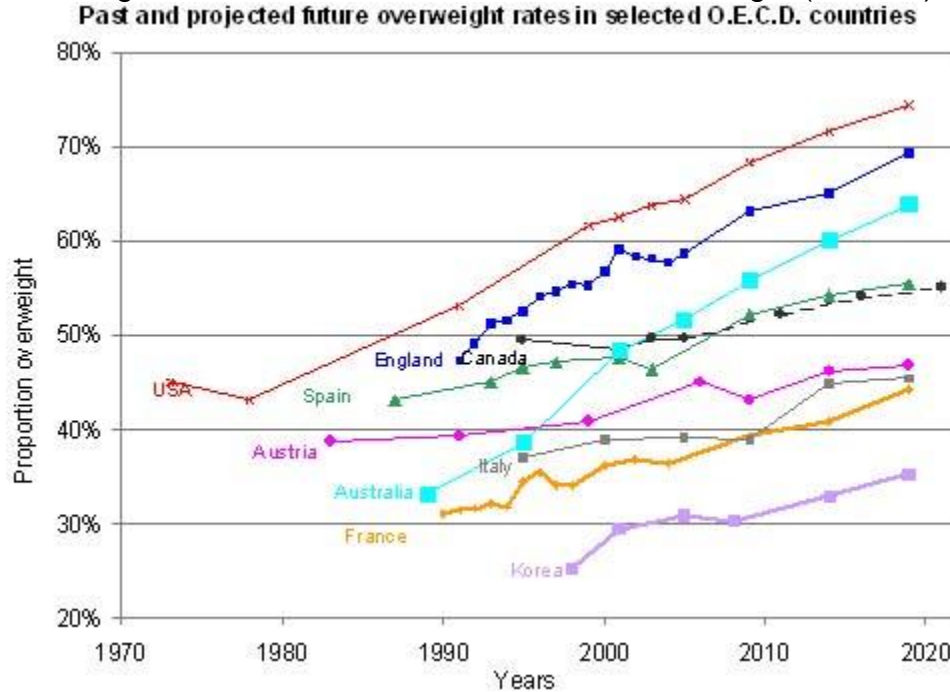


2 – Bi-Variate Data Analysis

Problem set 2-1

1. The graph below shows the percentage of the population in various countries that are overweight; this includes both those that are overweight (BMI>25) and obese (BMI>30).

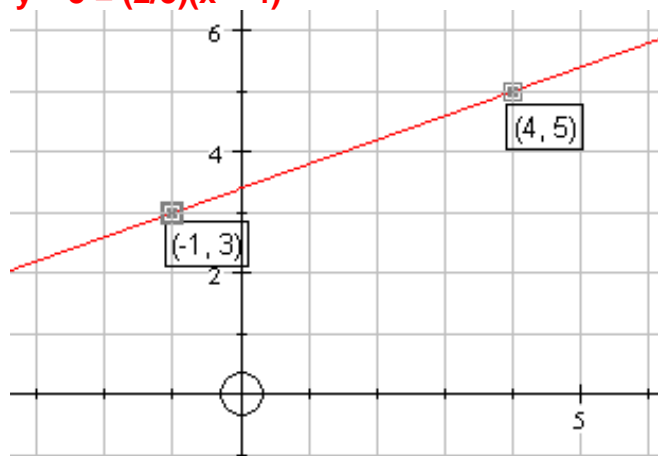


Source: *Global Nutrition Market, Obesity and World Health*. Global Sherpa, n.d. Web. 25 Feb. 2013.
<<http://www.globalsherpa.org/nutrition-market-obesity-malnutrition>>.

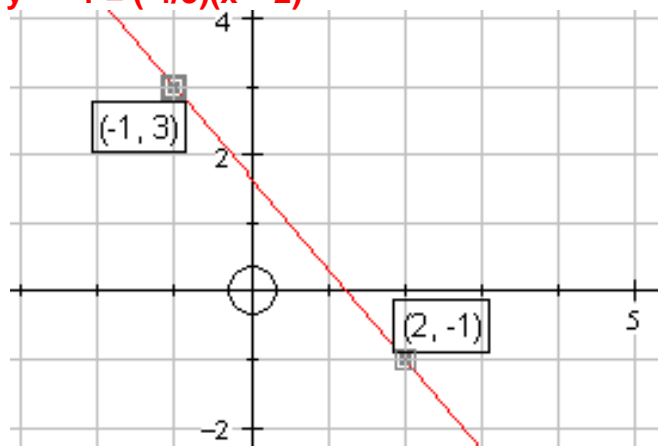
- a) The graph is cut off at about 2013. Based on the portion of the graph that is shown, about what percentage of the US population do you think will be overweight in 2020?
About 74% (Note: the portion of the graph from 2013-2023 is now revealed above.)
- b) What percentage of the US population do you think will be overweight in 2100? Show how you arrived at your answer and analyze it in the given context. **Since 1980, every 10 years the proportion of the US population that is overweight has gone up by about 7.5%. If the overweight population continues to grow linearly, then in 2100 the proportion would be 75% (10*7.5%) more than the proportion in 2000. That would make the proportion about 138%. Of course the proportion cannot be more than 100% and would unlikely ever be 100%. The overweight proportion of the population would have to level off (or decrease) at some point.**

In 2-4 below, find an equation for the line through the two given points in $y - k = a(x - h)$ form

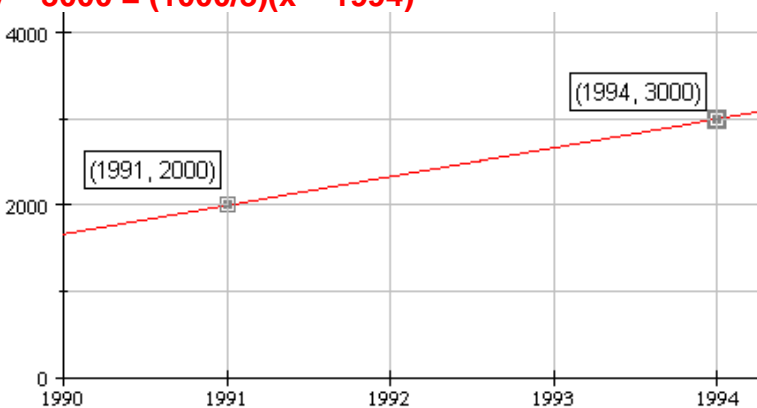
2. $y - 5 = (2/5)(x - 4)$



3. $y - -1 = (-4/3)(x - 2)$

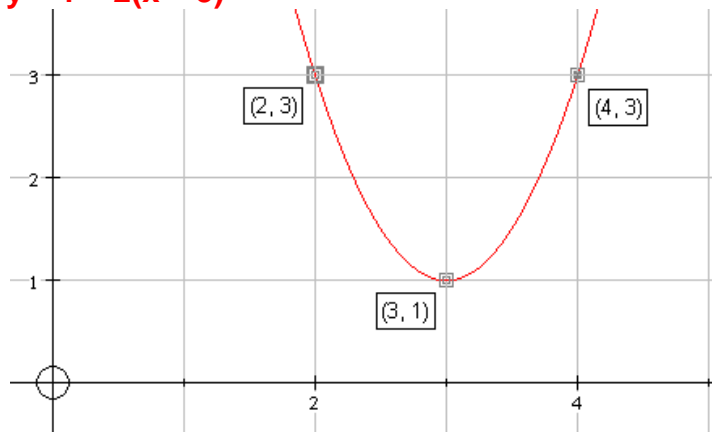


4. $y - 3000 = (1000/3)(x - 1994)$

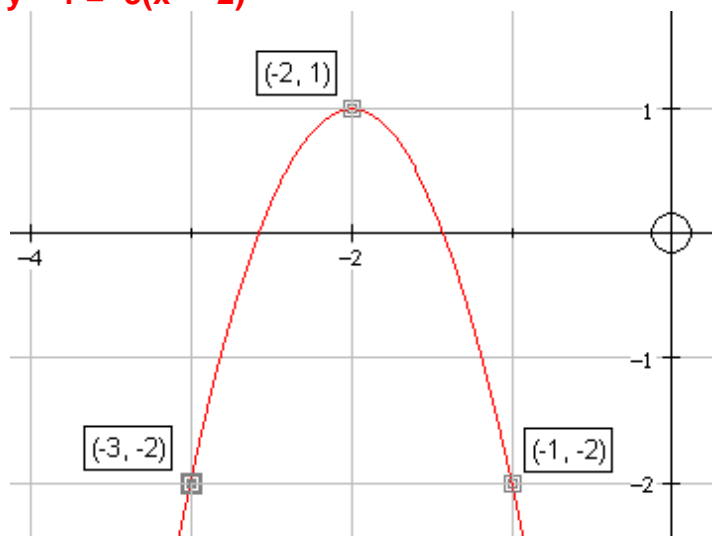


In 5-7, below, find an equation for the parabola that goes through the three given points in $y - k = a(x - h)^2$ form.

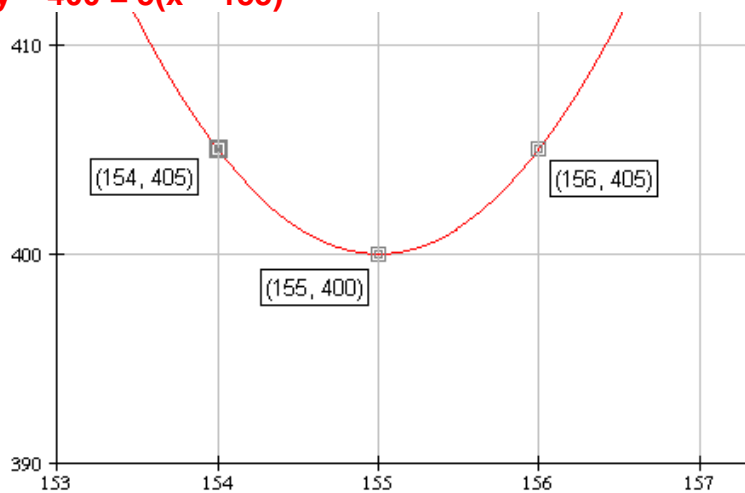
5. $y - 1 = 2(x - 3)^2$



6. $y - 1 = -3(x - 2)^2$



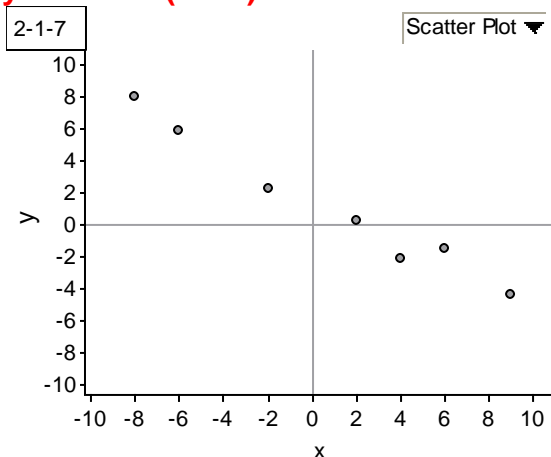
7. $y - 400 = 5(x - 155)^2$



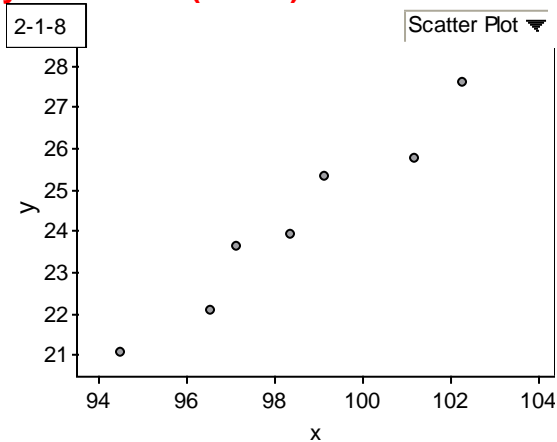
In 8-11, find an equation of a line in $y - k = a(x - h)$ form or a parabola in $y - k = a(x - h)^2$ form that models the given scatterplots.

Answers below are sample answers. Your answer will depend on which anchor point and rise and run you chose.

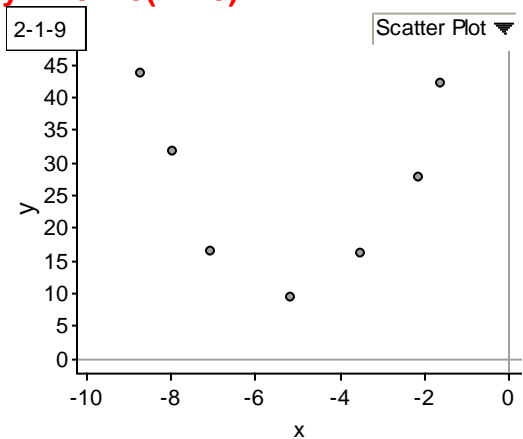
8. $y - 2 = -0.6(x - 0)$



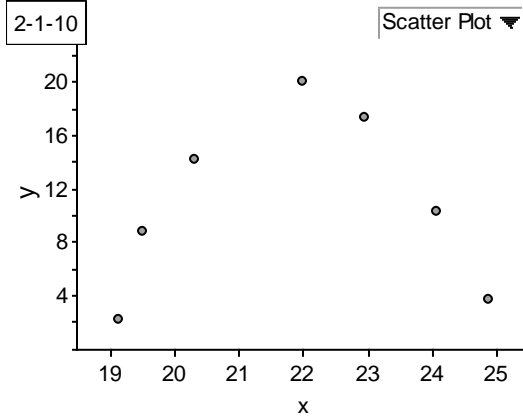
9. $y - 25 = 0.75(x - 99)$



10. $y - 10 = 3(x + 5)^2$



11. $y - 20 = -2(x - 22)^2$



12. Consider the scatterplot below of the height (in cm) vs. age (in months) of children in Kalama, Egypt.

a) Find an equation of a line in $y - k = a(x - h)$ form that models the scatterplot below.

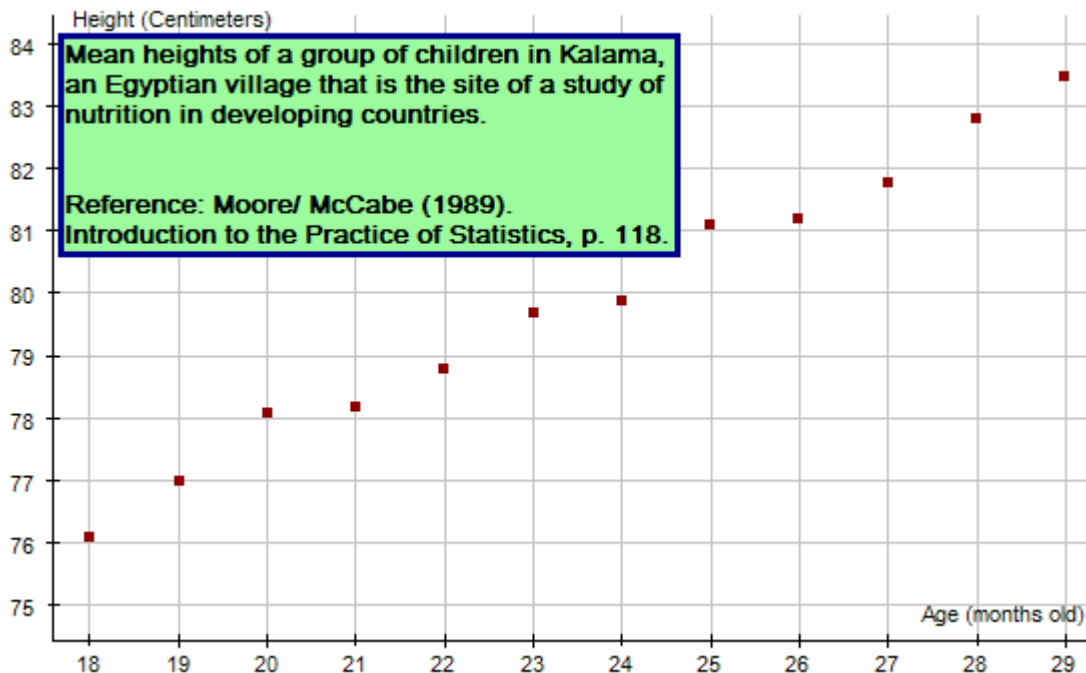
Remember, this answer is a sample: $y - 80 = 0.65(x - 24)$

b) Write a complete sentence that gives the meaning of the slope, a, in the context of the data set.

Slope: The slope of the given linear model is 0.65; so according to this linear model, we would expect an average increase of about 0.65 centimeters of height for each additional month of age for children in Kalama.

c) Write a complete sentence that gives the meaning of the point (h,k), in the context of the data set.

Point: This model goes through the point (24, 80); so according to this linear model a 24-month-old child in Kalama would be approximately 80 cm tall.

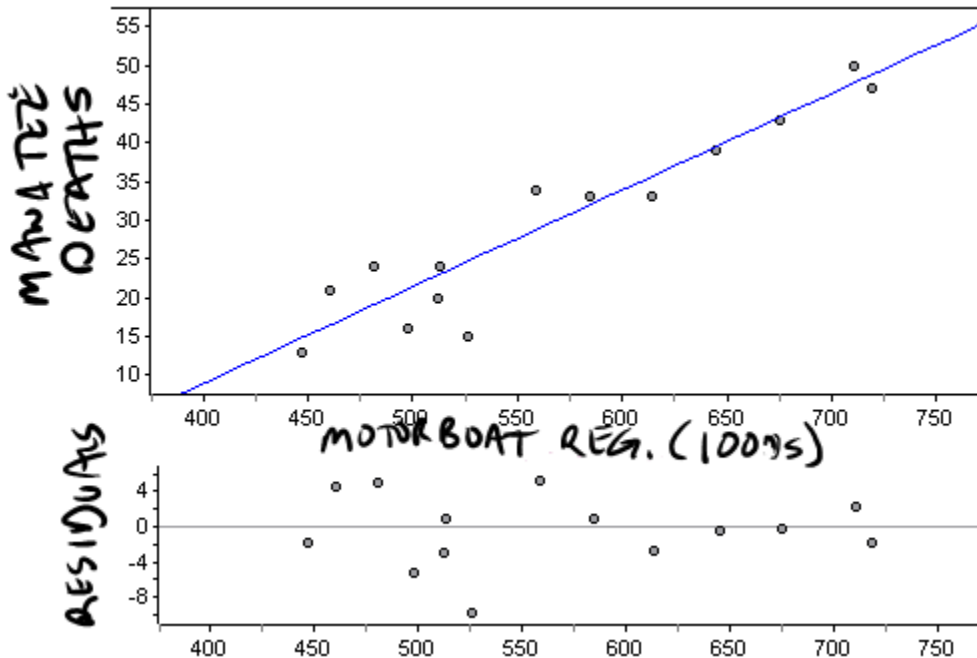


Problem set 2-2

1. Data Set: Manatee Deaths

Use Motorboat registration (thousands) as the independent variable and manatee deaths as the dependent variable.

a) "Analyze the data".



Model: $y = 0.125(x-593)+33$

x is Motorboat registrations (in thousands)

y is Manatee deaths

Comment on the residuals:

Pattern: No clear pattern

Magnitude: Fairly small


Comment on the model: The model is an appropriate model

b) Write a sentence that explains the meaning of the slope of your linear model within the context.

The slope of the linear model is 0.125; so, according to this linear model, we would expect an average increase of about 0.125 manatee deaths in a year for each additional one thousand motorboats registered in Florida.

c) Write a sentence that explains the meaning of the point (h,k) in your linear model within the context.

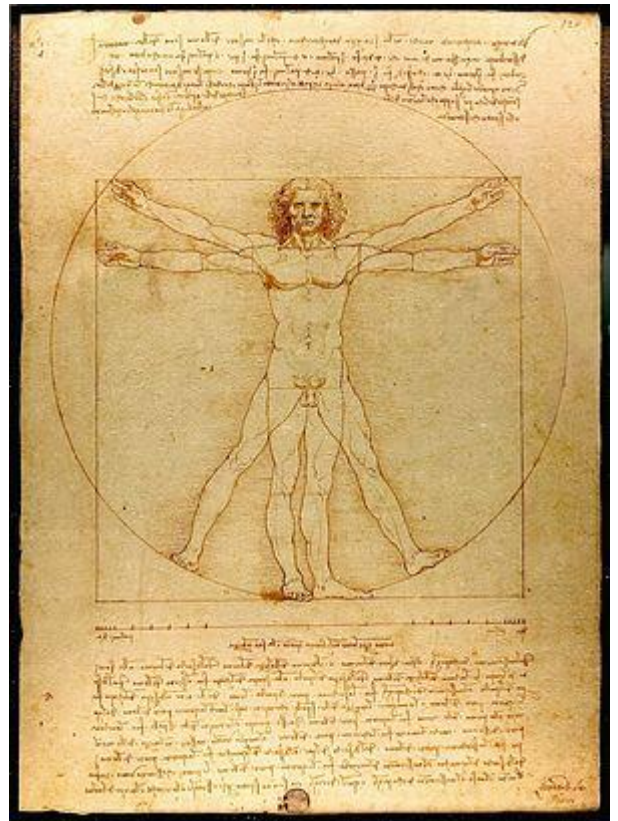
The model goes through the point (593, 33); so, according to this linear model when 593,000 motorboats were registered, there would be 33 manatee deaths in a year.

2.  Vitruvian Man (also called Canon of Proportions) is a famous drawing by Leonardo da Vinci that shows the proportions of man based on the Vitruvian text, “*On Symmetry: in Temples and in the Human Body*”. The class will collect an ordered pair of measurements for every student. The independent variable will be arm span in cm (to the nearest tenth). The dependent variable will be height (without shoes) in cm (to the nearest tenth).

- a) “Analyze the data”.
- b) Write a sentence that explains the meaning of the slope of your linear model within the context. **The slope of the linear model is ** (answers may vary); so, according to this linear model, we would expect an average increase of about ** cm in height for each additional one cm in arm span.**
- c) Write a sentence that explains the meaning of the point (h,k) in your linear model within the context.


The model goes through the point (##,&&)(answers may vary); so, according to this linear model when someone has an arm span of ## cm, we would expect their height to be approximately && cm.

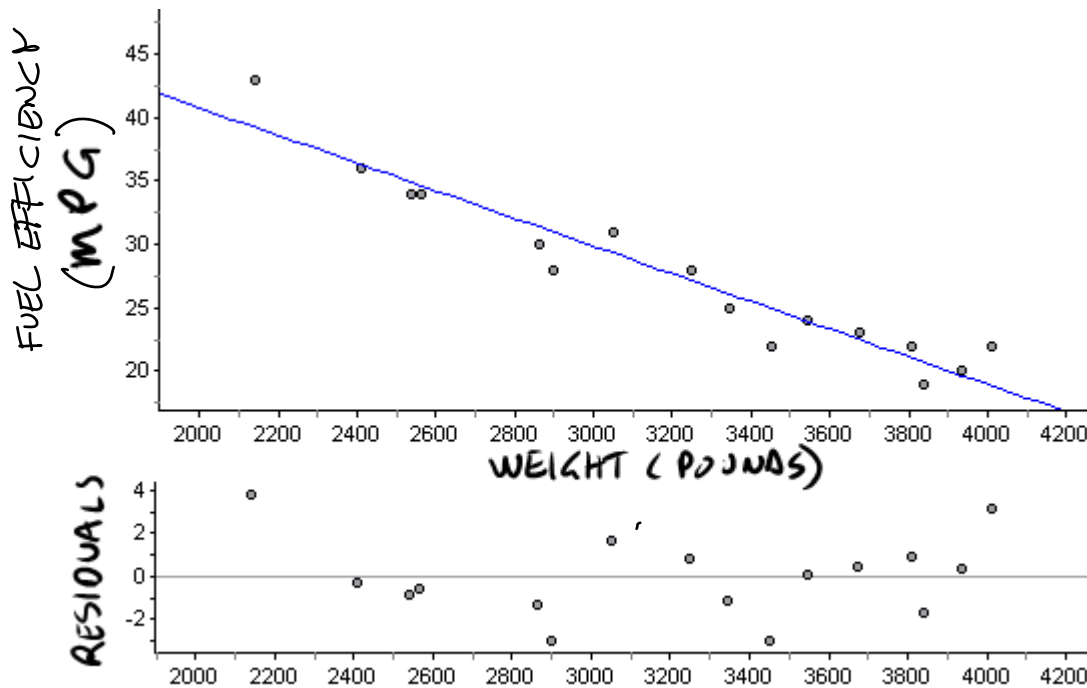
Ask your teacher for her/his arm span. Use your model (the equation, not the graph) to predict your teacher’s height.



On Symmetry: In Temples and in the Human Body. Vitruvius, n.d. Web. 24 Apr. 2013.
<http://faculty.txwes.edu/csmeller/human-experience/ExpData09/02GrecoRoman/GrRmRDs/rmR_Vitruvius-Sym.html>.

Image source: *Vitruvian Man*. Gallerie dell'Accademia, n.d. Web. 4 Feb. 2013.
<<http://www.gallerieaccademia.org/wp/wp-content/gallery/leonardo-luomo-vitruviano-fra-arte-e-scienza/leonardo.png>>.

3.  Data Set: Car MPG vs. Weight
 a) "Analyze the data".



Model: $y = -0.011(x-3208)+27.6$
 x is Weight (pounds)
 y is Gas mileage (MPG)

Residuals:
 Pattern: No clear pattern
 Magnitude: Small

Comment on the model: The model is an appropriate model

- b) Write a sentence that explains the meaning of the slope of your linear model within the context. **The slope of the given model is -0.011; so according to this linear model, we would expect an average decrease of about 0.011 gas mileage in MPG for each additional one pound of weight of the car.**
- c) Use your model to predict what the Fuel Efficiency (MPG) would be for a car weighing 2,700 pounds.
 $y = -0.011(2700-3208)+27.6 = 33.188$
According to the given model, a car weighing 2,700 pounds would get about 33 mpg.
- d) According to the slope of your model, what is the effect on fuel efficiency of taking on a passenger in your car that weighs 175 pounds?
 $-0.011 \text{ mpg/lb} * 175 \text{ lbs} = -1.925 \text{ mpg}$
According to the model, taking on a passenger weighing 175 decreases your fuel efficiency by 1.925 mpg.

4. 🖨️ Data Set: CO₂ Annual 2000-2007

Since the beginning of the industrial revolution there has been an ever increasing concentration of carbon dioxide (CO₂) in our atmosphere. High levels of CO₂ in our atmosphere result in a greenhouse effect which has led to global warming. CO₂ concentration in air is usually measured in parts per million by volume (ppmv). Just prior to the industrial revolution CO₂ concentrations were about 280 ppmv. By 2008 CO₂ concentrations reached about 385 ppmv. There is a growing consensus among scientists that specialize in global warming that 450 ppmv is a threshold above which disastrous, irrevocable changes would affect the planet.

a) “Analyze the data”.

b) Write a sentence that explains the meaning of the slope of your linear model within the context.

The slope of the given linear model is ** (answers will vary); so according to this linear model, we would expect an average increase of about ** ppmv of CO₂ each year.

c) According to your model, in what year will CO₂ concentration reach 450 ppmv?

About #### (answers will vary)


Sources:

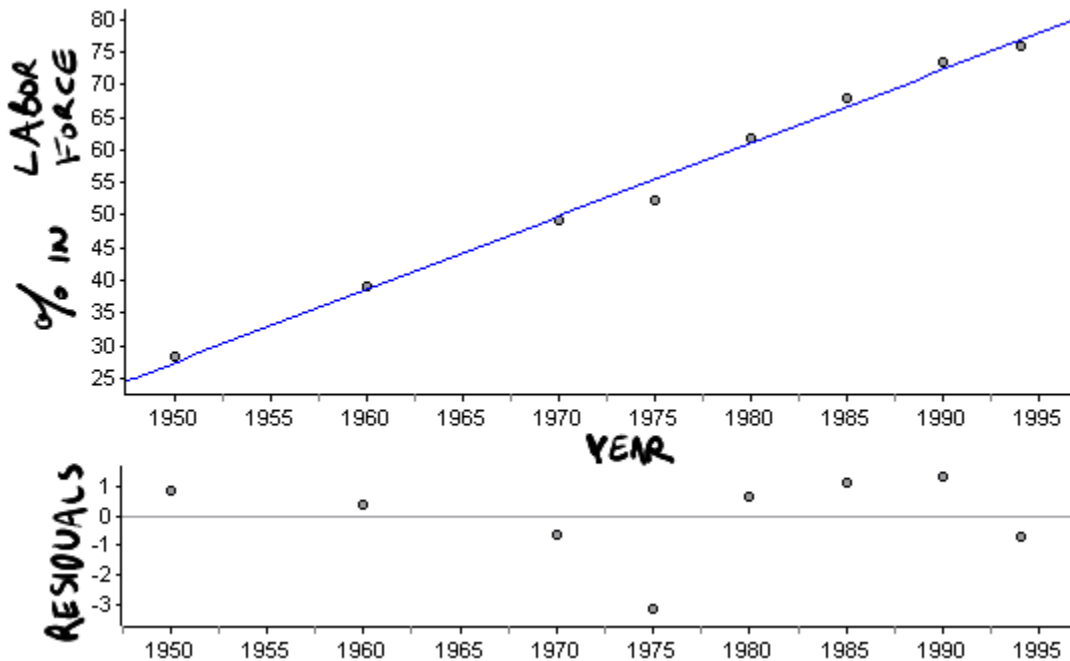
Fight Global Warming, *Global Warming: Facts, Consequences, and Solutions*, <http://www.fightglobalwarming.com/content.cfm?contentID=5113>, 2/21/2009.

Inter Press Service, *CLIMATE CHANGE: Oceans Passing Critical CO₂ Threshold*, <http://ipsnews.net/news.asp?idnews=44836>, 2/21/2009.

The Christian Science Monitor, *A Key Threshold is Crossed*, <http://www.csmonitor.com/2007/1011/p11s01-wogi.html>, 2/21/2009.

Problem set 2-3

1.  Data Set: Percentage of American Women in the Labor Force with Children
a) "Analyze the data".



Model: $y = 1.12(x-1975.5)+56.0$

x is Year

y is Percentage of American women in labor force

Residuals:

Pattern: No clear pattern

Magnitude: Very small

Comment on the model: The model is a very appropriate model

- b) According to your model, what percentage of American women that have children ages 6 to 17 will be in the labor force in the year 2030? Is it a realistic answer? If you said your model was an appropriate model then what went wrong?

$1.12(2030-1975.5)+56.0=117.04$, so according to my model about 117% of American women that have children ages 6 to 17 will be in the labor force in 2030. A maximum of 100% of women could be in the labor force, so the percentage would have to level off (or decrease) at some point.

- c) Write a sentence that explains the meaning of the slope of your linear model within the context.

Slope: The slope of the given linear model is 1.12; so, according to this linear model, we would expect on average 1.12% more women with children 6 to 17 to be in the labor force each year.


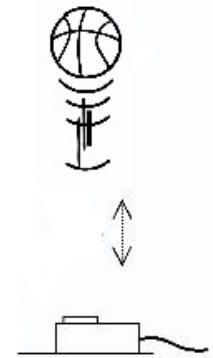
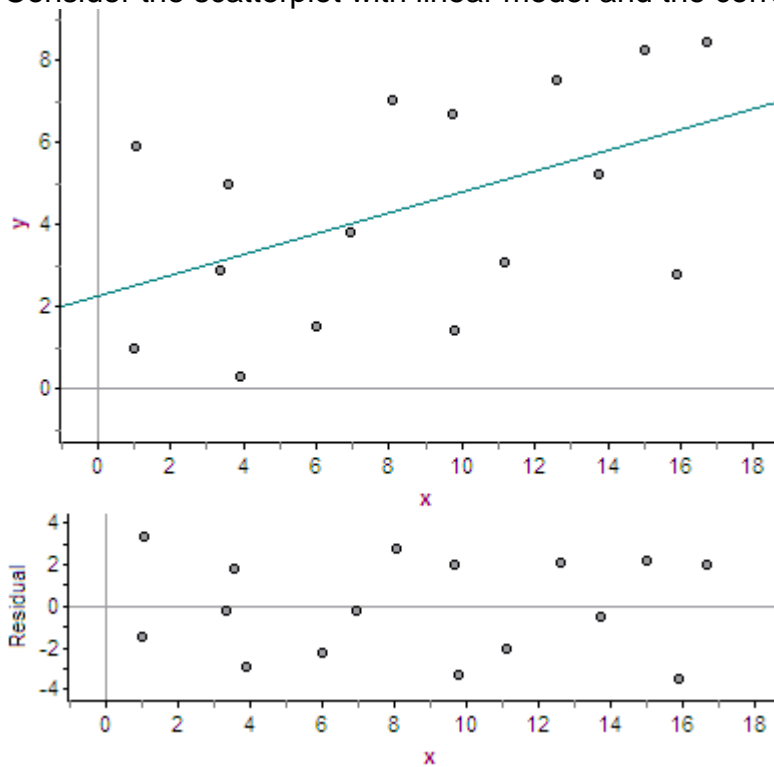
2.  Data Set: Ball Data (collected in class)
 “Analyze the Data”. Use time (since the ball was tossed upward) as the independent variable and height (vertical distance from the sensor) as the dependent variable.

Image source: Investigating a Mass on a Spring. Nuffield Foundation, n.d. Web. 22 Feb. 2013. <http://www.nuffieldfoundation.org/sites/default/files/images/Investigating%20a%20mass-on-spring%20oscillator_322.jpg>.



3. Consider the scatterplot with linear model and the corresponding residuals.



- a) Comment on the residuals.

Residuals:

Pattern: No clear pattern

Magnitude: large

- b) Comment on the appropriateness of the model.

Comment on the model:

The model is an appropriate model because there is no clear pattern to the residuals. The model is not useful, however, because the residuals are fairly large.

Problem set 2-4

No solutions are given for most of the rest of the chapter. Compare answers with your classmates before class or in class.

1. 📄 Data Set: Sheriff response time

A county sheriff has headquarters located five miles from highway mile marker 129. One day the sheriff recorded how long it took deputies to respond from headquarters to calls along the highway in the county. The table gives the mile marker from where the call was made and the corresponding response time in minutes.

a) “Analyze the data”

b) According to your model, how long would it take a deputy to respond to a call at mile marker 140?

2. 📄 Data Set: Car skids

Use Skid Length (ft) as the independent variable and Speed (MPH) as the dependent variable. “Analyze the data”.

Problem set 2-5

It is necessary to do all of Chapter 3 before continuing. When Chapter 3 is completed, repeat Problem Set 2-4 and then finish the chapter.

1. 🖨 Data Set: Pig weight gain. "A veterinarian was working for a large pig cooperative who was interested in increasing the weight of its pigs. Twenty-four randomly selected pigs were each given a daily dosage (in pellets) of a food supplement. Groups of three pigs each received the same dosage, and their percentage weight gain was averaged. The table given shows the average percentage weight gain in one month for each group of three pigs in relation to the dosage".

Functions, Statistics, and Trigonometry. 2nd ed. N.p.: SFAW, 1998. Print. University of Chicago School Mathematics Project. P 124

- a) "Analyze the data". Use dosage as the independent variable and weight gain as a percentage as the dependent variable.
- b) According to your model, what is the dosage that results in a maximum weight gain?
2. 🖨 Data Set: Pasta Data (collected in class)

"To estimate the amount of spaghetti to use for one serving, measure the amount you can enclose in the circle made by your thumb and forefinger".

BROADENING YOUR PASTA HORIZONS, Spike's Recipe Collection, <http://www.spike-jamie.com/recipes2/recipes230.html>, 10/30/04

Alternatively, you can use a pasta serving guide. The photo below is a guide for 1, 2, 3, and 4 servings of pasta. Image source: *Always Something Brilliant*, Measure Spaghetti, http://www.alwaysbrilliant.com/aa/measure/measure.aspx-products/pd-co098/bb/spaghetti_mesurer_spaghetti_portions.htm, 10/30/04

Collecting the data – In class you will use the spaghetti serving guide to make bundles of pasta that each represent one serving. Make 6 separate bundles such that each represents one serving. Measure the circumference of 1 serving (1 bundle) to the nearest millimeter. Because "error is inherent in measurement", measure the circumference three times and use the mean or median of the three measurements. Now add another bundle and measure the circumference of 2 servings. Keep adding bundles and measure the corresponding circumference and record the results in the table below.



- a) "Analyze the data".
- b) According to your model, if you wanted to make 100 servings, what would the bundle circumference be? How about 101 servings? Explain why these two circumferences are so close.
- c) If you had a bundle of spaghetti that had a circumference of 1,000 mm, how many servings would that represent?

Number of servings	Bundle circumference (mm)
1	
2	
3	
4	
5	
6	

Problem set 2-6

1. 🖨 Data Set: Nevada Population
 - a) “Analyze the data”.
 - b) According to your model, what was the population in 2000? Google “Census 2000 Data for the State of Nevada” to find out what the census population was in 2000.

2. 🖨 Data Set: Water Flow (data collected in class)

Put a piece of tape on the cylindrical portion of a 2-liter soda bottle and draw a mark every 10 mm from 0 mm at the bottom to 150 mm at the top. Drill a 5 mm (~3/16 inch) diameter hole next to the 0 mm mark. Fill the bottle to the top with water. Take off the cap and collect data using time since the water passed the 150 mm mark as the independent variable and height of the water in mm as the dependent variable. Collect one data point for each of the markings from 150 mm to 10 mm. Your first data point will be (0 sec, 150 mm) and the second will be something like (4 sec, 140 mm). You can use the on-line stop watch with “splits” at <http://www.online-stopwatch.com/split-timer/>.

- a) “Analyze the data”.
- b) Write a sentence that describes the relationship between time and water height.

The idea and for this problem come from:
The North Carolina School of Science and Mathematics. *Contemporary Precalculus Through Applications*. 2nd Edition, 1999 ed. N.p.: Everyday Learning Corp., n.d. Print.



3. 🖨 Data Set: Light intensity (data collected in class)

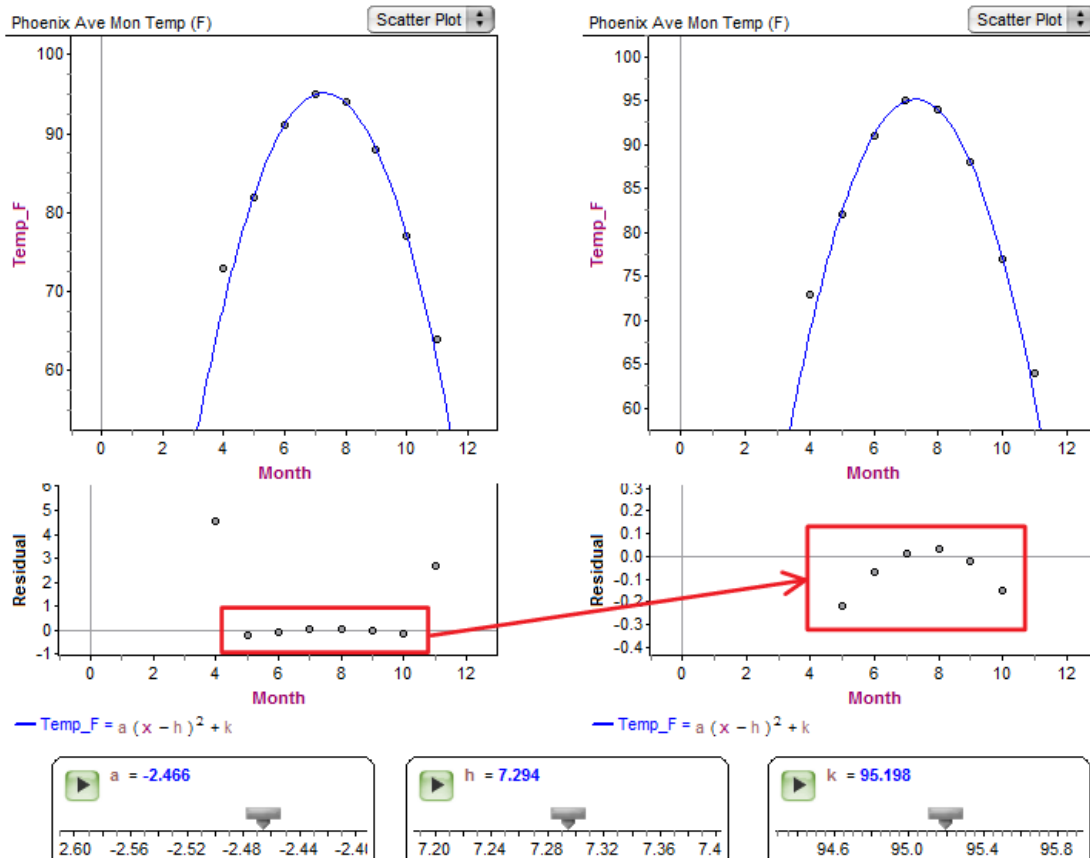
It makes sense that as you get further away from a light source the intensity of the light diminishes, but what is the relationship? The intensity of a light was measured in mW/cm^2 at distances from 120 cm to 210 cm away from the light source and the results are recorded in the table. (i)

 - a) “Analyze the data”.
 - b) According to your model, what would the intensity be when the light source is 220 cm away?

Problem set 2-7

1. 🖨 Data Set: Average Monthly Temperature for Phoenix AZ
 “Analyze the data”.

$y = -2.47(x-7.29)^2 + 96.20$



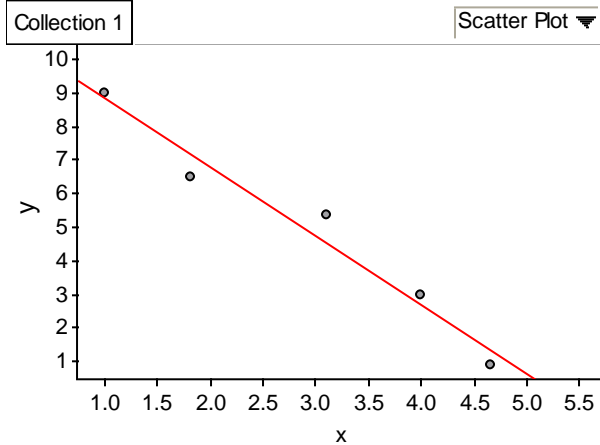
The residual plot on the left shows that two of the residuals, namely the first and last ones, are quite large. The residual plot on the right, which zooms in on months 5-10, shows a clear concave down pattern to the residuals. A quadratic model is not appropriate.

2. 🖨 Natural frequency of a tube (data collected in class)
 A tube has a natural frequency that is primarily dependent on the length of the tube. The data will be collected in class.
- “Analyze the data”.
 - Use your model to predict the natural frequency for a tube of length of 10 cm, 5 cm, and 0 cm. Based on your answers, would you say you have an appropriate model? If not, see if you can find a more appropriate model.
3. 🖨 Data Set: Car MPG vs. Weight
 The first time we analyzed the Car Weight vs. MPG data (2-2, #3), we created a linear model. If you consider end behavior a linear model seems inappropriate. “Analyze the data” again and find a more appropriate model.

Problem set 2-8

Statistical software packages like Fathom often give r^2 (or R^2) which is the Coefficient of Determination. The Coefficient of Determination tells the proportion of the variation of the dependent variable that is accounted for by a least-squares regression line. When Fathom gives r^2 , $\sqrt{r^2}$ gives $|r|$ and the slope of the least-squares line tells us the sign of r , that is, whether r is positive or negative.

1. Find r for the following scatter plot.



$$y = -2.03x + 10.85; r^2 = 0.96$$

In 2 – 4, complete the sentence.

2. If r (the correlation coefficient) is positive then, ...
3. If r^2 (the coefficient of determination) is 0.99 then, ...
4. If your statistics software gives you $r^2 = 4$, ...
5. From 2006-2012, monthly averages are available for the following two variables:

- The number of times “facebook,” was Googled
- US unemployment rates as a percentage of the total population

When a linear regression is performed on the data for these variables, $r = 0.9872$.

- a) What is the meaning of r in this case?

There is a very strong positive linear correlation between the number of times “facebook,” was Googled and US unemployment rates.

- b) Did people Googling “facebook,” **cause** unemployment? Explain.

NO! Correlation does not imply causation.

Source: *Google Correlate*. Google, n.d. Web. 5 Feb. 2013. <

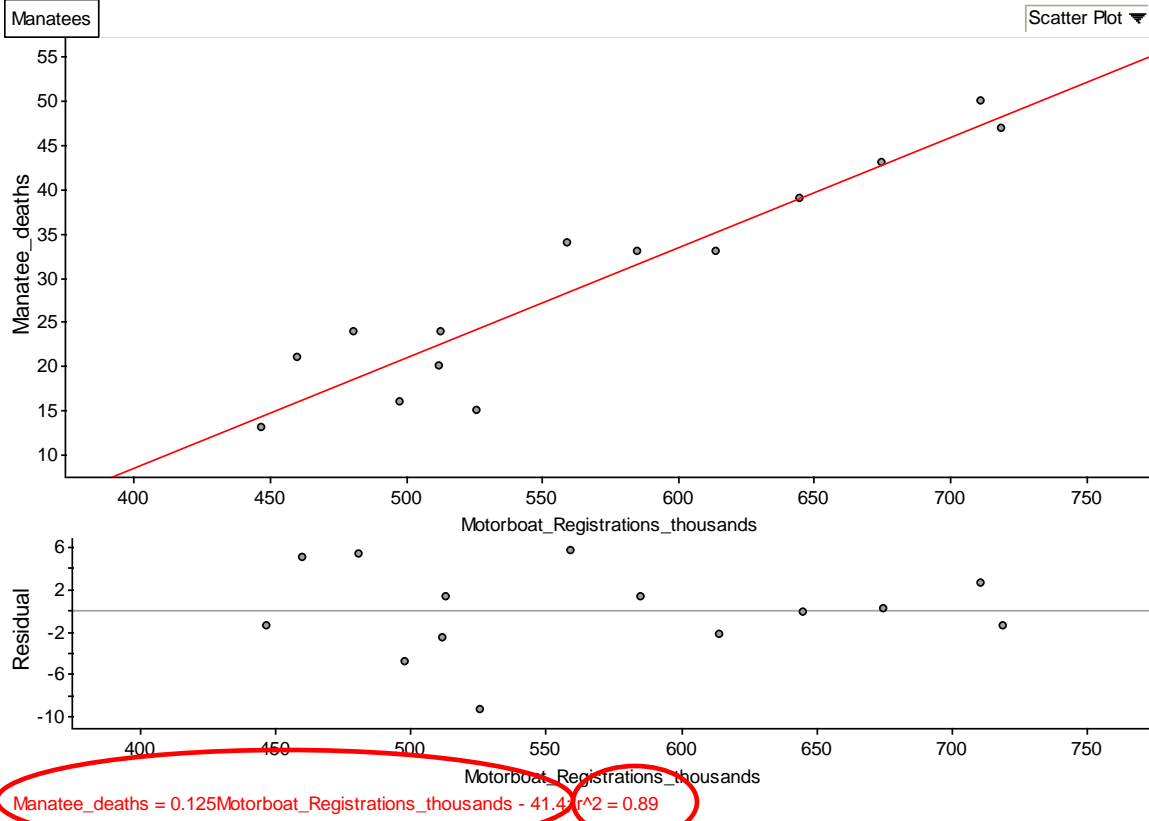
<http://www.google.com/trends/correlate/search?e=id%3ANWRPEIKPIBC&e=facebook%2C&t=monthly&p=us> >.

Source: *US Unemployment Rate*. Bureau of Labor Statistics, n.d. Web. 5 Feb. 2013.

<<http://data.bls.gov/timeseries/LNS14000000>>.

6. For the following data sets give:
- the least-squares line
 - the meaning of the slope and (if possible) y-intercept within the context of the data set
 - the correlation coefficient
 - What information does the correlation coefficient provide within the context of the data set?

a)  Manatee deaths



The equation of the least squares line is:
 $y = 0.125 * x - 41.4$

The meaning of the slope – The slope of the least squares line is 0.125 which means that according to the least squares line there is an increase of 0.125 manatee deaths per 1000 motorboats registered (or an increase of 1.25 manatee deaths per 10,000 motorboats registered) per year in Florida.

The meaning of the y-intercept – The y-intercept of the least squares line is -41.4 which means that according to the least squares line when there were 0 motorboats registered, there were -41.4 manatee deaths per year. This does not make sense, because it is not possible to have negative manatee deaths.

The correlation coefficient is $\sqrt{0.89} \approx 0.94$

The meaning of the correlation coefficient within the context – There is a very strong positive linear correlation between manatee deaths and motorboat registrations.

This is last question is not asked in the problem set, but...

What does the least squares line predict the number of manatee deaths will be if the number of motorboat registrations goes up to 750,000?

$$\text{Manatee deaths} = 0.125 \cdot (750) - 41.4$$

$$\text{Manatee deaths} = 52.35$$


The least squares model predicts there would be about 52 manatee deaths per year.

b)  Car MPG vs. Weight

c)  Nevada Population

7.  Data Set: Heart Attacks versus Wine Consumption

Use alcohol from wine as the independent variable and heart disease deaths as the dependent variable. Is the correlation positive or negative? Explain why. Is the correlation strong or weak? Explain why.

8.  Data Set: CO₂ Annual 2000-2007

a) What is the correlation coefficient?

b) What is the meaning of the correlation coefficient?

c) Comment on the residuals.

d) Comment on the appropriateness of the model.