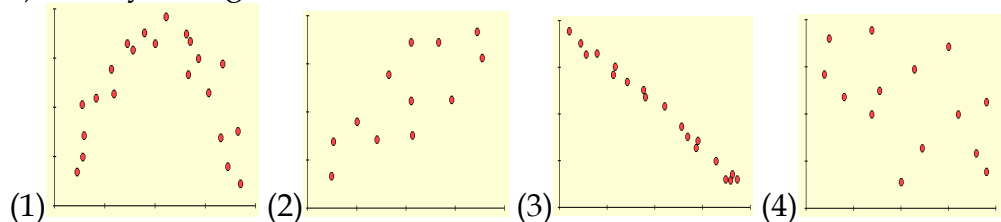


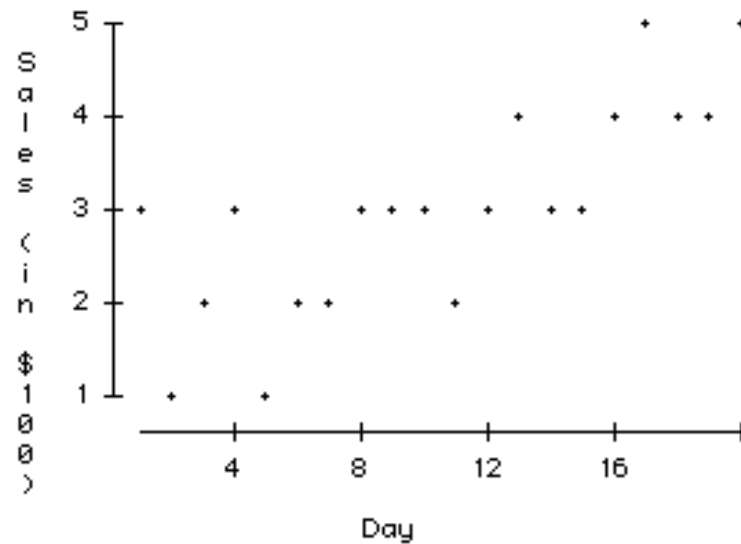
## Stat 101 Hw#3 Due Thursday Sept. 25

- Association** Suppose you were to collect data for each pair of variables. You want to make a scatterplot. Which variable would you use as the explanatory variable, and which as the response variable? Why? What would you expect to see in the scatterplot? Discuss the likely direction, form, and strength.
  - Apples: weight in grams, weight in ounces
  - Apples: circumference (inches), weight (ounces)
  - College freshmen: shoe size, grade point average
  - Gasoline: number of miles driven, gallons remaining in the tank
- Beanstalks** Beanstalk Clubs are social clubs for very tall people. To join a man must be over 6'2" tall, and a woman over 5'10". The National Health Survey suggests that heights of adults may be normally distributed with means heights of 69.3" for men and 64.1" for women. The respective standard deviations are 2.92" and 2.75".
  - You are probably not surprised to learn that men are generally taller than women, but what does the greater standard deviation for men's heights indicate?
  - Who is more likely to qualify for Beanstalk membership, men or women? Hint: Use z-scores !!
- Scatterplots** Which of the scatterplots below show:
  - little or no association?
  - a negative association?
  - a linear association?
  - a moderately strong association?
  - a very strong association?



- Coffee Sales** Owners of a new coffee shop tracked sales for the first 20 days, and displayed the data in a scatterplot (by day):

## Exploring Relationships Between Variables



- Make a histogram of the daily sales since the shop has been in business. (Enter the data by hand)
- State one fact that is obvious from the scatterplot, but not from the histogram.
- State one fact that is obvious from the histogram, but not from the scatterplot.

- Food consumption.** FAOSTAT, the Food and Agriculture Organization of the United Nations, collects information on the production and consumption of more than 200 food and agricultural products for 200 countries around the world. The data set `Food_consumption.jmp` contains, for each country, their per capita consumption of *Alcohol* and *Meat*. The United States leads in meat consumption with 267.30 lbs per person per year, while Ireland is the largest alcohol consumer at 55.80 gallons per person per year.

  - Which country's consumption is more remarkable, that of the U.S. or Ireland? Explain.
  - Using z-scores, find which country is the larger consumer of both meat and alcohol together
- Baldness and Heart Disease** Medical researchers followed 1435 middle aged men for a period of 5 years, measuring the amount of baldness present (none=1, little=2, some=3, much=4, extreme = 5) and presence of heart disease (No=0, Yes= 1). They found a correlation of 0.089 between the two variables. Comment on their conclusion that this shows that baldness is not a possible cause of heart disease.
- Pizza Sales.** The data set `Pizza_Sales.jmp` contains the total *Sales* volume (lbs of pizza) and average *Price* of frozen pizza for 39 different weeks in a U.S. city.

  - Describe the relationship between *Sales* and *Price*.

- b) Find the correlation. Do the assumptions and conditions for correlation appear to be met?
  - c) Does the scatterplot provide any support for the theory that as price is lowered, demand and hence sales increase? Does it prove it?
  - d) How would the correlation change if we measured the *Sales* volume in tons (2000 lbs to a ton) instead of pounds?
8. **Carbon Footprint** Carbon footprint. The data set *Carbon\_Footprint.jmp* contains, for 2008 cars, the *Carbon Footprint* (tons of CO<sub>2</sub> per year) vs. the new Environmental Protection Agency (EPA) *Highway Mileage* for 82 family sedans as reported by the U.S. government ([www.fueleconomy.gov/feg/byclass.htm](http://www.fueleconomy.gov/feg/byclass.htm)).
- a) Make a scatterplot and find the correlation between *Carbon Footprint* and *Highway Mileage*.
  - b) Do the assumptions and conditions for correlation appear to be met?
  - c) What's the car with the highest mileage (and lowest carbon footprint?). Label it on the graph. Hint: In the left part of the screen where the three variables *Car*, *Highway Mileage* and *Carbon Footprint* are listed, right click on *Car*, and select **Label/Unlabel**. Now just click on the point. To get it to stay, right click on the point and select **Row Label** – now you can print it.
  - d) How does the correlation change if you delete that car? Explain.
9. **Carbon Footprint, part 2.** Using the data in exercise 7,
- a) Find a linear regression to predict *Carbon Footprint* from *Highway Mileage*. Find the predicted value for the car mentioned in 7c. Hint: Fit the regression and then under **Linear Fit** select **Save Predicteds**. The predicted value will appear in the data set.
  - b) Delete that car (the car you deleted in 7c) and find the regression without it. How does it change the regression?
  - c) Now predict the *Carbon Footprint* of that car from the regression that was fit without it and compare to the value you got in a). Is it closer, or farther from the point? Explain.
10. **Life Expectancy** Data from the World Bank for 26 Western Hemisphere countries (*Life\_Expectancy\_2004.jmp* and shown below) can be used to examine the association between female life expectancy and the average number of children women give birth to.

## Exploring Relationships Between Variables

Country	Births/ woman	Life Exp.	Country	Births/ woman	Life Exp.
Argentina	2.5	77	Bahamas	2.2	77
Barbados	1.8	78	Belize	3.5	74
Bolivia	4.0	64	Brazil	2.2	71
Canada	1.5	82	Chile	2.2	79
Colombia	2.7	74	Costa Rica	2.5	79
Dom. Rep.	2.8	73	Ecuador	3.1	71
El Salv.	3.2	72	Guatemala	4.7	68
Jamaica	2.5	77	Honduras	4.0	72
Mexico	2.8	75	Nicaragua	3.6	71
Panama	2.5	76	Paraguay	4.0	72
Peru	3.1	71	Puerto Rico	3.1	71
U. S.	2.1	80	Uruguay	2.3	78
Venezuela	2.9	76	Virgin Is.	2.4	79

- Create a scatterplot and describe the association.
- Find the equation of the line of regression.
- Interpret the slope and the  $y$  - intercept of the line.
- Is the line an appropriate model? Describe what you see in the residuals plot. Hint: **Linear Fit → Plot Residuals**
- If government leaders wanted to increase life expectancy in their country should they encourage women to have fewer children? Explain.