<u>Monday, October 3</u>: 3.1 Describing Relationships

Read 141-142

Why do we study relationships between two variables?

Read 143-144

Alternate Example: Identify the explanatory and response variables

a) amount of rain; weed growth

b) winning percentage of a basketball team; attendance at games

c) resting pulse rate; amount of daily exercise

Read 144-149

How do you know which variable to put on which axis? Where do you start each axis?

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

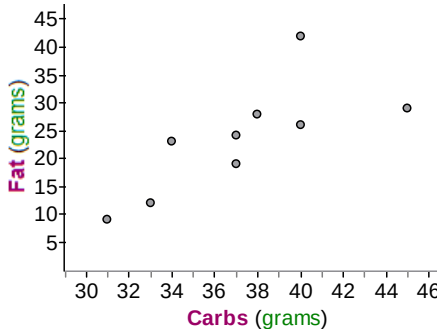
Alternate Example: Track and Field Day! The table below shows data for 13 students in a statistics class. Each member of the class ran a 40-yard sprint and then did a long jump (with a running start). Make a scatterplot of the relationship between sprint time (in seconds) and long jump distance (in inches).

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?

Alternate Example: The following scatterplot shows the amount of carbs (in grams) and amount of fat (in grams) of beef sandwiches from McDonalds.
(a) Describe the relationship between carbs and fat.
(b) Is there a cause-and-effect relationship between carbs and fat? Explain.



Read 149-150: Using technology to create scatterplots

HW #25: page 158 (1, 5, 7, 9, 11)

<u>Tuesday, October 4</u>: 3.1 Correlation

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

Read 150-151

What is the correlation *r*?

What are some characteristics of the correlation?

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?

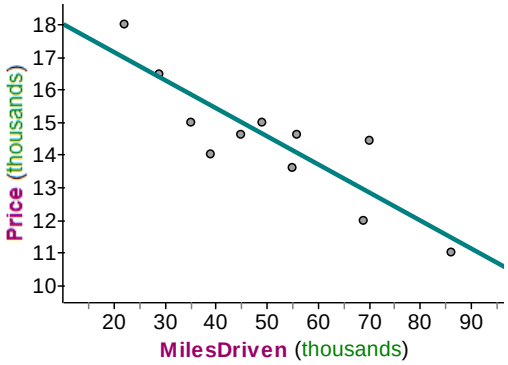
HW #26: page 160 (15-18, 21, 27-32)

Wed/Thurs, October 5/6: 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.

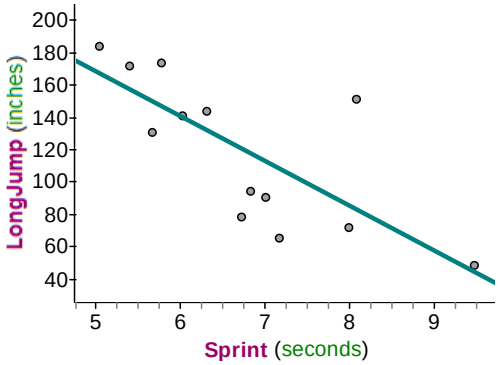
b) Predict the price of a used CR-V with 50,000 miles.

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.



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?

Alternate 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.

HW #26: page 191 (35, 37, 39, 41, 45)

Friday, October 7: 3.2 More about Residuals

Read 174-178

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 Honda CR-V data.

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.

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?

Using the McDonalds beef sandwich data:

a) Construct and interpret a residual plot

b) Calculate and interpret the standard deviation of the residuals

c) Calculate and interpret r^2.

HW #26 page 191 (53, 55, 57, 61)

<u>Monday, October 17</u>: Chapter 3 Review (so far)

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

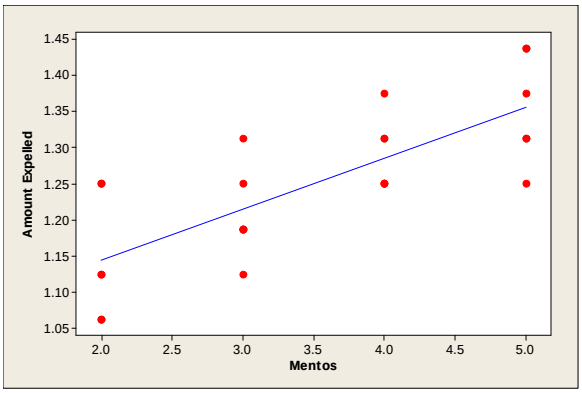
1. Explain why average driving distance should be the explanatory variable.
2. Draw a scatterplot for this association and discuss the noticeable features.
3. Calculate the equation of the least squares regression line and graph it on the scatterplot.
4. Interpret the slope and *y*-intercept in the context of the problem.
5. Calculate and interpret the value of the correlation coefficient.
6. If the distance was measured in feet instead of yards, how would the correlation change? Explain.
7. Calculate and interpret the residual for Lorena Ochoa.
8. Sketch the residual plot. What information does this provide?
9. Calculate and interpret the value of *s* in the context of the problem.
10. Calculate and interpret the value of r^2 in the context of the problem.
11. If you were to use the percentage of shots that land in the fairway instead of the proportion of shots that land in the fairway, how would the values of r^2 and *s* change?
12. Predict the driving accuracy for an LPGA golfer with an average driving distance of 300 yards. How confident are you in your prediction? Explain.

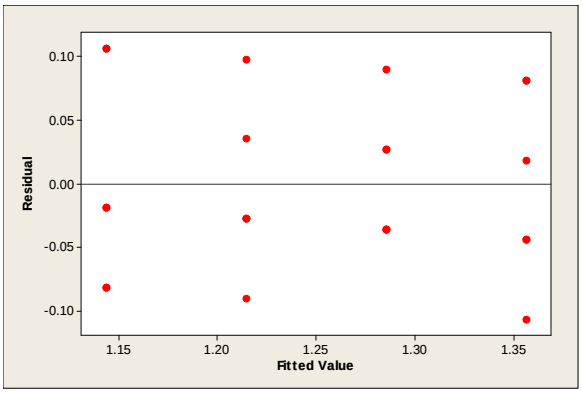
HW #27 page 193 (54, 56, 58, 59, 60)

<u>Tuesday, October 18</u>: 3.2 Computer Output/ Regression Wisdom

Read 181-183

Alternate Example: Mentos and Diet Coke
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) Calculate and interpret the residual for bottle of diet coke that had 2 mentos and lost 1.25 cups.

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

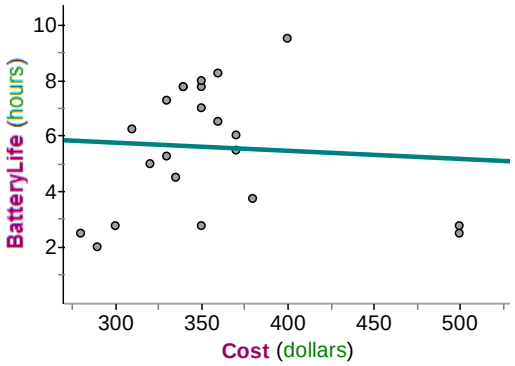
Read 183-188

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?

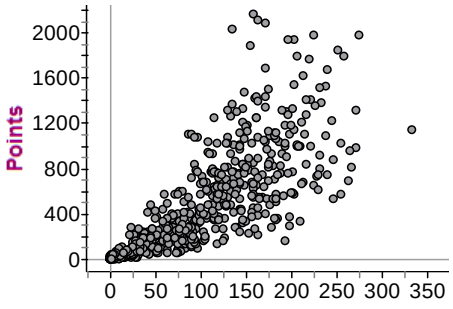
Here is a scatterplot showing the cost in dollars and the battery life in hours for a sample of netbooks (small laptop computers). What effect do the two netbooks that cost $500 have on the equation of the least-squares regression line, correlation, standard deviation of the residuals, and r^2? Explain.



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.

a) Describe the association.

b) Should NBA players commit more fouls if they want to score more points? Explain.



If time, show Hans Rosling video

HW #28 page 194 (63, 65, 71-78)

<u>Wed/Thurs, October 19/20</u>: 3.2 Regression to the Mean

Read 172-174

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

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

FRAPPY:

If time, show Hans Rosling video

HW #29 page 198 Chapter 3 Review Exercises

<u>Friday, October 21</u>: Review Chapter 3 / Projects

Discuss first semester projects

HW #30 page 200 Chapter 3 AP Statistics Practice Test

<u>Monday, October 24</u>: Review Chapter 3

<u>Tuesday, October 25</u>: Chapter 3 Test

HW #31 page 276 Cumulative AP Practice Test I

<u>Wed/Thurs, October 26/27</u>: Review for Midterm

<u>Friday, October 29</u>: Review for Midterm

Cootie problem!

<u>Monday, October 31</u>: Midterm

AP Statistics First Semester Project: Response Bias

The Project: You and your partner (or you by yourself) will design and conduct an experiment to investigate the effects of response bias in surveys. You may choose the topic for your surveys, but you must design your experiment so that it can answer at least one of the following questions:

- Can the wording of a question create response bias?
- Does providing additional information create response bias?
- Do the characteristics of the interviewer create response bias?
- Does anonymity change the responses to sensitive questions?
- Does manipulating the answer choices change the response?
- Can revealing other peoples' answers to a question create response bias?

Proposal (25 points):

- The proposal is due: _____. Late work will be penalized 20% per day, even if you are absent.
- The proposal will be worth 25% of the grade, so don't treat it casually.
- If the proposal isn't approved the first time, you will need to resubmit it for a reduced grade. You must attach the original proposal to any resubmissions.

In your proposal, you should:

- Describe your topic and state which type of response bias you are investigating
- Describe how you will obtain your subjects (minimum sample size is 50). This must be practical!! Your population does not need to be from CDO nor should you interrupt any classes.
- Describe what your questions will be and how they will be asked, including how you will incorporate the principles of a good experiment and avoid potentially confounding variables. You should also indicate what your hypotheses are. Convince me that you have a good design!

Poster (75 points):

- The poster is due: _____. Late work will be penalized 20% per day, even if you are absent.
- The key to a good statistical poster is communication and organization. Make sure all components of the poster are focused on answering the question of interest.
- The poster should be standard sized and not on foam board. Make sure the poster is light enough to be hung on the wall.

The poster should include:

- Title (in the form of a question).
- Introduction. In the introduction you should discuss what question you are trying to answer, why you chose this topic, and what your hypotheses are.
- Data Collection. In this section you will describe how you obtained your data. Be specific.
- Graphs and Summary Statistics. Make sure the graphs are well labeled, easy to compare, and help answer the question of interest.
- Discussion and Conclusions. In this section, you will state your conclusions. You should also discuss any errors you made, what you could do to improve the study next time, and any other comments based on your own critical reflection on the project.
- Live action pictures of your data collection in progress.

Presentation: Each pair (or individual) will be required to give a 5 minute oral presentation to the class. Both members need to participate equally and should be prepared to answer questions.

Response Bias Project	4 = Complete	3 = Substantial	2 = Developing	1 = Minimal
Intro	• Describes the context of the research • Has a clearly stated question of interest • Provides a hypothesis about the answer to the question of interest • Question of interest is of appropriate difficulty	• Introduces the context of the research and has a specific question of interest • Suggests hypothesis OR has appropriate difficulty	• Introduces the context of the research and has a specific question of interest OR has question of interest and a hypothesis	• Briefly describes the context of the research
Data Collection	• Method of data collection is clearly described • Includes appropriate randomization • Describes efforts to reduce bias, variability, confounding • Quantity of data collected is appropriate	• Method of data collection is clearly described • Some effort is made to incorporate principles of good data collection • Quantity of data is appropriate	• Method of data collection is described • Some effort is made to incorporate principles of good data collection	• Some evidence of data collection
Graphs and Summary Statistics	• Appropriate graphs are included (to help answer the question of interest) • Graphs are neat, clearly labeled, and easy to compare • Appropriate summary statistics are included • Summary statistics are discussed and correctly interpreted	• Appropriate graphs are included (to help answer the question of interest) • Graphs are neat, clearly labeled, and easy to compare • Appropriate summary statistics are included	• Graphs and summary statistics are included	• Graphs or summary statistics are included
Conclusions	• Uses the results of the study to correctly answer question of interest • Discusses what inferences are appropriate based on study design • Shows good evidence of critical reflection (discusses possible errors, shortcomings, limitations, alternate explanations, etc.)	• Makes a correct conclusion • Discusses what inferences are appropriate • Shows some evidence of critical reflection	• Makes a partially correct conclusion • Shows some evidence of critical reflection	• Makes a conclusion
Poster, Presentation, & Communi-cation	• Clear, holistic understanding of the project • Poster is well organized, neat and easy to read • Poster included pictures of data collection in progress and is visually appealing • Oral presentation is well organized	• Clear, holistic understanding of the project • Poster is unorganized, lacks pictures, or isn't visually appealing or oral presentation is not organized	• Poster and oral presentation are not well done or communication is poor	• Communicat ion and organization are very poor