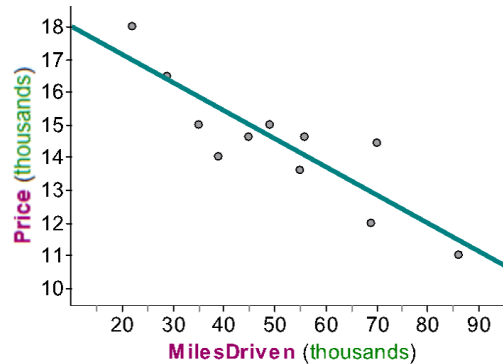


## Guided Notes 3.2 Least-Squares Regression

Read 164–167

What is the general form of a regression equation? What is the difference between  $y$  and  $\hat{y}$ ?

Alternate Example: Used Hondas  
The following scatterplot shows the number of miles driven (in thousands) and advertised price (in thousands) for 11 used Honda CR-Vs from the 2002-2006 model years. The regression line shown on the scatterplot is  $\hat{y} = 18773 - 0.08618x$ .



Miles Driven	Price
22000	17998
29000	16450
35000	14998
39000	13998
45000	14599
49000	14988
55000	13599
56000	14599
69000	11998
70000	14450
86000	10998

a) Interpret the slope and  $y$  intercept of a regression line.

b) Predict the price of a used CR-V with 50,000 miles.

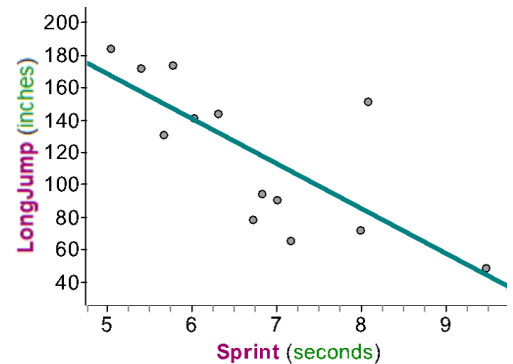
c) Predict the price of a used CR-V with 250,000 miles. How confident are you in this prediction?

What is extrapolation? Is it a good idea to extrapolate?

Alternate Example: Using the Track and Field data from earlier, the equation of the least-squares regression line is  $\hat{y} = 305 - 27.6x$  where  $y$  = long jump distance and  $x$  = sprint time.

a) Interpret the slope.

b) Does it make sense to interpret the  $y$ -intercept? Explain.



### 3.2 Residuals

Read 168–171

What is a residual? How do you interpret a residual?

Calculate and interpret the residual for the Honda CR-V with 86,000 miles and an asking price of \$10,998.

How can we determine the “best” regression line for a set of data?

Is the least-squares regression line resistant to outliers?

Example: McDonalds Beef Sandwiches.

Carbs (g)	31	33	34	37	40	40	45	37	38
Fat (g)	9	12	23	19	26	42	29	24	28

(a) Calculate the equation of the least-squares regression line using technology. Make sure to define variables! Sketch the scatterplot with the graph of the least-squares regression line.

(b) Interpret the slope and  $y$ -intercept in context.

(c) Calculate and interpret the residual for the Big Mac, with 45g of carbs and 29g of fat.

What is a residual plot? What is the purpose of a residual plot?

What two things do you look for in a residual plot? How can you tell if a linear model is appropriate?

- *Leftover patterns:*
- *Size of residuals*
- *Can't use correlation to assess linearity.*

Construct and interpret a residual plot for the Honda CR-V data.

**HW page 191 (40, 42, 46, 53, 55, 61)**

### **3.2 Standard deviation of the residuals and $r^2$**

Read page 177

What is the standard deviation of the residuals? How do you calculate and interpret it?

Calculate and interpret the standard deviation of the residuals for the Honda CR-V data.

Suppose that you see a used Honda CR-V for sale. Predict the asking price for this CR-V.

How much better would our predictions be if we knew how many miles it had been driven?

Read 179–181

What is the coefficient of determination  $r^2$ ? How do you calculate and interpret  $r^2$ ?

How is  $r^2$  related to  $r$ ? How is  $r^2$  related to  $s$ ?

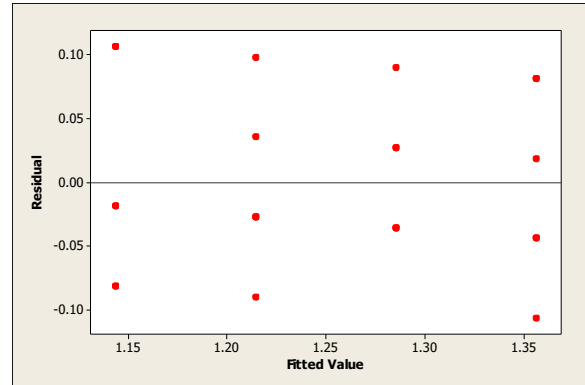
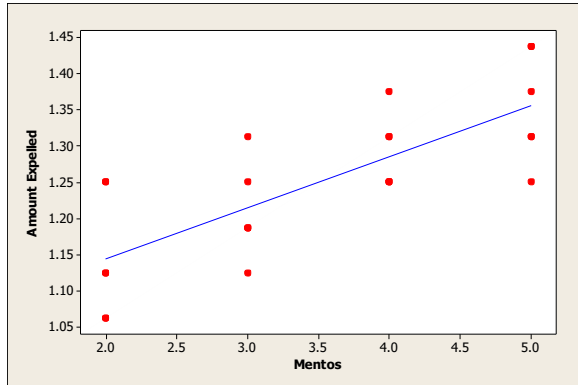
**HW: page 193 (54, 56, 58, 62)**

### **Interpreting Computer Output**

Read 181–183

When Mentos are dropped into a newly opened bottle of Diet Coke, carbon dioxide is released from the Diet Coke very rapidly, causing the Diet Coke to be expelled from the bottle. Will more Diet Coke be expelled when there is a larger number of Mentos dropped in the bottle?

Two statistics students, Brittany and Allie, decided to find out. Using 16 ounce (2 cup) bottles of Diet Coke, they dropped either 2, 3, 4, or 5 Mentos into a randomly selected bottle, waited for the fizzing to die down, and measured the number of cups remaining in the bottle. Then, they subtracted this measurement from the original amount in the bottle to calculate the amount of Diet Coke expelled (in cups). Output from a regression analysis is shown below.



(a) What is the equation of the least-squares regression line? Define any variables you use.

Predictor	Coef	SE Coef	T	P
Constant	1.00208	0.04511	22.21	0.000
Mentos	0.07083	0.01228	5.77	0.000

S = 0.0672442    R-Sq = 60.2%    R-Sq(adj)

(b) Interpret the slope of the least-squares regression line.

(c) What is the correlation?

(d) Is a linear model appropriate for this data? Explain.

(e) Would you be willing to use the linear model to predict the amount of Diet Coke expelled when 10 mentos are used? Explain.

(f) Calculate and interpret the residual for bottle of diet coke that had 2 mentos and lost 1.25 cups

(g) Interpret the values of  $r^2$  and  $s$ .

(h) If the amount expelled was measured in ounces instead of cups, how would the values of  $r^2$  and  $s$  be affected? Explain.

**HW page 194 (63, 64,)**

Using the top ten money winners from the 2009 LPGA Tour, we can investigate the relationship between average driving distance and driving accuracy using a scatterplot. Here we will use average driving distance (in yards) as the explanatory variable and driving accuracy (proportion of drives that land in the fairway) as the response variable.

Player	Average Driving Distance	Driving Accuracy
Jiyai Shin	246.8	0.824
Lorena Ochoa	265.2	0.718
Ai Miyazato	254.3	0.757
Cristie Kerr	263.7	0.714
Na Yeon Choi	255.5	0.733
Suzann Pettersen	268.1	0.660
Yani Tseng	269.2	0.654
In-Kyung Kim	249.3	0.748
Paula Creamer	248.6	0.811
Anna Nordqvist	245.7	0.770

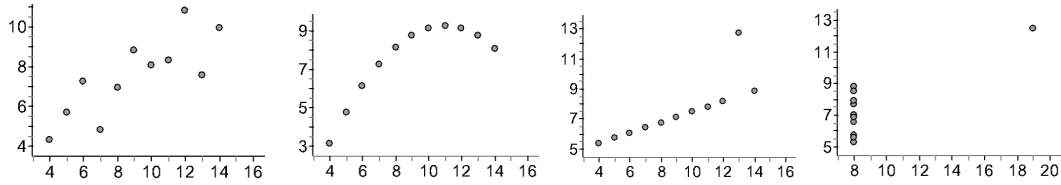
1. Draw a scatterplot for this association and discuss the noticeable features.
2. Calculate the equation of the least squares regression line and graph it on the scatterplot.
3. Interpret the slope and  $y$ -intercept in the context of the problem.
4. Calculate and interpret the value of the correlation coefficient.
5. If the distance was measured in feet instead of yards, how would the correlation change? Explain.
6. Calculate and interpret the residual for Lorena Ochoa.
7. Sketch the residual plot. What information does this provide?
8. Calculate and interpret the value of  $s$  in the context of the problem.
9. Calculate and interpret the value of  $r^2$  in the context of the problem.

**Regression Wisdom, etc.**

Read 183–188

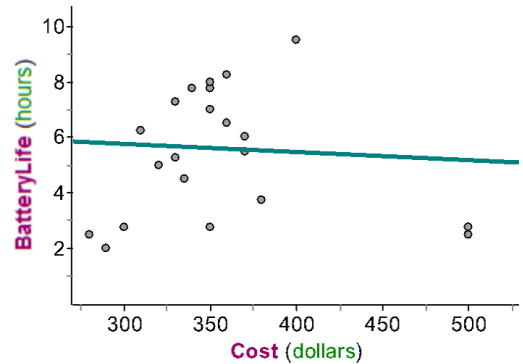
Does it matter which variable is  $x$  and which is  $y$ ?

Which of the following has the highest correlation?



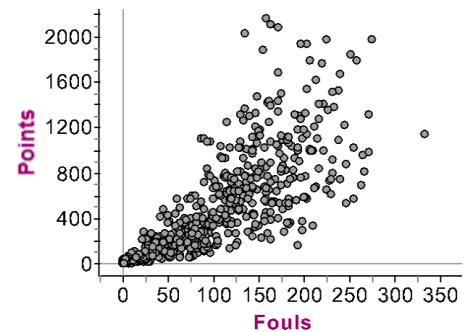
How do outliers affect the correlation, least-squares regression line, and standard deviation of the residuals? Are all outliers influential?

Here is a scatterplot showing the cost in dollars and the battery life in hours for a sample of netbooks (small laptop computers). What effect do the two netbooks that cost \$500 have on the equation of the least-squares regression line, correlation, standard deviation of the residuals, and  $r^2$ ? Explain.



Here is a scatterplot showing the relationship between the number of fouls and the number of points scored for NBA players in the 2010-2011 season.

- Describe the association.
- Should NBA players commit more fouls if they want to score more points?



Read 172-174

How can you calculate the equation of the least-squares regression line using summary statistics?

*These are on the formula sheet!*

*Note that the least-squares regression line always goes through  $(\bar{x}, \bar{y})$*

What happens to the predicted value of  $y$  for each increase of 1 standard deviation in  $x$ ?

**HW page 192 (47, 49, 65, 71–78)**