New technology and new applications of mathematics should encourage us to re-examine our course content – at the very least.

Applications that aren't usually part of the algebra curriculum:

Codes

C6 | =CODE(C5)-64

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Sample Process of Coding and Decoding											(Simple Version)						
2																		
3	Input Message to be Coded (IN ALL CAPS)																	
4																		
5	A	B	C	A	R	O	L	I	N	A	O	N	M	Y	M	I	N	D
6	4	1	3	1	18	15	12	9	14	1	15	14	13	25	13	9	14	4
7																		
8	Coded Message (Using Algorithm AX + B)																	
9	4	1	13	5	73	61	49	37	57	5	61	57	53	101	53	37	57	17
10																		
11	Decode the Message																	
12			3	1	18	15	12	9	14	1	15	14	13	25	13	9	14	4
13			C	A	R	O	L	I	N	A	O	N	M	Y	M	I	N	D

Data Mining

Example of a Market Basket Transaction

s{a}	=	10/12		
s{b, d}	=	3/12		
s{a, b, d}	=	2/12		
Support and Confidence for an Association Rule:				
s {b, d} → {a}	=	2/12	**17%**	
c {b, d} → {a}	=	$^{(2/12)}/_{(4/12)}$ = 2/4		**50%**
one-half of the transactions in which {b,d} was bought				
also bought {a}: this rule applies to 17% of all transactior				
s {a} → {b, d}	=	2/12	**17%**	
c {a} → {b, d}	=	$^{(2/12)}/_{(10/12)}$ = 2/10		**20%**
20% of transactions who bought {a} also bought {b,d}				
this rule applies to 17% of all transactions				

Customer ID	Transaction ID	Items Bought
1	0001	{a, b, c}
2	0024	{a, b, c, e}
2	0038	{a, b, e}
3	0012	{a, b, d, e}
3	0044	{a, c, e}
3	0088	{b, d, e}
4	0015	{a, b, c}
4	0055	{a, b, c, d, e}
5	0022	{a, b, c}
6	0035	{a, c, e}
6	0042	{a, d, e}
6	0080	{b, c, d, e}

<This is conditional probability in a new language/symbols.>

Problem: Provide all confidences that are greater than .8

Simulations

Periodic Dosing with Exponential Decay

This page allows you to analyze a model for drug delivery. Assume that we use uniform dosage and that the administration of a dose immediately results in the recipients concentration of the drug increasing by c_o units. N doses are administered at uniform interval of T time units. Furthermore, the body eliminates the drug according to an exponential decay model where $1/e$ is removed in τ time units. In a given dosage interval, let $c(t)$ be the concentration and let $t = 0$ when dosed, then the concentration obeys the differential equation $dc/dt = -c(t)/\tau$ with the initial condition $c(0) = c_o$. This initial-value problem has the unique solution $c(t) = c_o e^{-t/\tau}$.

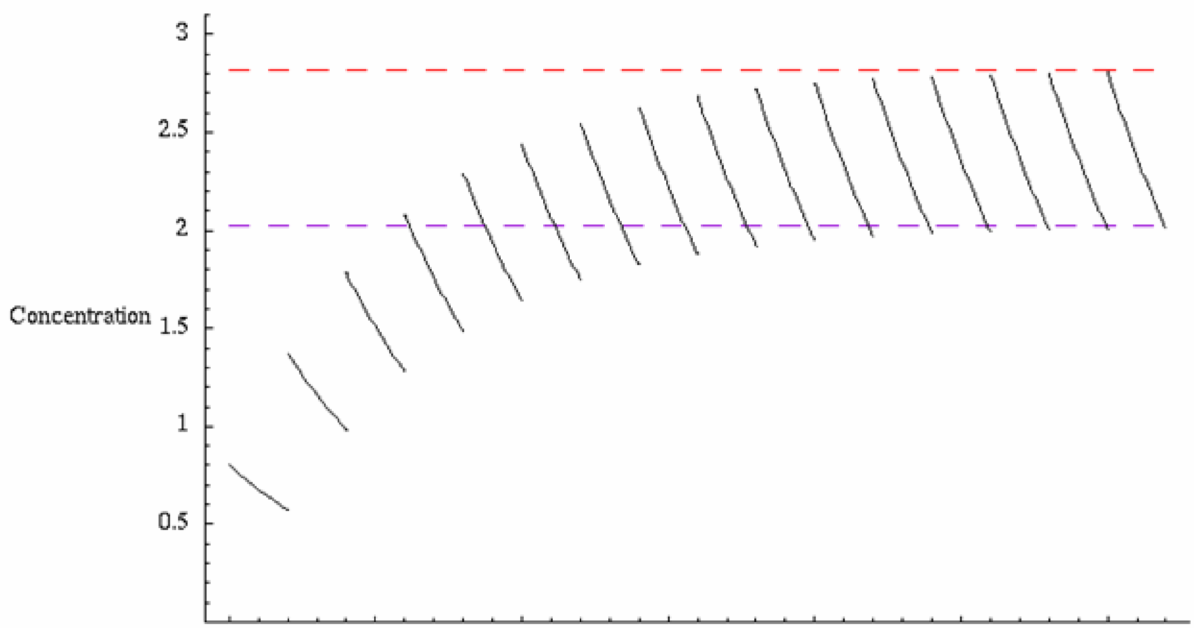
c_o = 0.8

T = 2

N = 16

τ = 6

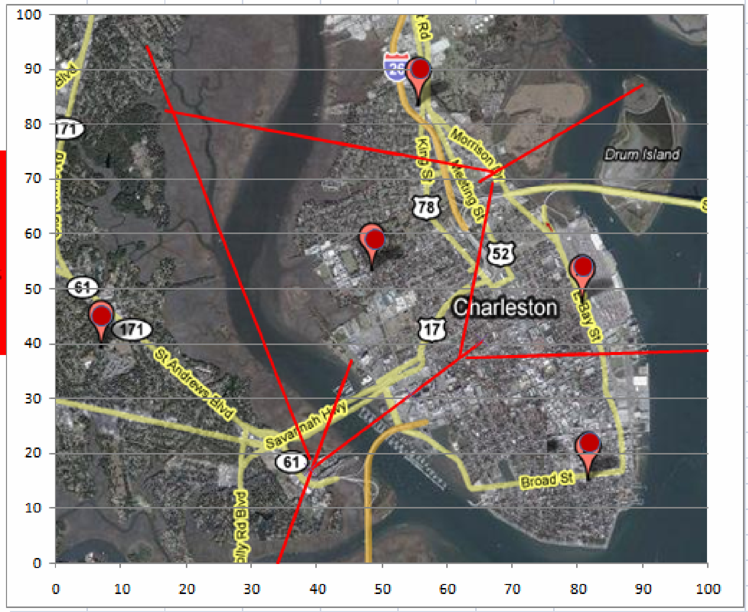
Evaluate



Voronoi Diagrams

	A	B
1	7	45
2	49	59
3	56	90
4	82	22
5	81	54

Charleston Post Offices



two points		midpoint		slope
7	45	28	52	0.3333333
49	59			

two points		midpoint		slope
49	59	52.5	74.5	4.4285714
56	90			

two points		midpoint		slope
49	59	65.5	40.5	-1.1212121
82	22			

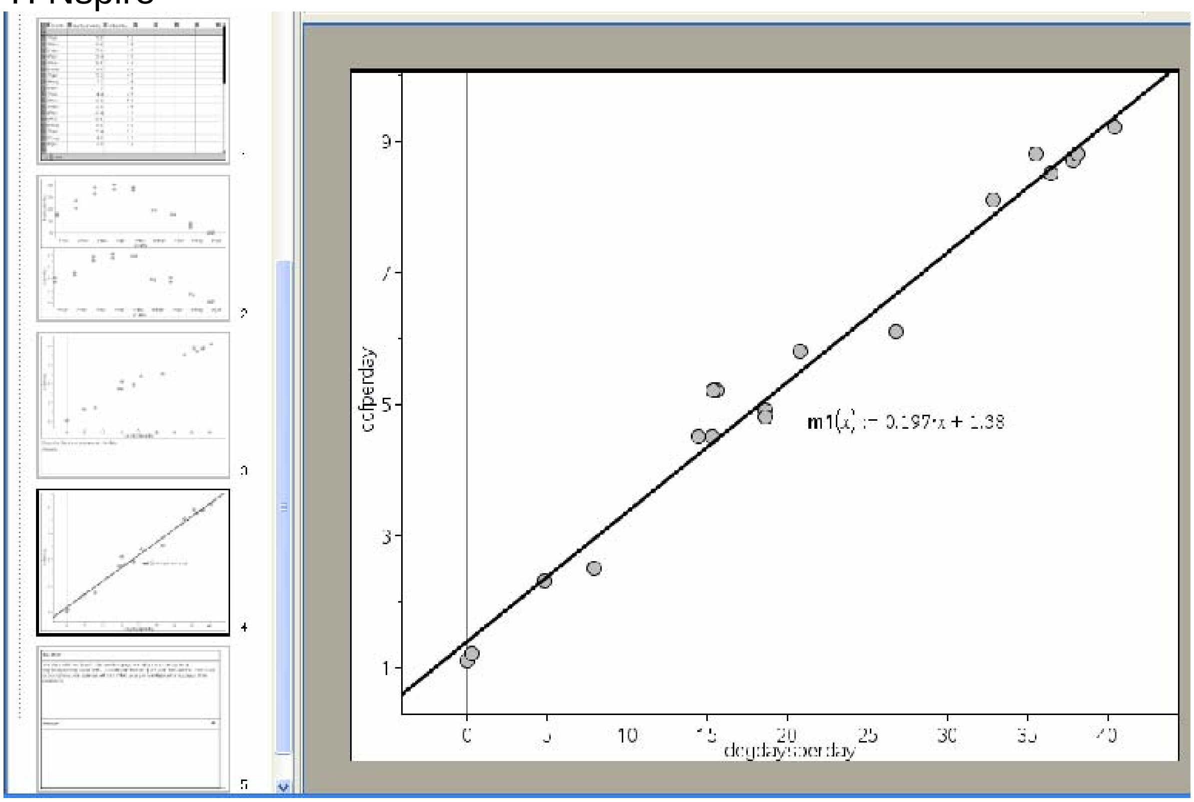
two points		midpoint		slope
7	45	44.5	33.5	-0.3066667
82	22			

two points		midpoint		slope
56	90	68.5	72	-1.44
81	54			

two points		midpoint		slope
81	54	81.5	38	-32
82	22			

two points		midpoint		slope
49	59	65	56.5	-0.15625
81	54			

TI Nspire



Overview

In this activity, you will explore data sets using technology. While doing so, you will

- Use vocabulary associated with functions and data analysis.
- Explore methods of representing data; numerically, symbolically, graphically, and verbally.

Introduction

A function defines a relationship between sets of data. Functions can be displayed using numerical, graphical, or symbolic representations. In some cases, the value of the independent variable accounts for changes in the value of the dependent variable. We are often concerned with the degree to which the value of the independent variable can predict the value of the dependent variable; and if so, how accurate that prediction is.

Directions

1. Problem 1

In order to learn the most from numerical data, we often view it as a graph.

The graph can show trends that we might not notice numerically. We can also analyze the data for correlation - the affect one variable has upon the other.

Consider the data shown in the spreadsheet (right).

On November 10, 1994 in Raleigh, the day started off delightful - sunny, pleasant, and getting warmer. However, about midmorning, a cold front passed through the area and things changed. (The time is shown using military format; 13:00 = 1:00 pm.)

	A time	B temp	C	D	E	F	G
1	600	61					
2	700	62					
3	800	64					
4	900	66					
5	1000	67					
6	1100	63					
7	1200	57					
8	1300	55					
9	1400	53					
10	1500	51					
11	1600	50					
12	1700	49					
13	1800	46					
14							
15							
16							
17							
18							
19							
20							

A1 600

a) At about what time did the temperature (temp) change from increasing to decreasing?
10:00

b) What was the average rate of change of temperature from 6:00 until 10:00?
=(67-61)/4 = 1.5 degrees per hour

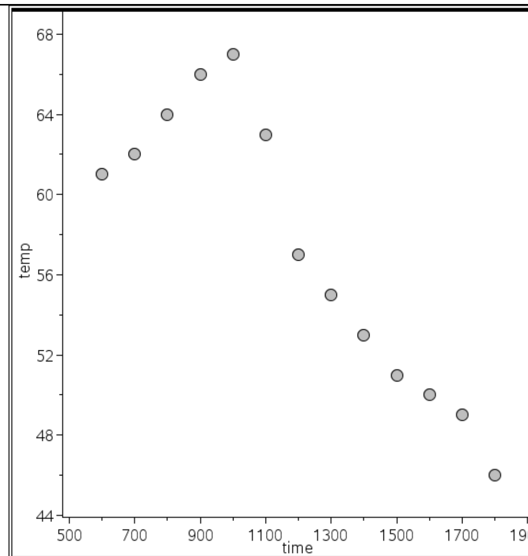
c) What was the average rate of change of temperature from 10:00 until 16:00?
=(50-67)/6 = -17/6 ~ -2.8 degrees per hour

d) Use the model to predict the temperature at 14:30. (The table of values represents the model – the relationship between the time and temperature.)
52 degrees

2. **Create a scatterplot of the data.**

   a) In which representation, the numerical representation (spreadsheet) or the graphical representation (scatterplot), is it easier for you to make observations about this data?
   Answers may vary, but the graph shows the overall trend in temperature more easily than does the numerical display.

   b) In which representation is it easier for you to ascertain exact information?
   Answers may vary, but the numerical data can be used to ascertain exact numbers more easily.



Place your scatter plot here.

3. **Problem 2**
   A vendor has several ice cream stands along a long stretch of beach. He believes he can use the weather to predict how much help he will need because the outside temperature has an effect on amount of ice cream purchased.

   The data shown in the spreadsheet was collected over the months of June and July one year. It shows the temperature at 2 pm, the average number of sales per hour he had that day (all locations combined), and the day of the week.

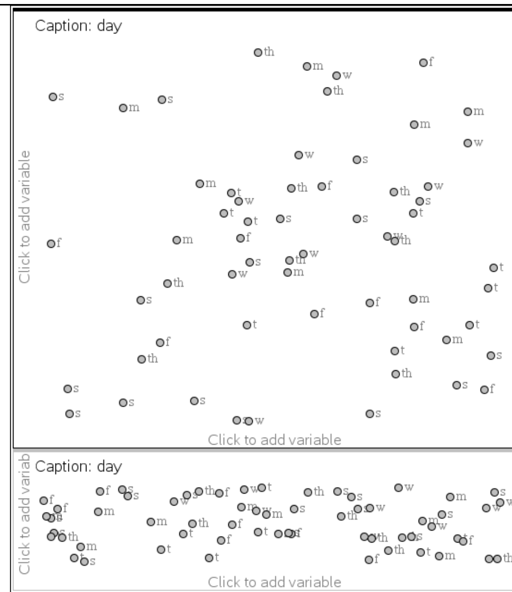
   Describe any trends you see?
   Students may say the sales per hour goes up and down. However be careful to point out that since the temperature data is not ordered, the changes in the sales aren't in order either.
   If students look closely, they can see that sales are higher on Saturday and Sunday and on days on which the temperature was highest.

	A tempat2	B salesperhr	C dayofweek	D	E	F
◆						
1	70	16	m			
2	78	21	t			
3	82	23	w			
4	83	25	th			
5	85	33	f			
6	84	37	sa			
7	85	38	s			
8	66	14	m			
9	78	21	t			
10	82	24	w			
11	88	26	th			
12	80	30	f			
13	85	31	sa			
14	71	28	s			
15	72	20	m			
16	78	19	t			
17	84	25	w			
18	89	26	th			
19	89	35	f			
20	93	36	sa			

A1 70

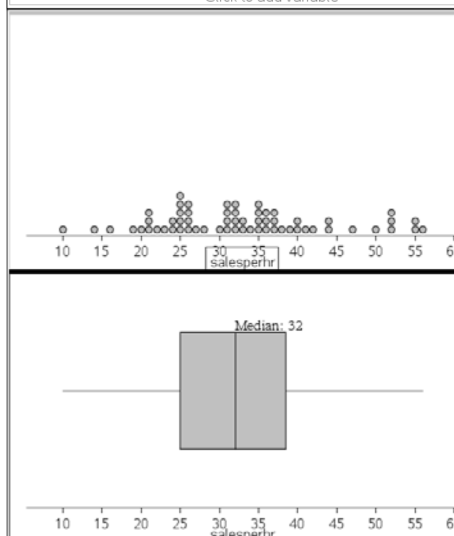
4. **There may be several trends or patterns you see in the data, but it is more difficult to see these patterns without re-organizing the data.**

   The unorganized data is shown (below). Perform the actions described and document any trends in the data you believe are worthwhile.



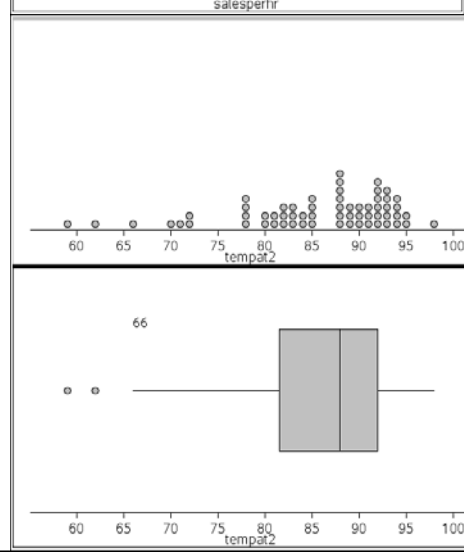
   a) In the larger window, click on the variable box on the horizontal axis and view the data arranged by "salesperhr". Estimate the median number of sales per hour over these two months.

   In the smaller window, click on the variable box on the horizontal axis and choose "salesperhr" and then create a Box Plot. How close was your estimate?
   Students should be able to estimate the median at around 32; between 30 and 35. They should see that sales per hour varied greatly. You should ask why the sales per hour might vary (temperature and day of the week)



   (b) In the larger window, click on the variable box on the horizontal axis and view the data arranged by "tempat2". Estimate the median temperature at 2 pm over these two months.

   In the smaller window, click on the variable box on the horizontal axis and choose "tempat2" and then create a Box Plot. How close was your estimate?
   The median is 88. You might ask the students to imagine what kind of day it was on which the temperature at 2 pm was 59, or 62, or 97.

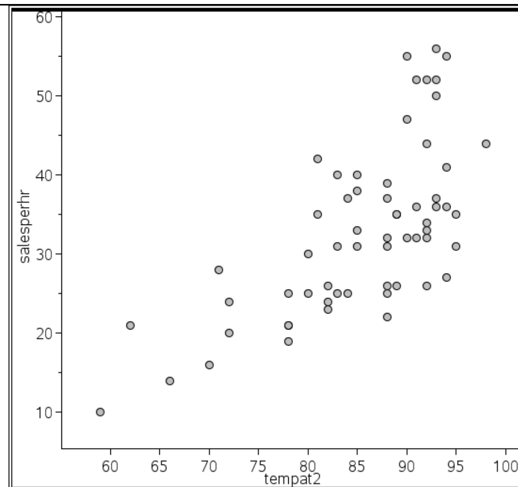


5. **The information you found in (4) may be interesting, but doesn't provide the vendor with the most useful information.**

   The data is organized by "tempat2". Click on the variable box along the vertical axis and view the data arranged by "tempat2" and "salesperhr".

   a) **Describe the trends do you see?**
   The higher the temperature the greater the sales per hour. There seems to be more variability in the sales per hour at the same temperature for higher temperatures.

   b) **What reasons may cause a great deal of variation in the number of sales per hour when the temperature is a warm 93 to 95 degrees?**
   The day of the week may be a major factor in the number of sales per hour. There may be more people on the beach on the weekends, if if the temperature isn't as high as on a week day.



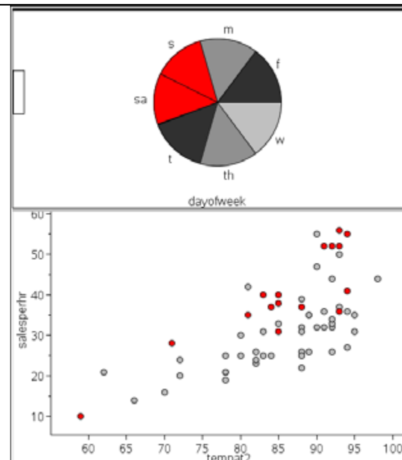
Scatterplot of tempat2 vs salesperhr.

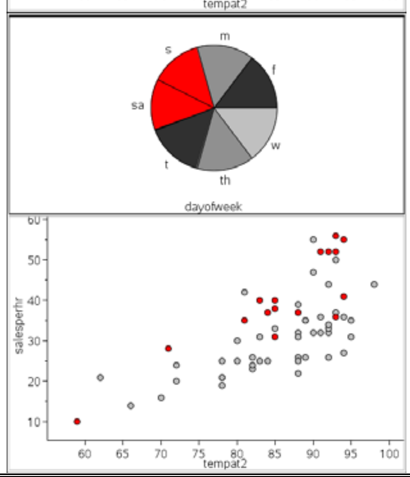
6. **The data is organized by "tempat2".**

   Click on the the "sa" and "s" slices (one at a time) in the pie chart. As you do, the data that comes from Saturdays and then Sundays is shown in red (or shaded).

   In what ways does this added dimension to the scatter plot help with seeing trends that might be useful to the vendor?
   It appears that by separating the data into these two categories, data from Saturday and Sunday and data from the weekdays, that the sales per hour correlate much better with the temperature. This will help the vendor to make decisions on the amount of help and supplies that might be needed based on the temperature.

   In what ways are the graphical displays of data much more useful in this example (ice cream vendor) than in the previous example (temperature in Raleigh)?
   The graphical display of the ice cream vendor data organizes the data so that we can make sense out of it. The numerical data doesn't allow that as easily.





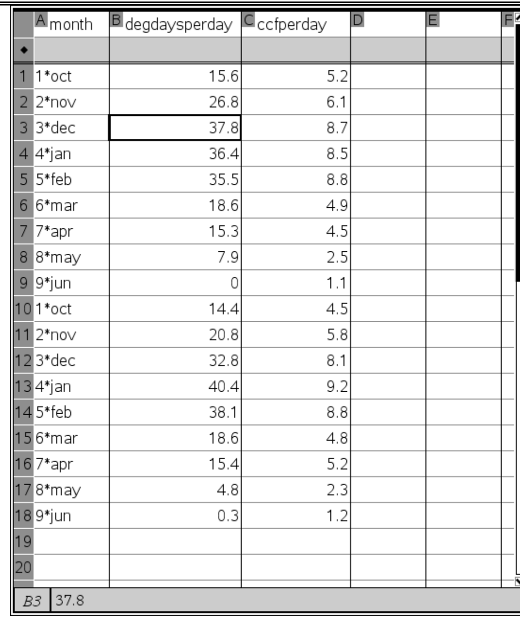
7. Problem 3

A home owner in Cary, NC is concerned with the amount of energy used to heat her home. She keeps a record of the natural gas consumed over a period of nine months over two years.

Because the months are not all equally long, she divides each month's gas consumption by the number of days in the month to get the cubic feet of gas used per day.

Then she goes to the weather service and records the number of degree-days for each month. (Degree-days are a measure of demand for heating: one degree-day is accumulated for each degree that a day's average temperature falls below 65 degrees F. See the article: http://en.wikipedia.org/wiki/Heating_degree_day) She divides this total by the number of days in the month giving the average number of degree-days per day.

Describe any trends you see in the data.
The numerical data indicates that more gas is used in the colder months, months during which the heating degree days are greater.



	A month	B degdaysperday	C ccfperday	D	E	F
♦						
1	1*oct	15.6	5.2			
2	2*nov	26.8	6.1			
3	3*dec	37.8	8.7			
4	4*jan	36.4	8.5			
5	5*feb	35.5	8.8			
6	6*mar	18.6	4.9			
7	7*apr	15.3	4.5			
8	8*may	7.9	2.5			
9	9*jun	0	1.1			
10	1*oct	14.4	4.5			
11	2*nov	20.8	5.8			
12	3*dec	32.8	8.1			
13	4*jan	40.4	9.2			
14	5*feb	38.1	8.8			
15	6*mar	18.6	4.8			
16	7*apr	15.4	5.2			
17	8*may	4.8	2.3			
18	9*jun	0.3	1.2			
19						
20						

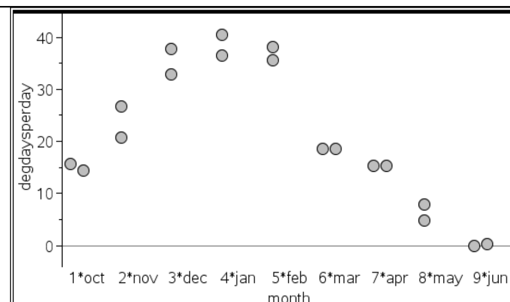
B3 37.8

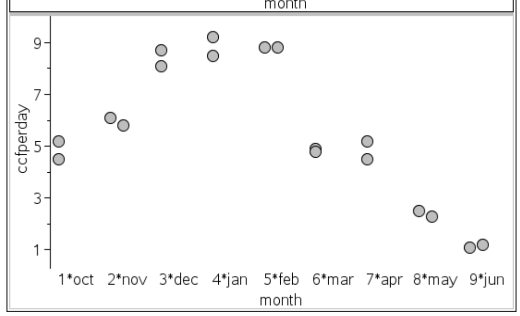
8. The data is organized by

(a) month and by "degreedaysperday" and
(b) by month and by "ccfperday".

Describe the trends you see in the two graphs?
The same as was noted in the numerical data; during November, December and January, when the number average number of degree days is highest, so is the amount of gas consumed.

(Note: you can probably see why the number was placed in front of each month – to order them. Otherwise, they would have been ordered alphabetically.)





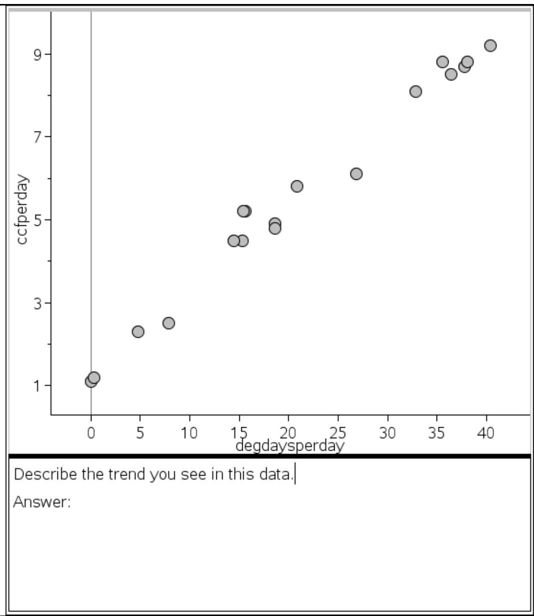
9. The pattern in each graph is similar. Each graph rises through the fall months and peaks in the winter months, falling through spring.

Because the two items, "ccfperday" and "degreedaysperday", are logically related, that is not surprising.

Explain the logic.
The colder the temperature the greater the degree days and the more gas that is needed to heat the house.

10. **The graph of "ccfperday" versus "degreedaysperday" is shown here. This graph shows the natural correlation between the two variables.**

    Describe the trend you see in the data.
    The points are fairly linear.



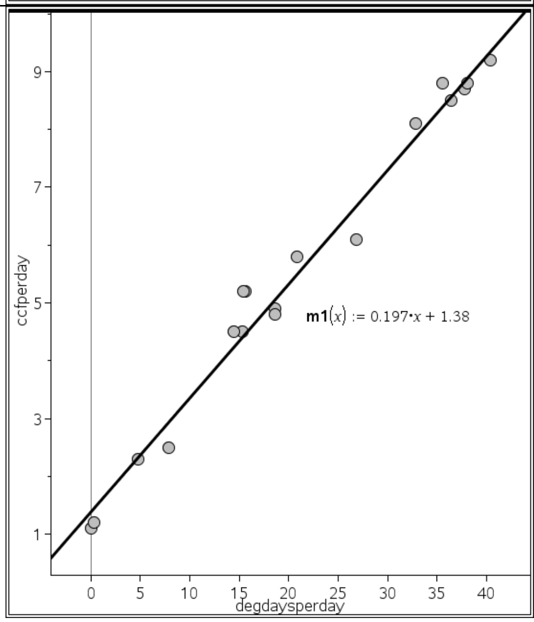
11. **Since the points are fairly linear, a line has been placed in the scatter plot.**

    Adjust the line so that it FITS the data as best you can.
    We are not after the line of best fit (regression line) but a line that looks like it is a reasonable fit.

    The relationship between the two variables is described by the formula of the line.

    Write this relationship using the variables (not x and m1(x)) and the numbers shown in the model.
    The number of cubic feet of gas used per day is a function of the number of heating degree days on that day such that the ccf is 1.97 times the number of heating degree days plus 1.38.

    Ask what .197 means. What are the units on .197? (The number of ccf of gas needed goes up by .197 for every unit change in the number of heating degree days; ccf/hdd.)
    Ask what 1.38 means and what are the units? 1.38 could be the number of ccf used even if heating wasn't needed; water heater, for example. Or it could just be due, in part, to the variability inherent in the model.)



12. **Use the model found in the previous page to predict the ccfperday for 40 degreedaysperday. (The graphical representation of data is just as much a model as a table of values or equation)**

    Explain your reasoning and your calculations.
    m1(40) = .197(40)+1.38 = 9.26

    How close do you believe your estimate will be?
    Because the points are reasonably close to the line (within .25 ccf) the approximation should be within .5 of the actual number of ccf required.

    What factors may influence the accuracy of the prediction?
    There are many factors that might influence the estimate: number of people in the house, the amount of sunshine, the amount of wind, whether it is a holiday or not, etc.

Problem 4

Thus far we have seen data that shows a trend; easily seen numerically and graphically.

We have also looked at a graphical display of data for which the numerical display was not so easily analyzed.

Last, we looked at data that showed a correlation between two variables. We saw the same trend in the data by month (an upside down U) and then saw an almost linear pattern that showed that as one variable increased the other did as well. However, the data was affected by many variables. If this model is used to make predictions it could not be relied on to entirely accurate.

Now we look at one more set of data.

13. This data shows the increase in the length of a spring as more weight is placed as a load on the spring.

	A weight	B length	C	D	E	F	G
◆							
1	20	4					
2	30	6					
3	40	8					
4	50	10					
5	60	12					
6	70	14					
7	80	16					
8	90	18					
9	100	20					
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

A1 20

The weight is in grams and the length in cm.
Ask you students about these units of measure.

What can you determine from the numerical display of the data and what do you expect to see in the graphical display?
There is a constant change in the length for a consistent change in the weight. Thus, we expect to see a linear relationship.

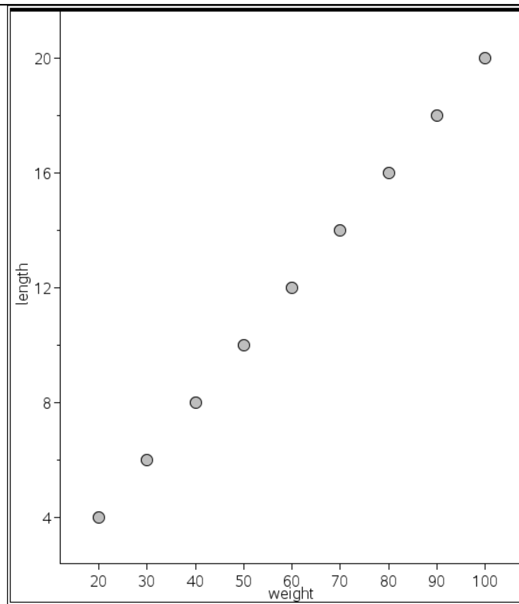
a) Predict the length of the spring when a load of 55 grams is placed on the spring.
11 cm.

How accurate is the prediction?
The prediction should be very accurate.

b) Predict the length of the spring when a load of 55 pounds is placed on the spring.
This value is beyond the limits of the spring (most likely).
Talk about why restrictions on the domain might occur.

How accurate is the prediction?

14. Create a scatterplot of the data with "weight" along the horizontal axis and "length" along the vertical axis.

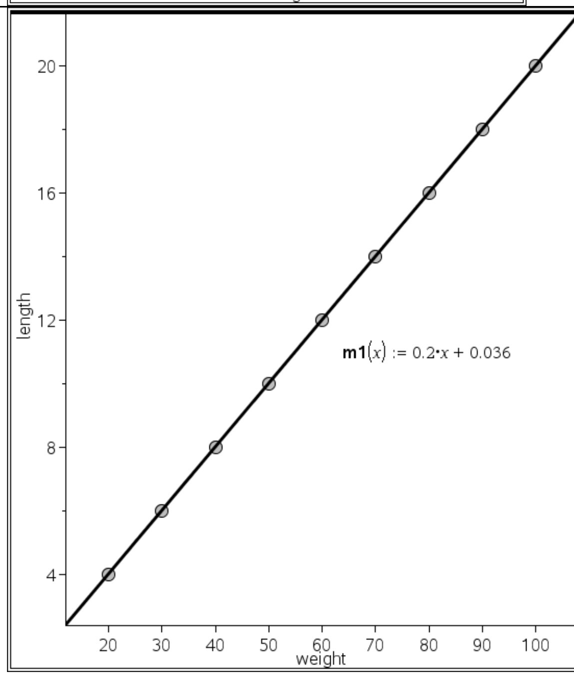


a) Why do you think "length" should be the dependent variable in this graph?
The weight will change the length so the length of the spring is dependent upon the load.

b) Describe the trend you see in the data when it is presented as a scatterplot.
The points are exactly linear.

15. A moveable line has been placed in the scatterplot.

Move the line to fit the data. (The equation of the line changes as you move it.)



a) Describe the model (the equation) using the names of the variables and not x and m1(x).
The length of the spring is a function of the load placed on it such that the length is .2 times the weight plus .036.

Ask what the .2 and .036 mean. (We would expect the constant term in the model to be zero, no weight

b) Explain why you believe predictions made with this model are more accurately than the predictions made with model found with degree days and gas use.
All the points are on the line so there is no variability. That is, the only factor that influences the length of the spring is the load on the spring.

Problem 5
Extension and Summative Assessment

Businesses are interested now, more than ever, in conserving the earth's natural resources and saving money. The recycling industry has grown to meet this demand. One large company produces a product that generates a great deal of scrap sheet metal. As they cut sheet metal to form boxes for electrical devices, a great deal of scrap is produced.

The metal strips vary from about 10 square inches to about 40 square inches in area. Each day, these strips are placed in a large 55-gallon drum. At the end of the workweek, the five drums have to be moved, weighed, recorded, and then sent to shipping where they are then picked up by a recycling business. The drums are hardly ever completely filled to capacity, since each drum is the result of one day's work.

	A height	B weight	C	D	E	F	G
◆					=LinRegM		
1	35	800		Title	Linear R...		
2	22	525		RegEqn	m*x+b		
3	42	975		m	22.2065		
4	41	950		b	31.8222		
5	24	550		r²	0.99511		
6	45	1000		r	0.997552		
7	41	950		Resid	{-9.0505...		
8	37	850					
9	35	800					
10	28	650					
11	42	975					
12	39	900					
13	41	950					
14	45	1025					
15	31	725					
16	37	850					

E1 ="Linear Regression (mx+b)"

Extension

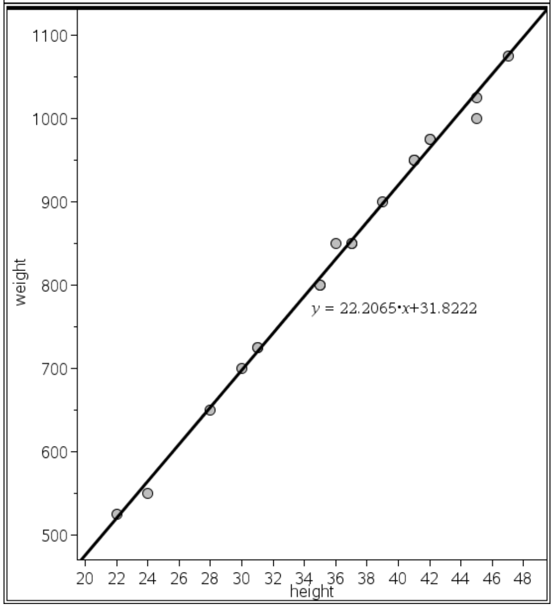
Scrap metal in a barrel: height in inches and weight in pounds.

It took a great deal of time and energy at the end of each week to weigh the barrels of sheet metal. One astute supervisor believed the weight could be estimated from the height of the sheet metal in the barrel. An estimate, say within 50 pounds, was good enough since the price per pound was so low.

They began to collect data to determine a reasonable model.

The data, collected over four weeks, is shown in the spreadsheet.

Suppose you are the supervisor. You are to write a report to management showing that the data you collected demonstrates that a model can be used to predict the weight of scrap metal in the barrel. Use graphics to support your conclusions.



The model predicts that the weight of the barrel can be predicted by the formula
Weight = 22.2 * Height + 31.8
The correlation between these two variables is very high so the predictions should be fairly accurate; the average error in the 20 data points was only 8.14 pounds.