The performance task should be spread out over the entire unit.

Other Evidence (Formative and Summative Assessments)

- Exit Slips
- Class work
- Homework assignments
- Math journal
- Unit 5 Test

Hurricanes

Each year tropical storms that form in the Atlantic Ocean are given names. The first named storm starts with “A”, the second starts with “B”, and so on. A tropical storm becomes a hurricane if its wind speed reaches 74 miles per hour.

2005 was the most active year for hurricanes on record. In July of 2005, Cindy was the first tropical storm to become a hurricane. In August of 2005, Katrina made headlines worldwide as it wreaked havoc on the city of New Orleans. This chart shows the maximum wind speed, in miles per hour, for each of the fifteen Atlantic Ocean hurricanes of 2005.

2005 Atlantic Ocean Hurricanes		
Name	**Dates**	**Max Wind Speed (mph)**
Cindy	7/3 - 7/7	75
Dennis	7/4 - 7/13	150
Emily	7/11 - 7/21	160
Irene	8/4 -8/18	105
Katrina	8/23 - 8/30	175
Maria	9/1 - 9/10	115
Nate	9/5 - 9/10	90
Ophelia	9/6 - 9/17	85
Philippe	9/17 - 9/23	80
Rita	9/18 - 9/26	180
Stan	10/1 - 10/5	80
Vince	10/8 - 10/11	75
Wilma	10/15 - 10/25	185
Beta	10/26 - 10/31	115
Epsilon	11/29 - 12/8	85

In previous courses you learned about three statistics called **measures of center**. These statistics are described in the box at the right.

1. For the maximum wind speeds, find the:

   a. mean

   b. median

   c. mode

Three Measures of Center
The **mean** is the average that you're used to, where you add up all the data values and then divide by the number of values.
The **median** is the "middle" value in a list of numbers. To find the median, first list the numbers in numerical order. Then, if the number of values is odd, the median is the number in the middle. If the number of values is even, the median is the mean of the two numbers in the middle
A **mode** is a value that occurs most often. If no number is repeated, then there is no mode for the list. Some lists of numbers may have more than one mode.

2. For these data, which measure of center is larger, the mean or the median? Why do you suppose this is?

3. What difficulty did you have answering question 1(c) above? What does that tell you about the mode of a set of values?

Hurricane Categories

Hurricanes are classified based on their maximum wind speed according to the Saffir-Simpson Hurricane Scale shown in this chart.

Saffir-Simpson Hurricane Scale

Category	Max Wind Speed (mph)
1	74–95
2	96–110
3	111–130
4	131–155
5	155+

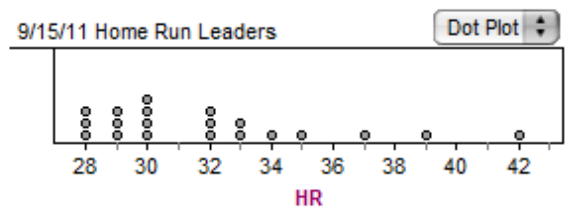
4. Use the Saffir-Simpson Hurricane Scale to categorize the hurricanes in the chart below.

2005 Atlantic Ocean Hurricanes			
Name	**Dates**	**Max Wind Speed (mph)**	**Category**
Cindy	7/3 - 7/7	75	1
Dennis	7/4 - 7/13	150	4
Emily	7/11 - 7/21	160	
Irene	8/4 -8/18	105	
Katrina	8/23 - 8/30	175	
Maria	9/1 - 9/10	115	
Nate	9/5 - 9/10	90	
Ophelia	9/6 - 9/17	85	
Philippe	9/17 - 9/23	80	
Rita	9/18 - 9/26	180	
Stan	10/1 - 10/5	80	
Vince	10/8 - 10/11	75	
Wilma	10/15 - 10/25	185	
Beta	10/26 - 10/31	115	
Epsilon	11/29 - 12/8	85	

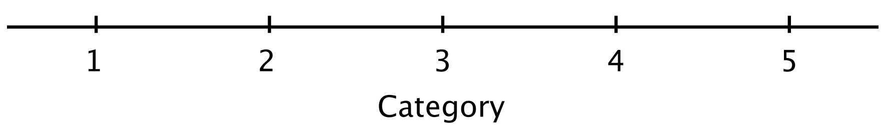
5. Find the mean, median, and mode of the category data.

6. Compare your results from questions 1 and 5. Describe any patterns you observe.

7. Display the category data with a dot plot. The categories are shown on the number line. For every hurricane, place a dot above the appropriate location on the number line. (An example of a dot plot is given at the right. This dot plot shows the major league home run leaders in September 2011.)

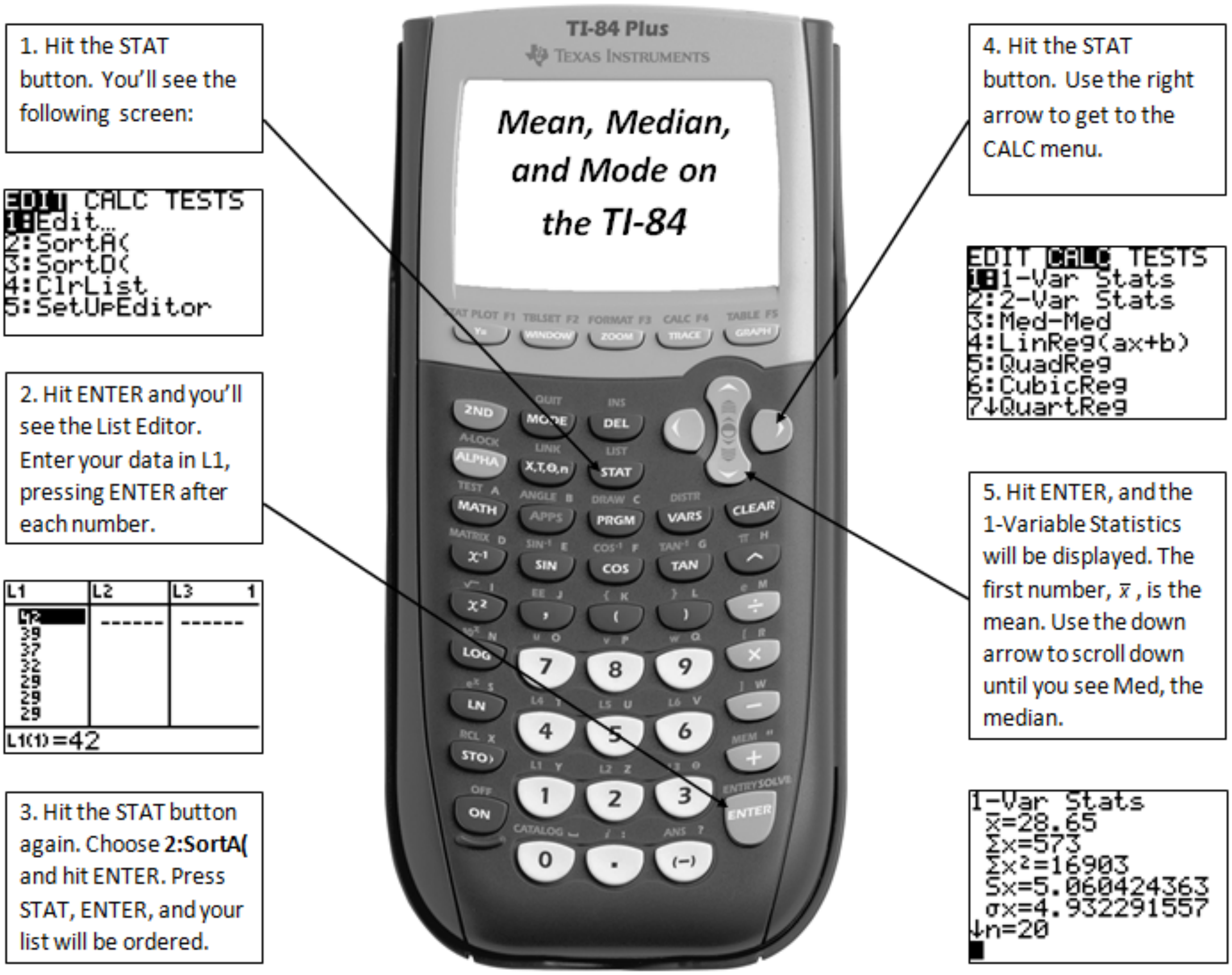


Distribution of Hurricane Categories in 2005



8. On the dot plot locate the three measures of center for the category data. Place an asterisk (*) on the mean, put a square around the median, and put a circle around the mode.

9. Write three sentences about the conclusions you can make from the dot plot and your analysis of the data.



Using Technology to Calculate Statistics

Enter the maximum wind speed data into L1 as shown above. (Press STAT then select EDIT. Then select STAT → CALC and 1-Var Stats. Find the statistics for the list L1.)

The screen on the left will appear:

$\bar{x}$ =	n =
Σx =	minX =
Σx^2 =	Q1 =
Sx =	med =
σx =	Q3 =
n =	max =

10. The first number to appear is the mean. (It appears as $\bar{x}$, which is called "*x* bar.") Does this number agree with the mean you calculated in question 1(a)?

11. For now, ignore the four numbers below the mean. The last number on the screen is *n*. It tells you how many data values you have. What is the value of *n*? Is this correct?

    Notice the arrow to the left of *n*. It suggests that you can scroll down. Scroll down until you see the statistics shown in the image at the bottom right of the previous page.

12. On this screen you will find the minimum (Min), the maximum (Max) and the median (Med). Do these results make sense? Explain.

13. The **range** of a dataset is the difference between the maximum and minimum values. The range measures the amount of spread in a dataset.

$$Range = Maximum - Minimum$$

    Find the range for the maximum wind speed data.

Histograms

One way to display one-variable data is with a histogram. A **histogram** is like a dot plot, except that the values are grouped into intervals called *bins*. To create a histogram, we must have a **frequency table**. A frequency table contains a set of bins and the number of data values contained in each bin. The number of data values in a bin is called the *frequency* of the bin. We can draw a graph and represent each bin with a bar. The height of a bar shows the frequency of the bin.

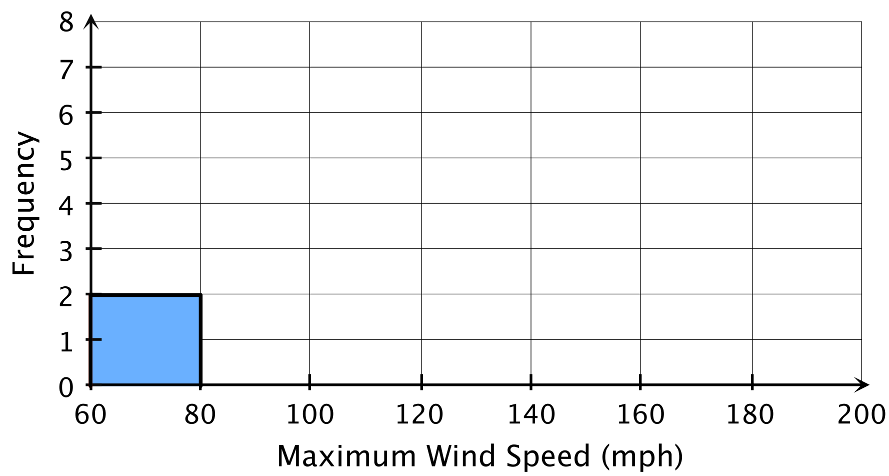
It is important that all the bins are the same width. All data values must be placed in one of the bins.

14. A frequency table for the maximum wind speeds is shown below. Each bin has a width of 20 miles per hour. We must identify the number of data values in each bin. The data values contained in the first two bins and the frequency of the first two bins have already been identified. Fill in the rest of the table.

Bin	Maximum Wind Speeds	Frequency
$60 \le x < 80$	75, 75	2
$80 \le x < 100$	90, 85, 80, 80, 85	5
$100 \le x < 120$		
$120 \le x < 140$		
$140 \le x < 160$		
$160 \le x < 180$		
$180 \le x < 200$		

15. Use the data in the table to make a histogram of the maximum wind speeds. Notice the bins are on the *x*-axis and the frequencies are on the *y*-axis. The first bar is drawn for you.

Maximum Wind Speeds in 2005



16. Now use your calculator to make the same histogram. Follow these steps:

- In the Stat Plot menu, select Plot 1, turn it on, and select the histogram icon. Xlist should be L1 and Freq = 1.
- In the Window menu set Xmin = 60, Xmax = 200, Xscl = 20, Ymin = 0, Ymax = 8, Yscl = 1.
- Press Graph. Describe what you see.

17. Now draw a histogram with a narrower bin width of 10 miles per hour. To change the bin width to 10, go to the Window menu, and set Xscl = 10. Then press Graph. What do you notice? How are the two histograms alike? How are they different?

Hurricanes in 2012

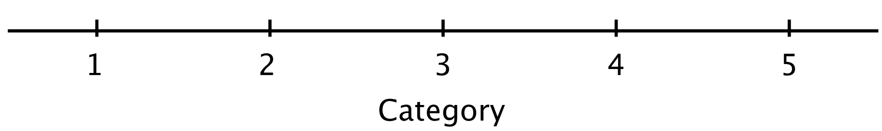
Now let's do the same analysis with the hurricanes of 2012. The data in the table below are incomplete. Your teacher will give you the complete data set or you can find it yourself at Wikipedia.org/wiki/2012_Hurricane_Season.

2012 Atlantic Ocean Hurricanes			
Name	**Dates**	**Max Wind Speed (mph)**	**Category**
Chris	6/19-6/22	75	
Ernesto	8/1-8/10	85	
Gordon	8/15-8/20	110	
Isaac	8/21-9/1	80	
Kirk	8/28-9/2	105	
Leslie			
Michael			
Nadine			
Rafael			
Sandy			

18. Find the mean, median, and range for the maximum wind speeds of the 2012 hurricanes.

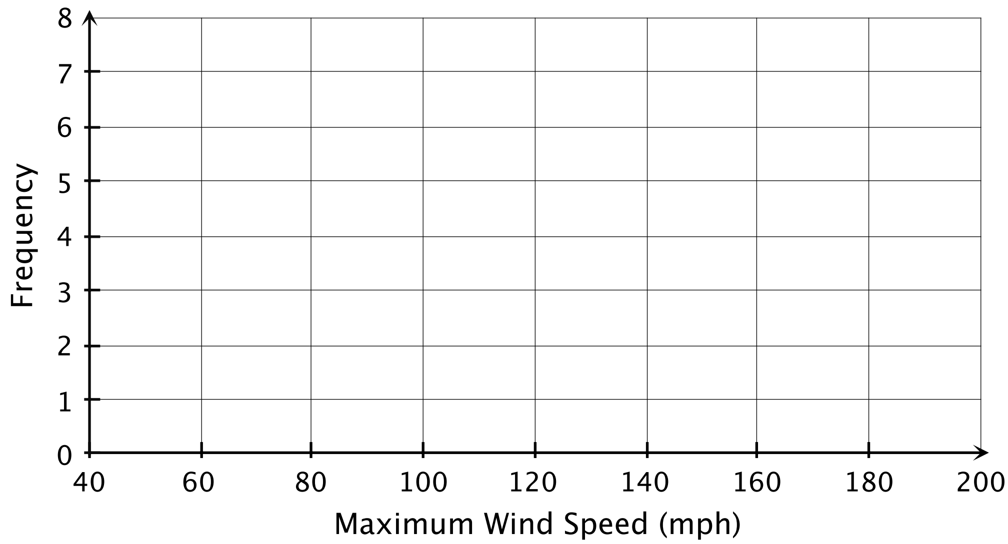
19. Make a dot plot of the categories of the 2012 hurricanes.

Distribution of Hurricane Categories in 2012



20. Make a histogram for the maximum wind speeds of the 2012 hurricanes.

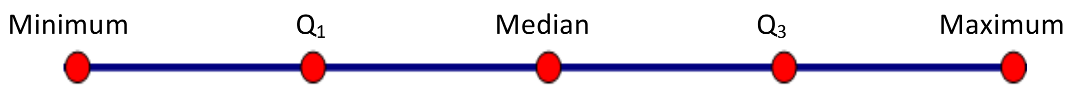
Maximum Wind Speeds in 2012



21. Write a paragraph contrasting the 2005 and 2012 hurricane seasons. In your paragraph refer to the tables, the statistics (mean, median, and range), and your data displays (dot plots and histograms).

The Five-Number Summary

We often are interested in how much *spread* there is in a data set. The spread, or variability, of data describes how far apart the data values are. A set of statistics that help us see the amount of spread in a data set is the **five-number summary.** The five-number summary consists of the minimum, Q_1, median, Q_3, and maximum. Q_1 and Q_3 are the first and third quartiles. The median equals Q_2. Quartiles divide a data set into four quarters.



To create the five-number summary of a data set, start by ordering the data set into increasing or decreasing order. Then, find the median (middle) of your data set. The median divides the data set into two halves. To find the quartiles, find the median of the lower half and find the median of the upper half.

Example: Below are the arm-spans (in cm) of 15 Algebra I students:

148 152 152 152 154 154 154 162 163 164 164 170 172 180 181

Solution: The data values are already ordered. There are an odd number of values, so the median is the middle number in the list. The lower half and upper half are shown in boxes. Each half has 7 data values. The median of the lower half is 152, and the median of the upper half is 170.

[148 152 152 152 154 154 154] 162 [163 164 164 170 172 180 181]

Median of the lower half (Q_1)

Median of the upper half (Q_3)

Middle of the data set (Median = Q_2)

1. Use the statistics above and the minimum and maximum to complete the table.

	minimum	Q_1	median	Q_3	maximum
Arm-spans					

2. Find the five-number summary for the maximum wind speeds of the named hurricanes in 2005.

Name	Max Wind Speed (mph)
Cindy	75
Dennis	150
Emily	160
Irene	105
Katrina	175
Maria	115
Nate	90
Ophelia	85

Name	Max Wind Speed (mph)
Philippe	80
Rita	180
Stan	80
Vince	75
Wilma	185
Beta	115
Epsilon	85

a. Write the maximum wind speeds in increasing order.

b. Fill in the five-number summary.

	minimum	Q_1	median	Q_3	maximum
Maximum Wind Speeds					

Rules for Finding the Median & Quartiles

When you have an even number of data values, the **median equals the average of the middle two numbers**. If the lower half and upper half of the data set also have an even number of values, Q_1 and Q_3 will be the average of the middle two numbers in the lower half and upper half, respectively.

3. Find the five-number summary of the speeds of the land animals below.

Land Animals	**Speed (mph)**
Cheetah	70
Pronghorn antelope	61
Thomson's gazelle	50
Quarter horse	48
Elk	45
Coyote	43
Ostrich	40
Greyhound	39
Rabbit(domestic)	35
Giraffe	32
Reindeer	32
Cat(domestic)	30
Grizzly bear	30
White-tailed deer	30
Human	28
Elephant	25
Black mamba snake	20
Squirrel	12
Pig (domestic)	11
Chicken	9

a. Write the speeds in increasing order.

b. Fill in the five-number summary.

	minimum	Q_1	median	Q_3	maximum
Speeds of Land Animals					

4. The following table lists the top 25 all-time highest grossing movies as of 9/16/11. Find the five-number summary of the 25 highest box office revenues (in millions of dollars).

#	Movie Title and Year	$
1	Avatar (2009)	761
2	Titanic (1997)	601
3	The Dark Knight (2008)	533
4	Star Wars: Episode IV (1977)	461
5	Shrek 2 (2004)	436
6	E.T.: The Extra-Terrestrial (1982)	435
7	Star Wars: Episode I (1999)	431
8	Pirates of the Caribbean (2006)	423
9	Toy Story 3 (2010)	415
10	Spider-Man (2002)	404
11	Transformers (2009)	402
12	Star Wars: Episode III (2005)	380
13	Harry Potter - Deathly Hallows (2011)	377

#	Movie Title and Year	$
14	The Lord of the Rings: Return of King (2003)	377
15	Spider-Man 2 (2004)	373
16	The Passion of the Christ (2004)	370
17	Jurassic Park (1993)	357
18	Transformers: Dark of the Moon (2011)	351
19	The Lord of the Rings: 2 Towers (2002)	340
20	Finding Nemo (2003)	340
21	Spider-Man 3 (2007)	337
22	Alice in Wonderland (2010)	334
23	Forrest Gump (1994)	330
24	The Lion King (1994)	328
25	Shrek the Third (2007)	321

a. Write the movie revenues in increasing order.

b. Fill in the five-number summary.

	minimum	Q_1	median	Q_3	maximum
Movie Revenues (in millions)					

Evolution of the Telephone

Most people know that Alexander Graham Bell is commonly credited as the inventor of the telephone. In 1876, Bell was the first to obtain a patent for an "apparatus for transmitting vocal or other sounds telegraphically". Most people don't know that the first cordless phone was patented in 1959 by Dr. Raymond P. Phillips Sr., an African American inventor from Terrell, Texas. The first cellular network in the world was built in 1977 in Chicago, IL. In 1979 the first cellular network (the 1G generation) was launched in Japan. With the evolution of cell phones, the sales of corded phones slowly phased out.

Consumer Electronics and Electronic Components – Factory Sales by Product 1990 to 2005 [In millions of dollars (11,021 represents $11,021,000,000).]

CATEGORY	1990	1995	2000	2001	2002	2003	2004	2005
Home office products, total	11,021	24,140	36,854	34,924	33,505	38,282	41,770	45,032
Cordless telephones	842	1,141	1,562	1,960	1,261	1,268	1,134	1,017
Corded telephones	638	557	386	294	266	256	259	261
Telephone answering devices	827	1,077	984	1,062	1,060	1,210	1,247	1,426
Caller ID devices	(NA)	(NA)	54	35	20	12	13	9
Personal computers	4,187	12,600	16,400	12,960	12,609	15,584	18,233	18,215
Computer printers	(NA)	2,430	5,116	5,245	4,829	4,734	4,053	3,858
Aftermarket computer monitors	(NA)	879	1,908	2,173	1,670	1,856	2,214	2,315
Modems/fax modems	(NA)	770	1,564	1,564	1,445	1,419	1,465	1,525
Computer peripherals	1,980	816	1,950	2,150	2,256	2,707	3,032	3,575
Computer software	971	2,500	4,480	5,062	4,961	5,060	5,162	5,250
Personal word processors	656	451	240	97	36	13	6	5
Fax machines	920	919	386	349	297	242	186	128
Digital cameras	(NA)	(NA)	1,825	1,972	2,794	3,921	4,739	7,468

Source: 2007 Statistical Abstract of the United States, U.S. Census Bureau

1. Use the **corded telephones** sales (in millions of dollars) to complete the table below.

Year	Years since 1990	Sales of Corded Phones (in millions of dollars)

2. Use technology to find the linear regression line. Identify the slope and y-intercept below. Round each number to the nearest hundredth.

   (Slope) $a =$ (y-intercept) $b =$

3. Identify the equation of the linear regression line. The equation is of the form $y = ax + b$.

4. Use the equation to estimate the sales in 1997.

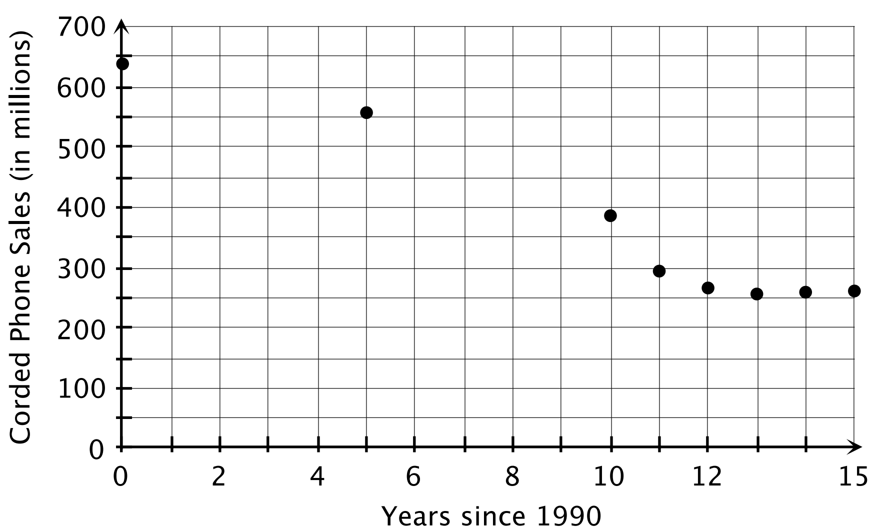
5. Is the prediction in question (4) an interpolation or an extrapolation? Explain.

6. What is the value of the correlation coefficient r?

7. What does the value of r indicate about the strength and direction of the data?

8. What does r tell us about the use of corded phones since 1990? Explain.

A graph of the corded phone sales over time is shown below.



9. Draw a trend line on the graph above that you feel fits the data points.

10. Identify two points on your line.

11. Find the equation of your line.

12. Compare the line you just created to the linear regression line which you calculated in question (3) by answering the following questions:

    a. Compare the *y*-intercepts. How close is your line's *y*-intercept to the regression line's *y*-intercept?

    b. Compare the slopes. How close is your line's slope to the regression line's slope?

    c. Is your line to steep or too flat? Explain how you came to your conclusion.

    d. Based on these answers, do you believe you drew a good line of fit? Explain.

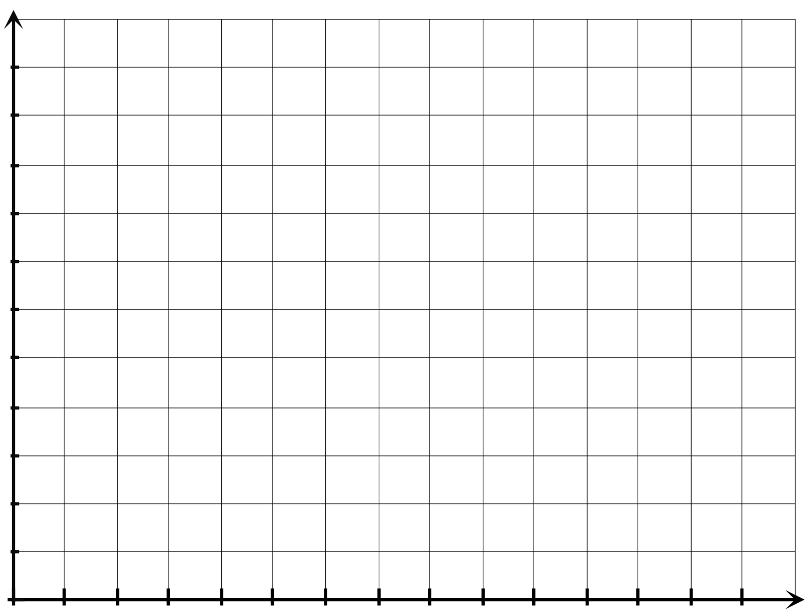
13. Use the equation you created to determine the year when the sales of corded phones will drop to 100 million dollars? Is this interpolation or extrapolation?

Go back to the table on page 1 and select a different product. Complete the table below using the sales information for this product. Write the name of the product above the table.

Product:

Year	**Years since 1990**	**Sales (in millions of dollars)**

14. Draw a scatterplot of the sales of the product over time in the coordinate plane below. Label and scale the axes appropriately.



15. Draw a trend line on the graph above that you feel fits the data points.

16. Circle the word(s) that describe the correlation between the sales of the product and the years since 1990. Explain your choice.

strong weak none

positive negative

17. Identify two points on your line.

18. Find the equation of your line.

19. Predict the sales of this product in the year 2015.

20. If you are offered a job to be the head of a marking department that sells this product, would you accept it based on research you just did? Do you think your job at this company would be secure for a long time?

21. Use technology to find the equation of the linear regression line.

22. Calculate the correlation coefficient r. Does this value of r confirm your prediction in question (14)?

23. How does your line compare to the linear regression line which you found in question (16)?

24. Did you make a good trend line? Explain.

Forensic Anthropology

While excavating for the new school building, construction workers found partial skeletal human remains. Who is this person? Is the person a male or a female? When did the person die? Forensic anthropologists were called in on the case to analyze the bones and help answer these questions. Police want to know the person's height in order to match the person with missing persons on file. Can you estimate the person's height from his or her skeleton bones?

The long bones such as the femur (thigh), tibia (shin) and ulna (forearm) predict height better than the shorter bones. The only intact long bone is the ulna, which is 28.5 centimeters in length. Your job as the forensic anthropologist's assistant is to estimate the height of the mystery person whose bones were found.

Step 1: STATE in a full sentence what you are going to find.

Step 2: COLLECT DATA on the height and ulna length from everyone in your class.

- Measure and record the ulna (forearm) lengths and heights of each of your classmates.
- Place your elbow on the table with the thumb pointed toward your body.
- Have the classmate measure from the round bone in your wrist, just below your pinky finger, to the bottom of your elbow, which is resting on the desk.
- People may measure differently, so it is best to have two different people measure your ulna. Use the average of the two measurements.
    - Measurement 1 = ________________
    - Measurement 2 = ________________
    - Average of 2 Measurements = ____________

Name	Ulna Length (cm)	Height (cm)

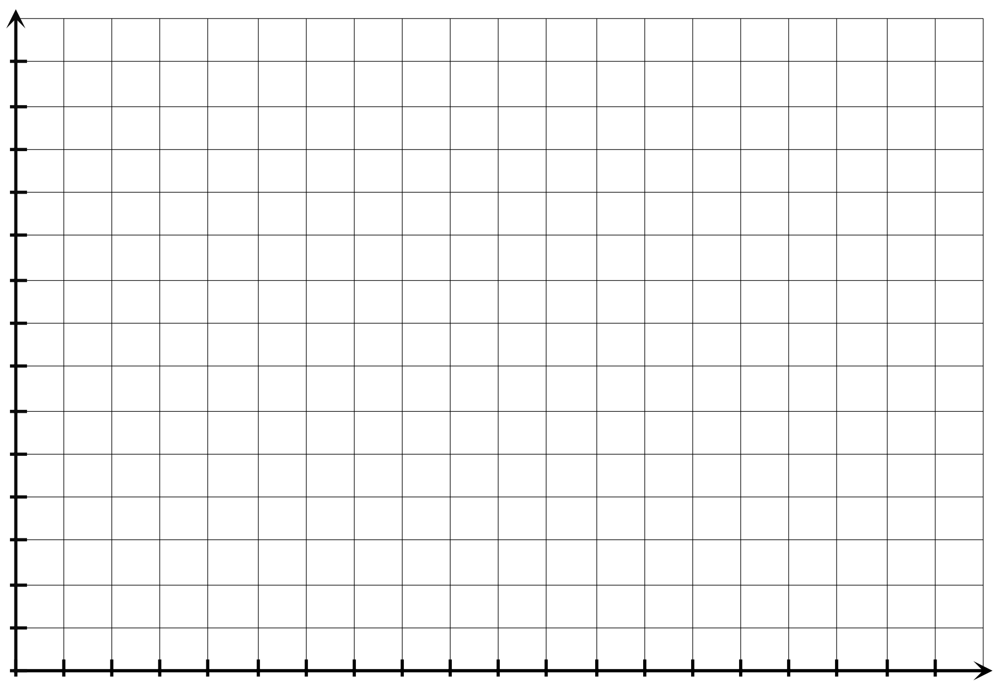
Name	Ulna Length (cm)	Height (cm)

Step 3: PLOT the data and find a trend line by hand from the graph.

a. Which is the independent variable?

b. Which is the dependent variable?

c. Graph the data on the coordinate axis below. Create a nice window for your graph with an appropriate scale. Label the axes and give the graph a title.



d. Does the data appear linear or not? Explain.

e. Sketch a trend line on your scatterplot by hand or using a straight edge. Find the equation of the trend line by hand. Show your work here.

f. Is your trend line equation similar to that of your classmates? Why or why not?

Step 4: USE TECHNOLOGYTO CALCULATE the regression line and correlation coefficient.

a. Write the regression line you found using technology.

b. Write the correlation coefficient *r* you found using technology.

c. Based on your correlation coefficient, comment on the direction and the strength of the linear relationship between ulna lengths and height.

d. Compare the hand-calculated trend line from step 3e with the regression line you found using technology in step 4a. How close are the two slopes and *y*-intercepts? What are the advantages or disadvantages of doing the work by hand or using technology?

Step 5: Make a **PREDICTION**.

a. Use the linear regression line that you found using technology to estimate the height of the missing person. Remember that the found ulna bone is 28.5 centimeters long. Show your work.

b. How accurate or reliable is your prediction? Explain your answer by referring to the graph and the correlation coefficient *r*.

<u>EXTENSIONS:</u>

1. Your boss asked you to do some research online to see if there is already a formula or maybe a table that can be used to determine someone's height from his or her ulna bone length. What is this formula?

2. Use the anthropologist's formula to predict the height of the mystery person. How close was your prediction to the anthropologist's prediction?

3. What might explain the differences in the height that you found using your regression line and the height you found using the anthropologists formula?

Present Your Results

Write a letter to the commissioner of the police department informing him of the estimated height of the mystery person.

- ✓ Describe the steps you took to find the height of the person.
- ✓ Analyze how well the model scatterplot and your equation represent the data you collected.
- ✓ Is there a correlation or causation between ulna length and height? Explain
- ✓ Discuss whether or not you could use *your* model to estimate or predict the height of the mystery person.
- ✓ Compare your estimation of the height of the person to the estimation using the anthropologist's equation.
- ✓ Include in your report equations, graphs, estimates, and predictions.

Population and Representation

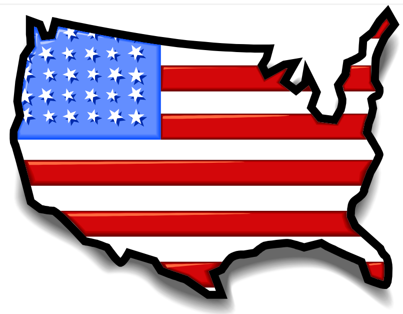
Each person (or pair) is going to research the relationship between the population of a state and the number of representatives that state elects to the United States House of Representatives.

Doing the Research

You will go to a search engine (i.e. Google, Yahoo, etc.) and research two topics:

- The **current population** for each state
- The **current number of elected representatives** for each state

(To save time you only have to find the population and number of elected representatives for 15 states. If you find the data for more states, your predictions will be more accurate.)



Writing about the Results

1. Explain the two variables you are comparing. What did you type into the search bar to get your results? Cite the websites you used.

2. Make a table of the data points that you collected.

3. Create a scatter plot on a piece of graph paper or using technology. Draw the regression line on the scatterplot.

4. Describe the correlation.

5. Write the equation of the regression line.

Predict *y*, given *x*

6. Choose a value for your x-variable.

7. Use your equation to calculate the approximate number of elected representatives (y) for the value of x you chose above.

8. Describe your prediction in a sentence.

Predict *x*, given *y*

9. Choose a value for your y-variable.

10. Use your equation to calculate the approximate population of a state (x) for the value of y you chose above.

11. Describe your prediction in a sentence.

12. Write a paragraph explaining what you learned about the relationship between your variables and how this knowledge could be useful to someone.

Conducting an Experiment

Each group is going to examine the relationship between a certain type of aerobic exercise and a person's heart rate.

Create an Experiment

Your group needs to think of an aerobic exercise that you can do in the classroom or in the hallway. Your group's job is to come up with an experiment that shows how a teenager's heart rate changes based on a type of aerobic exercise.

Describe the Experiment

1. Explain the two variables you are comparing. Then, in detail, list the steps that will allow you to perform an experiment to study the relationship between these two variables.

Write about the Results

2. Make a table of the data points that you collected.

3. Create a scatter plot on a piece of graph paper or using technology. Draw the regression line on the scatterplot.

4. Describe the correlation.

5. Write the equation of the regression line.

<u>Predict *y*, given *x*</u>

6. Choose a value for your *x*-variable.

7. Use your equation to calculate the heart rate of a person (*y*) for the value of *x* you chose above.

8. Describe your prediction in a sentence.

<u>Predict *x*, given *y*</u>

9. Choose a value for your *y*-variable. (Do not choose a heart rate value that is less than your resting heart rate.)

10. Use your equation to calculate the value of *x* for the value of *y* you chose above.

11. Describe your prediction in a sentence.

12. Write a paragraph explaining what you learned about the relationship between your variables and how this knowledge could be useful to someone.

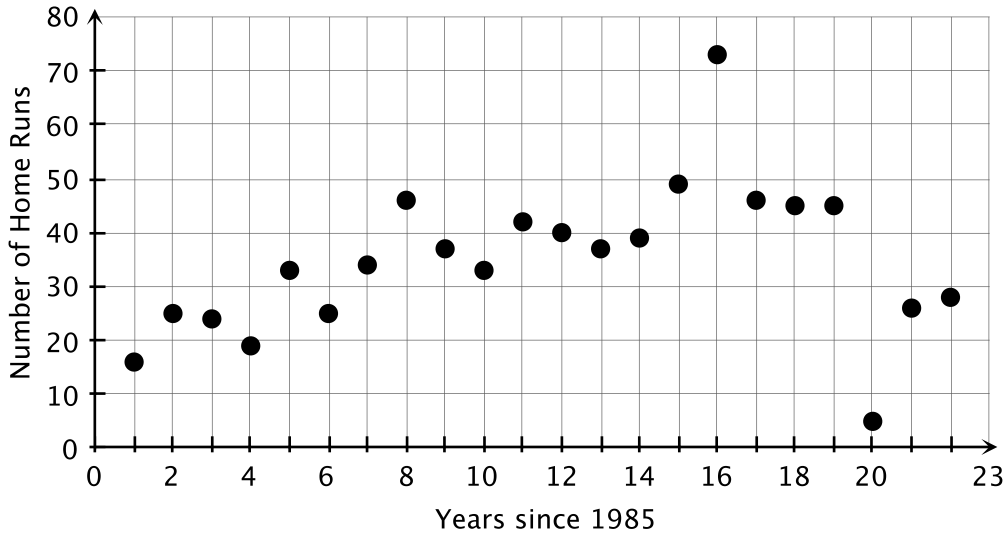
Barry Bonds' Home Runs

Many consider San Francisco Giants' slugger Barry Bonds to be one of the greatest baseball players of all time. In 2001, he hit 73 home runs and broke the record for the most home runs hit in a single season. However, in 2011, Bonds was found guilty of obstruction of justice due to his testimony on steroid use which was found to be evasive and misleading. His possible steroid use has led many to question whether his home run record should be allowed to stand.

Barry Bonds played in the major leagues for 22 years. His home run and games played statistics for each season are displayed in the table below.

Season	Home Runs	Games Played
1986	16	113
1987	25	150
1988	24	144
1989	19	159
1990	33	151
1991	25	153
1992	34	140
1993	46	159
1994	37	112
1995	33	144
1996	42	158
1997	40	159
1998	37	156
1999	39	102
2000	49	143
2001	73	153
2002	46	143
2003	45	130
2004	45	149
2005	5	14
2006	26	130
2007	28	126

Below is a scatterplot showing the number of home runs Barry Bonds hit each season.



1. From the table and the graph, identify two points that appear to be outliers.

2. Give a possible cause for each outlier.

3. In 2005 Barry Bonds missed most of the season due to a knee injury. Use the table or the scatter plot to explain whether this might have affected his performance in 2006 and 2007.

4. Calculate the regression line for 1986 through 2004, which are the years before his knee injury. (Suggestion: Let the independent variable be years since 1986.)

5. Determine the correlation coefficient *r* for the data from 1986 to 2004.

6. Use your regression line to predict the number of home runs Barry Bonds would hit in 2001? How does this compare with the actual value?

7. Now eliminate the data for the year 2001. Calculate the regression line for 1986 through 2004, excluding the year 2001. (Suggestion: Let the independent variable be years since 1986.)

8. Determine the correlation coefficient *r* for the data from 1986 to 2004, excluding the year 2001.

9. Compare the two regression lines:

   a. Which has the greater slope?

   b. Which shows a stronger correlation?

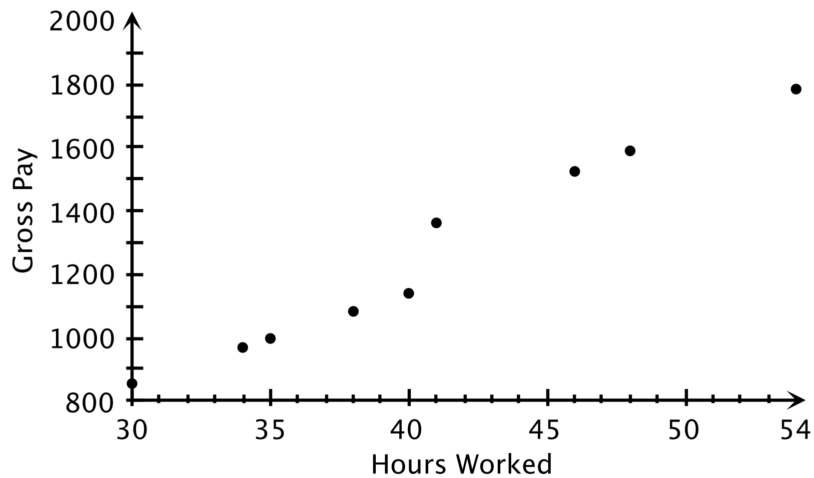
   c. Explain the differences.

10. If we include all 22 years of Bonds' career in finding a regression line, how would this affect the slope and the strength of the correlation? Make a prediction and then test it using a calculator or appropriate software.

Paychecks & Triathlons

Steve works for a theatrical lighting design company. He works many hours during the summer months when the company is very busy. The following table lists his gross pay (before taxes and other deductions) based on the number of hours he worked in a week.

Hours worked	Gross Pay
30	855
34	969
35	997.50
38	1083
40	1140
41	1182.75
46	1396.50
48	1482
54	1738.50



1. Does one line fit the data well, or will it take more than one line to obtain a good fit?

2. When does the pattern change? What do you know about pay rates that might explain this change?

3. Use technology to find two regression lines. Line 1 contains the points with *x*-values from 30 to 40, and Line 2 contains the points with *x*-values from 41 to 54.

Line 1:	Line 2:
What does the slope mean?	What does the slope mean?

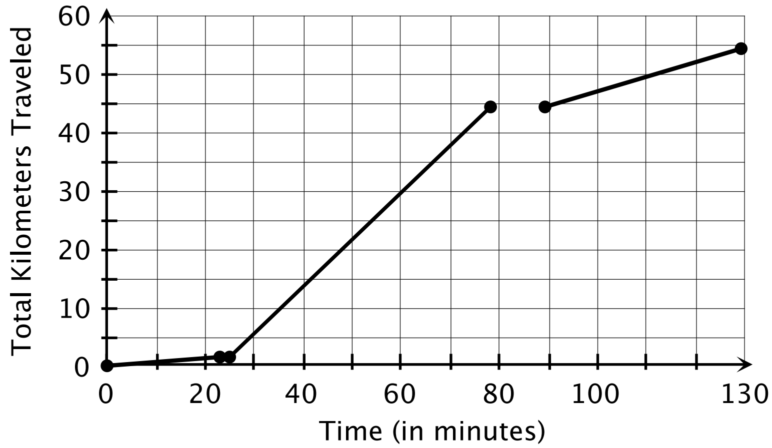
4. A piecewise function contains two or more rules with each rule acting on a different part of the function's domain. Label Steve's weekly pay function $f(x)$ and write each rule on its own line, along with the part of the domain where that rule is applied.

$$f(x) = \begin{cases} \_\_\_\_\_\_\_\_\_\_ \; when \; \_\_\_\_\_\_\_\_\_\_ \\ \_\_\_\_\_\_\_\_\_\_ \; when \; \_\_\_\_\_\_\_\_\_\_ \end{cases}$$

5. Find $f(32)$ and $f(62)$. Make sure you use the correct rule.

Helen Jenkins, of Great Britain, finished first in the Dextro Energy Triathlon in London on August 6, 2011. The race consisted of a 1.5 km swim, a 42.9 km bike ride, and a 10 km run. Below is a table and graph of her time and distance as she completed each leg of the race.

Time≤	Distance
0	0
24	1.5
24.6	1.5
78.2	44.4
89.3	44.4
129.3	54.4



6. Find an equation for the total distance Helen traveled from the time she started during each event (swimming, biking, running). Find each equation by hand. (Assume that she traveled at a constant rate during each event.)

a. Swimming b. Biking c. Running

7. How much time did Helen have to rest:

   a. Between swimming and biking?

   b. Between biking and running ?

8. Complete the piecewise function that models the distance Helen traveled during the race

$$f(x) = \begin{cases} \_\_\_\_\_\_\_\_\_\_\_\_ \, when \, \_\_\_\_\_\_\_\_\_\_\_\_ \\ \_\_\_\_\_\_\_\_\_\_\_\_ \, when \, \_\_\_\_\_\_\_\_\_\_\_\_ \\ \_\_\_\_\_\_\_\_\_\_\_\_ \, when \, \_\_\_\_\_\_\_\_\_\_\_\_ \end{cases}$$

9. Find $f(10)$.

10. Find $f(50)$.

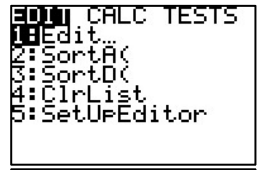
11. Find $f(100)$.

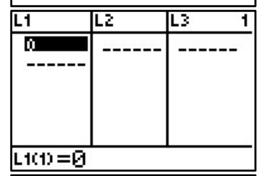
Graphing Calculator Instructions

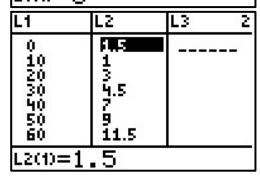
This document contains TI-83 and TI-84 calculator instructions for:

- Entering in a data set or multiple data sets
- Clearing an entire list in the List Editor
- Setting up the List Editor
- Calculating single-variable statistics on a data set
- Creating a scatterplot
- Calculating the slope and intercept of a linear regression line and the correlation coefficient
- Evaluating a function
- Solving an equation

Entering Data into Lists





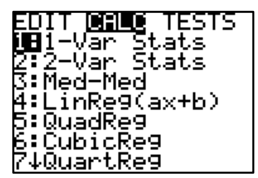


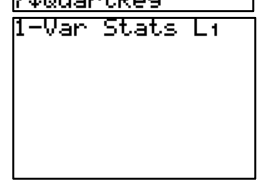
a. Press **STAT** and select **1:Edit** to go to the list editor.
b. Move the cursor to the list that you wish to edit. Type in a number. Press **ENTER**.
c. Continue entering numbers into a list. Press **ENTER** after typing in each number.
d. Then (if necessary) enter another data set into another list.
e. You can amend each single entry with the **DEL** key.

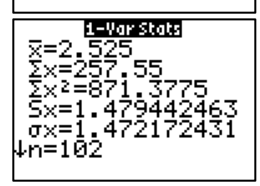
You can sort a list that you entered by using the **SortA** command. Press **STAT**, and select **2:SortA**. **SortA(** will appear on the home screen. Then enter the name of the list you want sorted (by pressing **2nd 1** (for L_1)), and press **ENTER**.

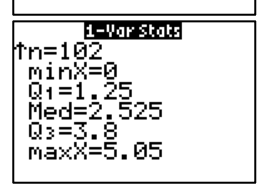
If you enter in two lists containing paired data, be sure that the lengths (also called *dimensions*) of the two lists are the same. Otherwise, you will get the *Dimension Mismatch* error message when you try to graph the data or perform regression.

Calculating Single-Variable Statistics on a Data Set





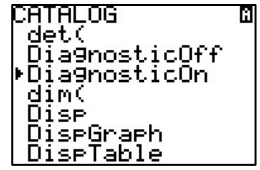




a. You must have a list entered into L_1.
b. Start at the home screen. Clear the home screen.
c. Press **STAT**, move the cursor to the right to highlight **CALC**.
d. Select **1:1-Var Stats**.
e. On the home screen you will see the **1-Var Stats** command. You must assign this command a list.
f. Press **2nd 1 (L_1)** to place the list name L_1 next to the command.
g. Press **ENTER**.

You will see the output of the **1-Var Stats** command on the home screen. The output contains several statistics. The statistics include the mean, population standard deviation, sample standard deviation, sample (or population) size, minimum value, first quartile, median, third quartile, and maximum value.

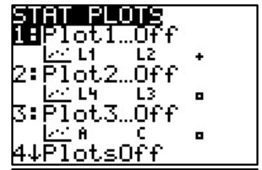
Turning on the Diagnostics to Obtain the Correlation Coefficient *r*

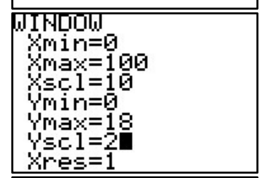


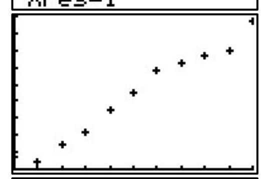
a. Clear the home screen.
b. Press **2nd 0 (CATALOG)** and scroll down to **DiagnosticOn** then press **ENTER**.
c. Home screen now shows **DiagnosticOn**. Press **ENTER**.
d. Home screen shows **Done**.

If the Diagnostic is turned off, the **LinReg(ax+b)** command (discussed below) will not output the correlation coefficient.

Drawing a Scatterplot

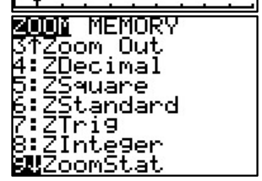




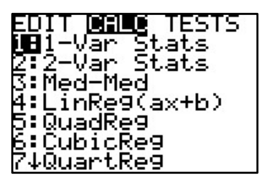


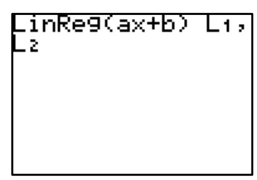
a. You must have two lists with the same dimension in L_1 and L_2.
b. Press **2nd Y= (STAT PLOT)**
c. Select **Plot 1** by highlighting **Plot 1** and pressing **ENTER**.
d. Set up the STAT PLOT by selecting **On**, selecting the first **Type** (the first type is a scatter plot), making the **Xlist** equal to $\mathbf{L_1}$, making the **Ylist** equal to $\mathbf{L_2}$, and selecting the **Mark** you wish to use.
e. Choose an appropriate window to view the scatterplot based on the minimum and maximum values of the data in the two lists. Press **WINDOW** and enter the minimum and maximum values for *x* and *y*. Choose a scale for each axis using **Xscl** and **Yscl**. Set **Xres** = 1 for the screen resolution.
f. Press **GRAPH** to see the scatterplot.

The calculator can create a window which contains the entire scatterplot automatically. Instead of manually defining the window manually, press **ZOOM**, then select **9**:**ZoomStat. ZoomStat** automatically sets a window that captures all data points in the scatterplot.



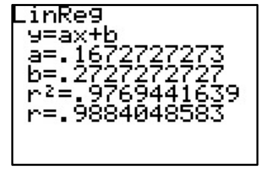
Calculating the Slope and Intercept of a Linear Regression Line and the Correlation Coefficient





a. You must two lists with the same dimension in L_1 and L_2.
b. Start at the home screen. Clear the home screen.
c. Press **STAT**, move the cursor to the right to highlight **CALC**.
d. Select **4:LinReg(ax+b)**.
e. On the home screen will be the command **LinReg(ax+b)**.
f. Press **2nd 1 ($\mathbf{L_1}$)** to get the list, L_1, that contains the independent variable.
g. Press **,** (the comma key), which is to the right of the x^2 key.
h. Press **2nd 2 ($\mathbf{L_2}$)** to get the list, L_2, that contains the dependent variable.
i. Press **ENTER**.

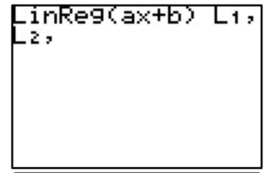
You will see the slope, *a*, and the *y*-intercept, *b*, of the regression line. You also will see the correlation coefficient, *r*, and the coefficient of determination, r^2, if you have turned the **Diagnostic** on. (See **Turning on the Diagnostic to Obtain the Correlation Coefficient *r***)

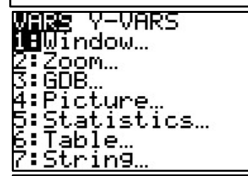


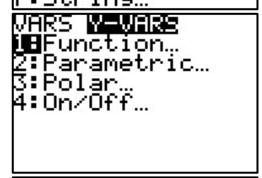
Graphing a Regression Line and Scatterplot in the Same Window

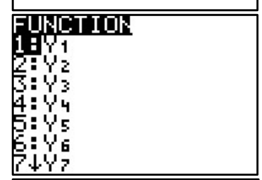
a. You can store the regression line into a function in the calculator by running the **LinReg(ax+b)** command. Select the command **LinReg(ax+b)**.
b. Type in $\mathbf{L_1}$ **,** (comma) $\mathbf{L_2}$ **,** (comma) .
c. You now must enter in a function name (such as Y_1 or Y_2). To enter the function name Y_1, press **VARS**, move the cursor right to **Y-VARS**, select **1:Function**, then select **1:**$\mathbf{Y_1}$.
d. The home screen should now show the command **LinReg(ax+b)** $\mathbf{L_1}$, $\mathbf{L_2}$, $\mathbf{Y_1}$ which will calculate the regression line and store the regression line into Y_1.
e. Press **ENTER** to execute the command.
f. Press **Y=** key to verify that the regression line is indeed pasted into Y_1. Make sure that **Plot 1** at the top is highlighted. This specifies that the scatterplot in STAT PLOT 1 and the regression line in Y_1 will be plotted simultaneously. You can remove the highlight from **Plot 1** if you want to only show the regression line.
g. Press **GRAPH**.

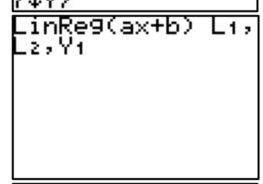
You can also enter the regression line in for Y_1 by hand. Press **Y=** and type in the regression equation which the calculator obtained.

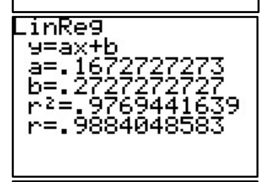


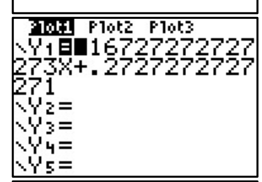


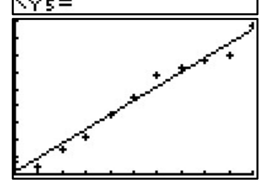




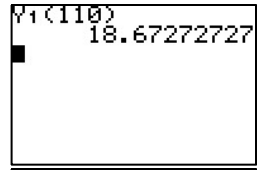


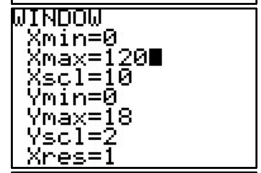


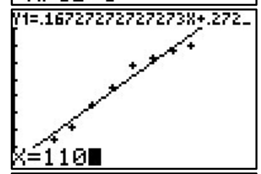


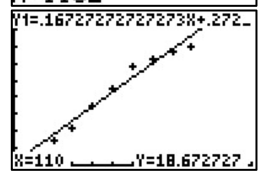


Evaluating a Function



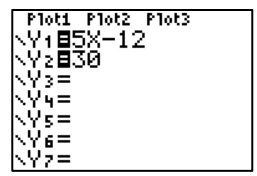


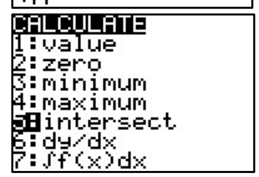


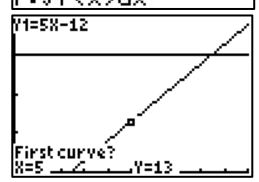


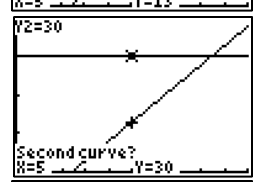
a. To evaluate a function, store a function $y = f(x)$ in Y_1.
b. To evaluate Y_1, you must bring the function name Y_1 to the home screen.
c. Press **VARS**, move to cursor to the right to **Y-VARS**, then select **1:Function**, then select **1: $\mathbf{Y_1}$**. This will make $\mathbf{Y_1}$ show up on the home screen.
d. Then enter **(110)** to evaluate $Y_1(110)$.
e. Press **ENTER**.
f. You can also use the table or graph feature to evaluate a function.
   - **Table feature**: You may want to choose "ask" mode for the independent variable. This mode lets you choose any input. When you press **TABLE** a blank table will appear. Enter in any value you want for the independent variable, press **ENTER**, and the corresponding dependent variable will appear.
   - **Graph feature**: When the function is graphed, press **2nd Trace (CALC)**. In the Calculate menu, select **1: value**. Enter a value for *x* and press **ENTER**. Make sure that you are in a graphing window which includes the input value which you want to use.

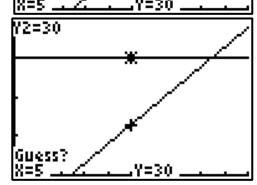
Solving an Equation

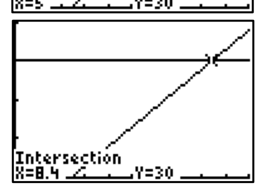








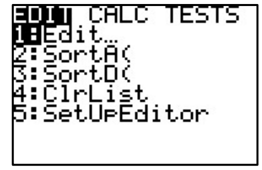




Consider the equation $5x - 12 = 30$. To solve this equation using the graphing calculator you must treat each side of the equation as a function. Then you can find the intersection point (if it exists) of the two functions.

a. Enter the expression $5x - 12$ into Y_1 and enter 30 into Y_2.
b. You must set a window that contains the intersection point of the two functions.
c. Press **GRAPH**.
d. Press **2nd TRACE (CALC)**.
e. Select **5: intersect**.
f. The calculator asks you to select the first curve. One of the functions will be highlighted. Press **ENTER**.
g. The calculator asks you to select the second curve. The other function will be highlighted. Press **ENTER**.
h. You are asked for a guess. You can leave this prompt blank. Press **ENTER**.
i. The intersection point (solution of the equation) is displayed at the bottom of the screen.

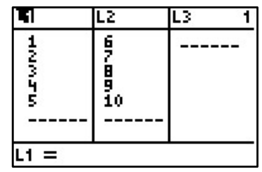
Setting Up the Lists in the Editor



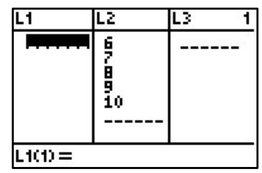
a. Press **STAT**.
b. Select **5:SetUpEditor** and press **ENTER**.
c. The home screen will show **SetUpEditor**.
d. Press **ENTER**.
e. The home screen will show **Done**.

You only need to run SetUpEditor if you are missing a list in the List Editor or if your lists are out of order. You do not need to do this every time.

Clearing an Entire List in the Editor



a. Press **STAT** and select **1:Edit**.
b. Place the cursor over the list name at the top so the list name is highlighted.
c. Press **CLEAR** and then press **ENTER** to clear an entire list. *Note*: you must highlight the list name at the top, not the first entry.



You can replace entries in a list with new data by simply typing over the existing data in the list.