<u>Day #1</u>: 3.1 Describing Relationships

Read 141-142

Why do we study relationships between two variables?

Read 143-144

Page 144: Check Your Understanding

Read 144-149

How do you know which variable to put on which axis?

What is the easiest way to lose points when making a scatterplot?

Alternate Example: Track and Field Day! The table below shows data for 13 students in a statistics class. Each member of the class ran a 40-yard sprint and then did a long jump (with a running start). Make a scatterplot of the relationship between sprint time (in seconds) and long jump distance (in inches).

Sprint Time (s)	5.41	5.05	9.49	8.09	7.01	7.17	6.83	6.73	8.01	5.68	5.78	6.31	6.04
Long Jump Distance (in)	171	184	48	151	90	65	94	78	71	130	173	143	141

What four characteristics should you consider when interpreting a scatterplot?

Does a strong association between two variables indicate a cause-and-effect relationship?

Page 149: Check Your Understanding

Read 149-150: Using technology to create scatterplots

HW #1: page 158 (1, 5, 7, 9, 11)

<u>Day #2</u>: 3.1 Correlation / 3.2 Regression Lines

Just like two distributions can have the same shape and center with different spreads, two associations can have the same directions and form, but very different strengths.

Read 150-151

How do we measure the strength of a linear relationship between two quantitative variables?

What are some characteristics of correlation?

Activity page 152

Can you determine the form of a relationship using only the correlation?

Is correlation a resistant measure of strength?

Read 152-155

Do you need to know the formula for correlation?

Read 155-156

What are some additional facts about correlation?

Read 164-166

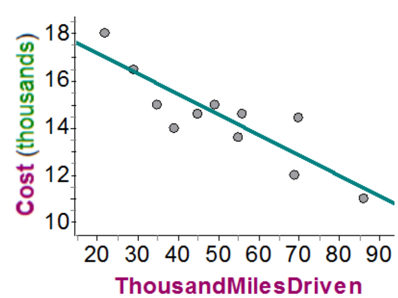
What is a regression line?

What is regression to the mean?

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

How do you interpret the slope and y intercept of a regression line?

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 – 86.18x. Interpret the slope and y intercept in context.



Read 166-167

What is extrapolation? Why is it dangerous?

Page 167 Check Your Understanding

HW #2 page 160 (15-18, 21, 27-32) page 191 (35, 37)

<u>Day #3</u>: 3.2 Residuals

Read 168-171

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

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

Is the least-squares regression line resistant to outliers?

Alternate Example: Track and Field Day! The table below shows data for 13 students in a statistics class. Each member of the class ran a 40-yard sprint and then did a long jump (with a running start).

Sprint Time (s)	5.41	5.05	9.49	8.09	7.01	7.17	6.83	6.73	8.01	5.68	5.78	6.31	6.04
Long Jump Distance (in)	171	184	48	151	90	65	94	78	71	130	173	143	141

(a) Calculate the equation of the least-squares regression line using technology.

(b) Interpret the slope in context. Does it make sense to interpret the *y* intercept?

(c) Calculate and interpret the residual for the student who had a sprint time of 8.01 seconds.

Read 172-174

How can you calculate the least-squares regression line from summary statistics?

What is the relationship between the correlation and the slope?

Read 174-177

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?

Construct and interpret a residual plot for the Track-and-Field Day data.

Suppose that a student was absent on the day these data were collected. Predict this student's long jump distance.

Read 177-181

What is *s*? How do you calculate and interpret *s*?

What is r^2? How do you calculate and interpret r^2?

How are *s* and r^2 the same? Different?

HW #3 page 191 (43, 47, 49, 53, 55, 57, 61)

<u>Day #4</u>: Chapter 3 Review (so far)

The anticipation of the first blooms of spring flowers is one of the joys of April. One of the most beautiful is that of the Japanese cherry tree. Experience has taught us that, if the spring has been a warm one, the trees will bloom early, but if the spring has been cool, then the blossoms will appear later. Mr. Yamada is a gardener who has observed the date in April when the blossoms first appear for the last 24 years. His son, Hiro, went on the internet and found the average March temperature (in degrees Celsius) in his area for those years. For example, for the first year in the data table, the average March temperature was 4°C and the first bloom happened on the 14th day of April.

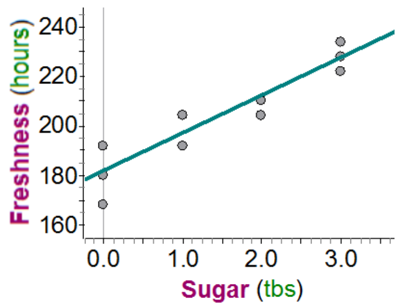
Temp	Date
4	14
5.4	8
3.2	11
2.6	19
4.2	14
4.7	14
4.9	14
4	21
4.9	9
3.8	14
4	13
5.1	11
4.3	13
1.5	28
3.7	17
3.8	19
4.5	10
4.1	17
6.1	3
6.2	3
5.1	11
5	6
4.6	9
4	11

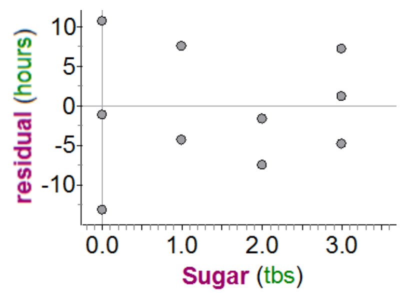
1. Why should temperature be the explanatory variable?
2. Draw a scatterplot and discuss the noticeable features.
3. Calculate the least squares regression line and graph it on the scatterplot.
4. Interpret the slope in the context of the problem.
5. Interpret the *x*- and *y*-intercepts in the context of the problem. Are these reasonable values in this context?
6. Calculate and interpret the value of the correlation coefficient.
7. If the temperature was measured in degrees Fahrenheit, how would this value change?
8. If *r* is high, can we conclude that warmer temperatures cause earlier blooms?
9. Calculate and interpret the residual for the first year in the data table.
10. Sketch the residual plot. What does it tell you?
11. Calculate and interpret the values of r^2 and *s* in the context of the problem.
12. If you were to use number of hours in April instead of number of days in April for the date of first bloom, how would the values of r^2 and *s* change?
13. Predict the date of first bloom for a year with an average March temperature of 3°C. How confident are you in your prediction?

HW #4 page 193 (54, 56, 58, 59, 60)

<u>Day #5</u>: 3.2 Computer Output/ Regression Wisdom
Read 181-183

Alternate Example: Fresh Flowers
For their second semester project, two AP Statistics students decided to investigate the effect of sugar on the life of cut flowers. They went to the local grocery store and randomly selected 12 carnations. All the carnations seemed equally healthy when they were selected. When they got home, the students prepared 12 identical vases with exactly the same amount of water in each vase. They put 1 tablespoon of sugar in three vases, 2 tablespoons of sugar in three vases, and 3 tablespoons of sugar in three vases. In the remaining 3 vases, they put no sugar. After the vases were prepared and placed in the same location, the students randomly assigned one flower to each vase and observed how many hours each flower continued to look fresh. A scatterplot, residual plot, and computer output from the regression are shown.






```
Predictor      Coef  SE Coef      T      P
Constant    181.200    3.635  49.84  0.000
Sugar        15.200    1.943   7.82  0.000

S = 7.52596   R-Sq = 86.0%   R-Sq(adj) = 84.5%
```

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

(b) What is the correlation?

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

(d) Calculate and interpret the residual for the flower that got 3 tablespoons of sugar and lasted 222 hours.

Read 183-188

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

What should you always do before calculating the correlation or least-squares regression line?

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

Does association imply causation?

HW #5 page 194 (63, 68, 69, 71-78)