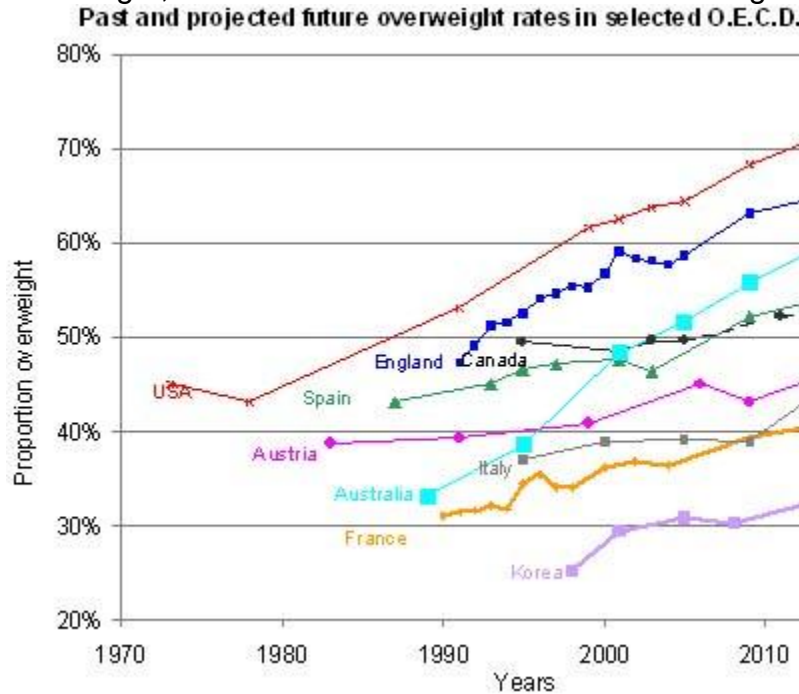


2 – Bi-Variate Data Analysis

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).

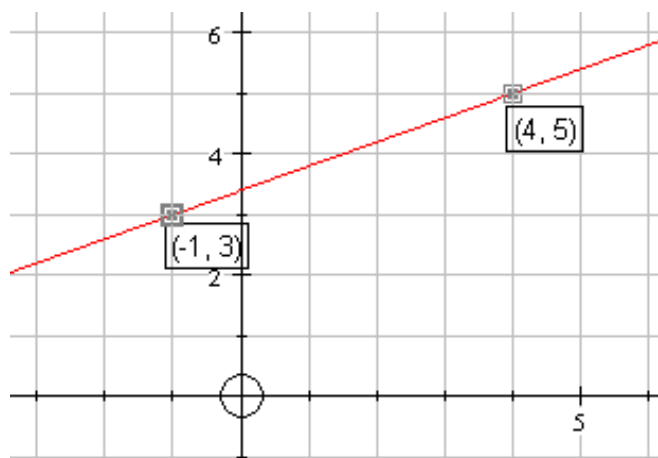


Source: *Global Nutrition Market, Obesity and World Health*. Global Sherpa, n.d. Web. 25 Feb. 2013.
<<http://www.globalsherpa.org/nutrition-market-obesity-malnutrition>>.

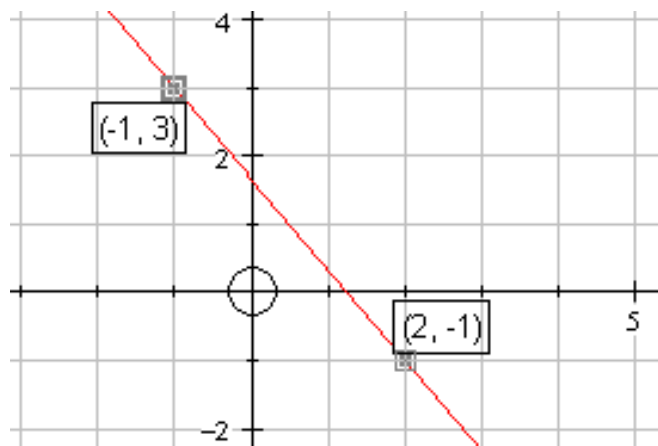
- a) The graph is cut off at about 2013. Based on the portion of the graph that is shown, about what percentage of the US population do you think will be overweight in 2020?
- b) What percentage of the US population do you think will be overweight in 2100? Show how you arrived at your answer and analyze it in the given context.

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

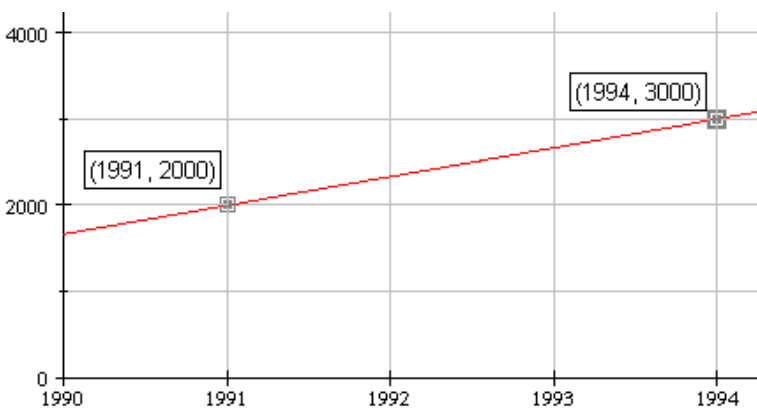
2.



3.

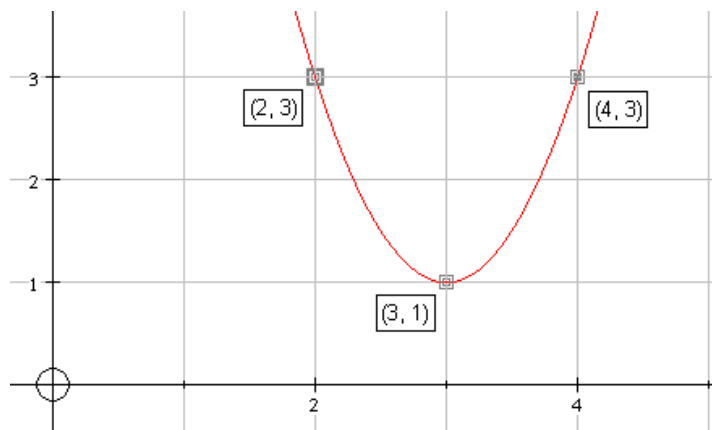


4.

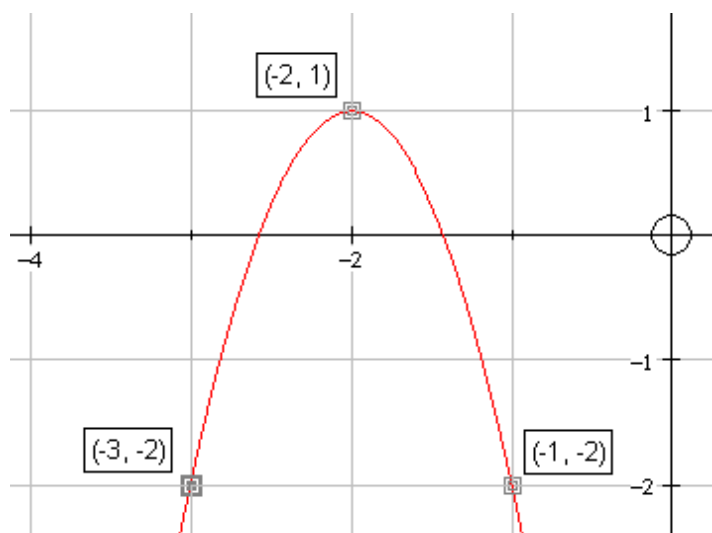


In 5-7, below, find an equation for the parabola that goes through the three given points in $y - k = a(x - h)^2$ form.

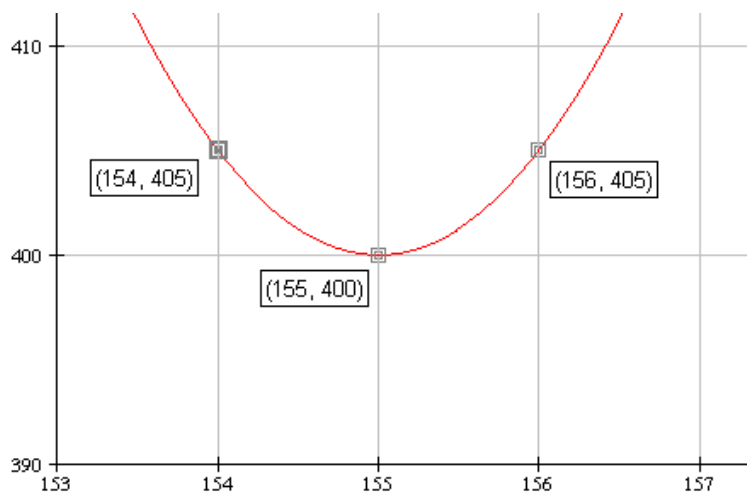
5.



6.

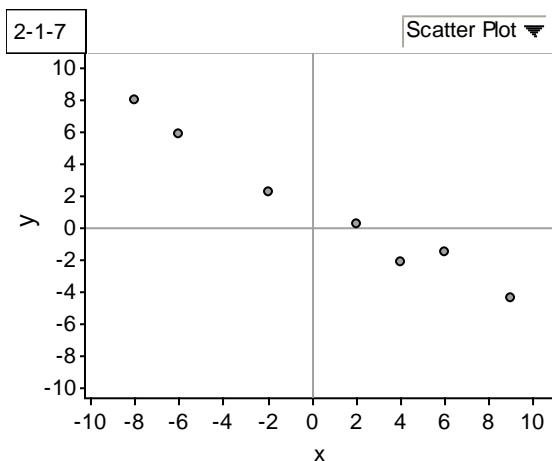


7.

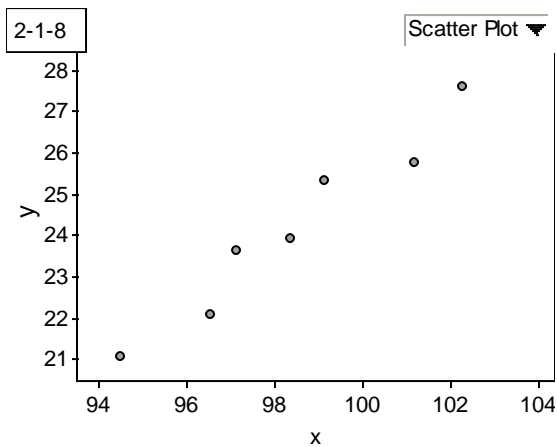


In 8-11, find an equation of a line in $y - k = a(x - h)$ form or a parabola in $y - k = a(x - h)^2$ form that models the given scatterplots.

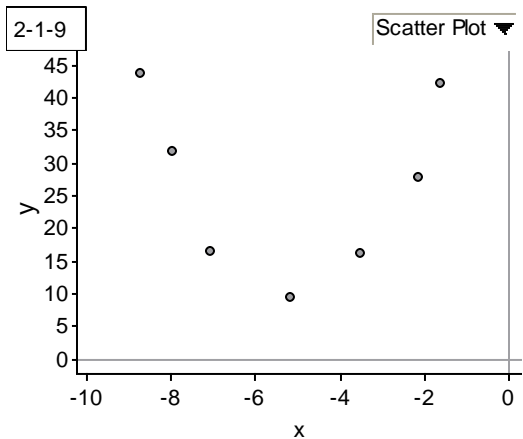
8.



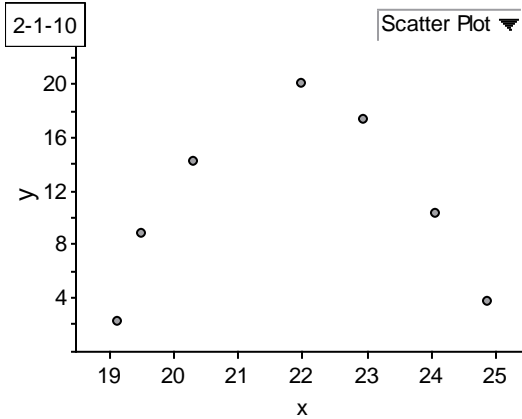
9.



10.

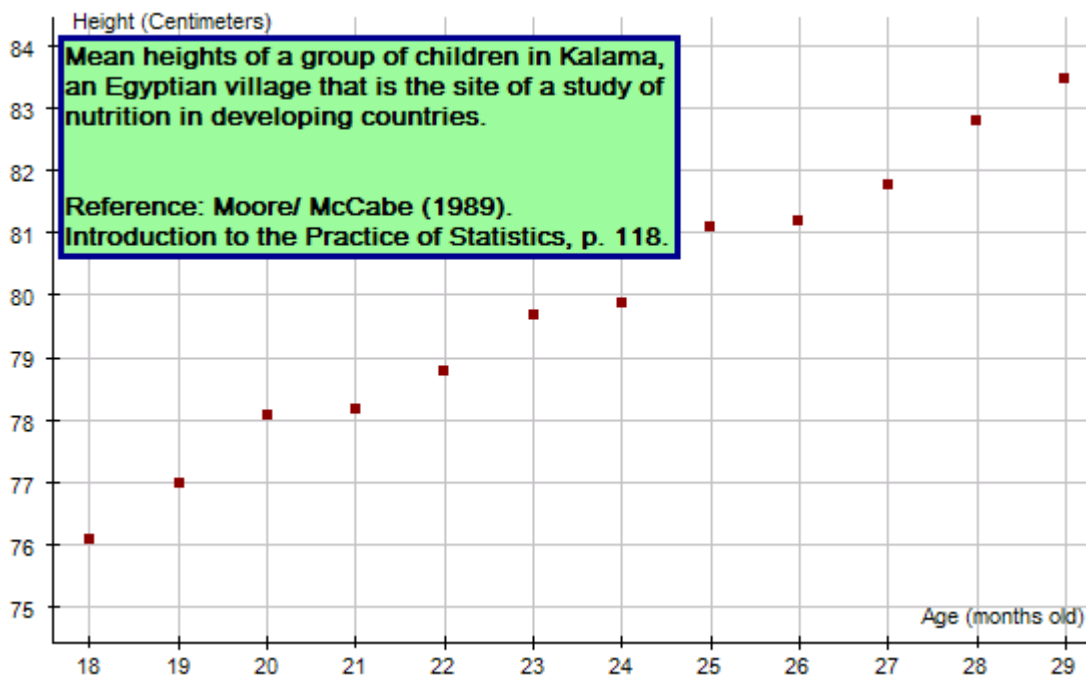


11.



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

- Find an equation of a line in $y - k = a(x - h)$ form that models the scatterplot below
- Write a complete sentence that gives the meaning of the slope, a , in the context of the data set.
- Write a complete sentence that gives the meaning of the point (h, k) , in the context of the data set.



Problem set 2-2

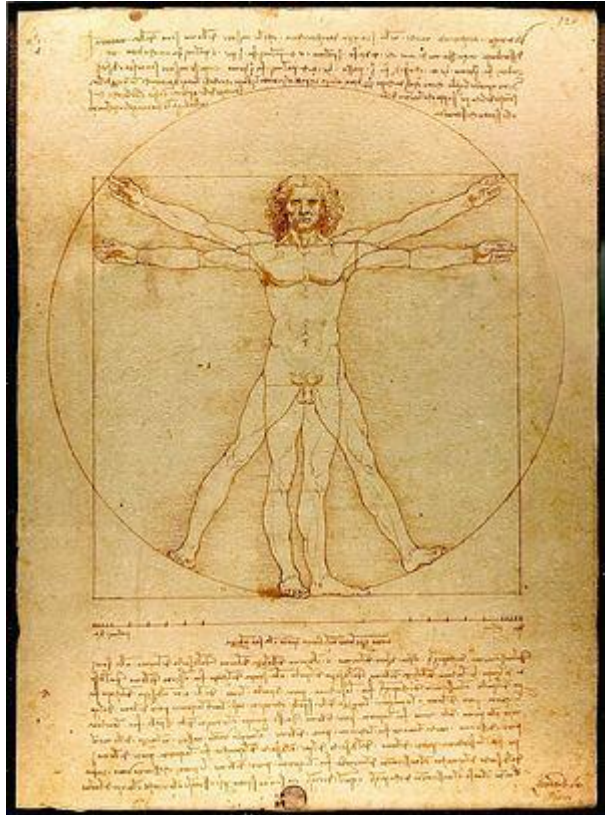
1. 🖨 Data Set: Manatee Deaths

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

- “Analyze the data”.
- Write a sentence that explains the meaning of the slope of your linear model within the context.
- Write a sentence that explains the meaning of the point (h,k) in your linear model within the context.

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).

- “Analyze the data”.
- Write a sentence that explains the meaning of the slope of your linear model within the context.
- Write a sentence that explains the meaning of the point (h,k) in your linear model within the context.
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".
- b) Write a sentence that explains the meaning of the slope of your linear model within the context.
- c) Use your model to predict what the Fuel Efficiency (MPG) would be for a car weighing 2,700 pounds.
- 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?

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


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

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.

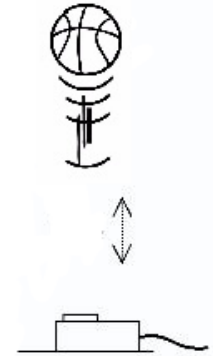
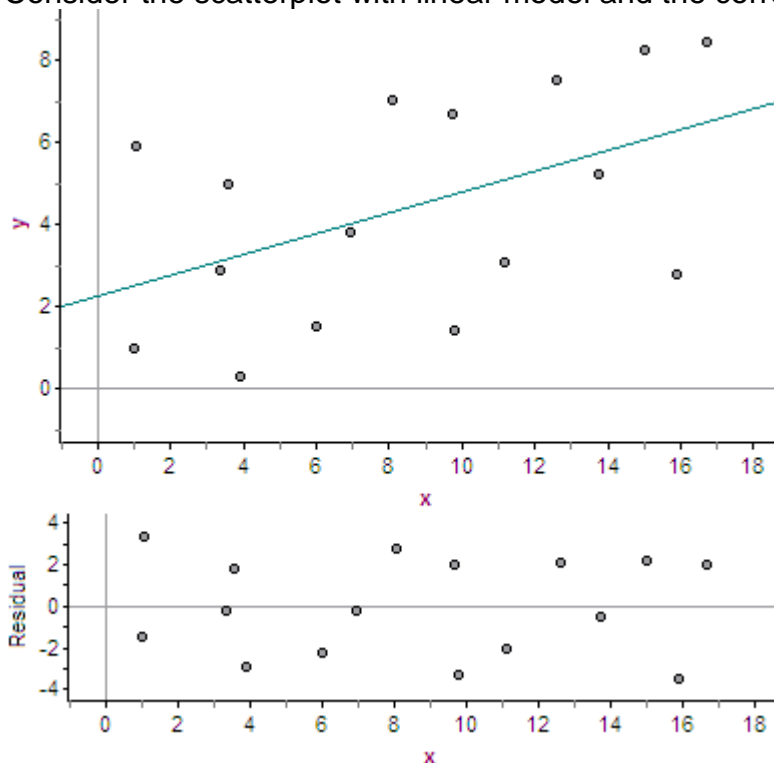


Image source: Investigating a Mass on a Spring. Nuffield Foundation, n.d. Web. 22 Feb. 2013. <http://www.nuffieldfoundation.org/sites/default/files/images/Investigating%20a%20mass-on-spring%20oscillator_322.jpg>.

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



- a) Comment on the residuals.
- b) Comment on the appropriateness of the model.

Problem set 2-4

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

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

- "Analyze the data". Use dosage as the independent variable and weight gain as a percentage as the dependent variable.
- 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_measurer_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.



- "Analyze the data".
- 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.
- 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

- “Analyze the data”.
- 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/>.

- “Analyze the data”.
- 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)

- “Analyze the data”.
- 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”.

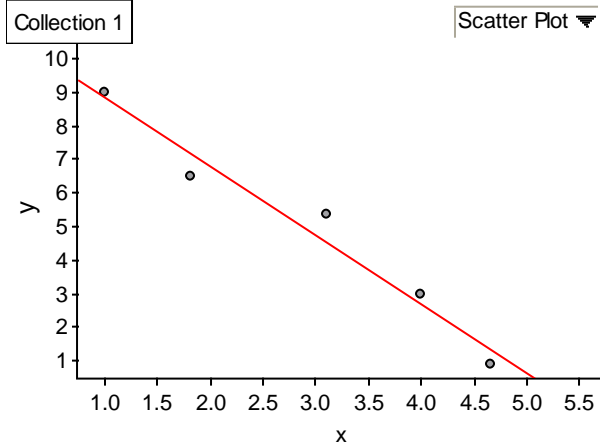
2. 📄 Natural frequency of a tube (data collected in class)
A tube has a natural frequency that is primarily dependent on the length of the tube. The data will be collected in class.
 - a) “Analyze the data”.
 - b) Use your model to predict the natural frequency for a tube of length of 10 cm, 5 cm, and 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.



In 2 – 4, complete the sentence.

- If r (the correlation coefficient) is positive then, ...
- If r^2 (the coefficient of determination) is 0.99 then, ...
- If your statistics software gives you $r^2 = 4$, ...
- 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, $r = 0.9872$.






- What is the meaning of r in this case?
- Did people Googling “facebook,” **cause** unemployment? Explain.

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
 - 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.