Chapter 7, 8 and 9 Notes:

Chapter 7 Scatterplots, Association, and Correlation:

#2 (Taken from Bock, Velleman, and DeVeaux book, p. 164): 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.

a) T-shirts at a store: price each, number sold.

b) Scuba diving: depth, water pressure.

c) Scuba diving: depth, visibility.

d) All elementary school students: weight, score on a reading test.

#5 (Taken from Bock, Velleman, and DeVeaux book, p. 164): Scatterplots. Which of the scatterplots at the top of the next column show

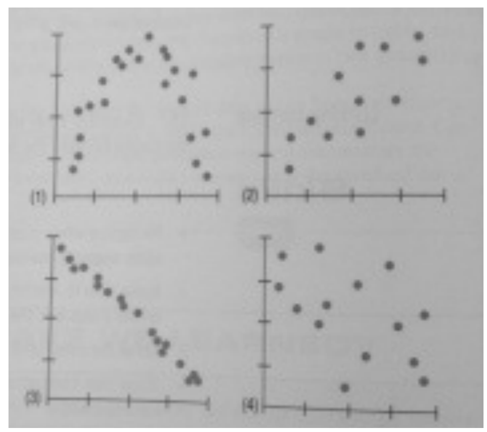
a) little or no association?

b) A negative association?

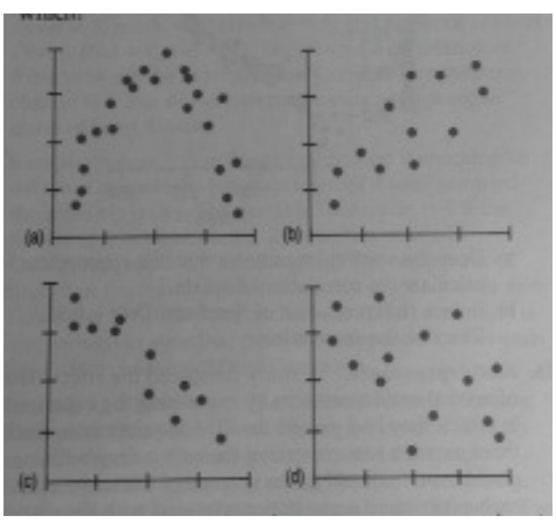
c) A linear association?

d) A moderately strong association?

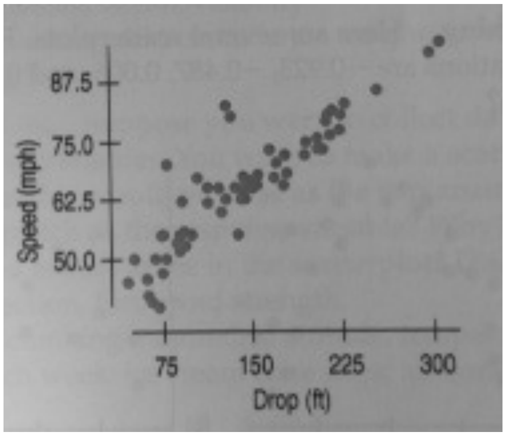
e) A very strong association?



#11 (Taken from Bock, Velleman, and DeVeaux book, p. 165): Matching. Here are several scatterplots. The calculated correlations are -0.923, -0.487, 0.006, and 0.777. Which is which?



#15 (Taken from Bock, Velleman, and DeVeaux, p. 166): Roller Coasters. Roller coasters get all their speed by dropping down a steep initial incline, so it makes sense that the height of the drop might be related to the speed of the coaster. Here's a scatterplot of top *Speed* and the largest *Drop* for 75 coasters around the world.



a) Does the scatterplot indicate that it is appropriate to calculate the correlation? Explain.

b) In fact, the correlation of *Speed* and *Drop* is 0.91. Describe the association.

#25 (Taken from Bock, Velleman, and DeVeaux, p. 167): Height and reading. A researcher studies children in elementary school and finds a strong positive linear association between height and reading scores.

a) Does this mean that taller children are generally better readers?

b) What might explain the strong correlation?

<u>Chapter 8:</u>

#35 (Taken from SYM book, p. 191): What's my line? You use the same bar of soap to shower each morning. The bar weighs 80 grams when it is new. Its weight goes down by 6 grams per day on the average. What is the equation of the regression line for predicting weight from days of use?

#36 (Taken from SYM book, p. 191): What's my line? An eccentric professor believe that a child with IQ 100 should have a reading test score of 50, and that reading score should increase by 1 point for every additional point of IQ. What is the equation of the professors, regression line for predicting reading score from IQ?

#37 (Taken from SYM book, p. 191): Gas mileage. We expect a car's highway gas mileage to be related to its city gas mileage. Data for all 1198 vehicles in the government's *2008 Fuel Economy Guide* give the regression line predicted highway mpg = 4.62 + 1.109 (city mpg).

a) What's the slope of this line? Interpret this value in context.

b) What's the intercept? Explain why the value of the intercept is not statistically meaningful.

c) Find the predicted highway mileage for a car that gets 16 miles per gallon in the city. Do the same for a car with city mileage 28 mpg.

#38 (Taken from SYM book, p. 191): IQ and reading scores. Data on the IQ test scores and reading test scores for a group of fifth-grade children give the following regression line: predicted reading score = -33.4 + 0.8882 (IQ score).

a) What's the slope of this line? Interpret this value in context.

b) What's the intercept? Explain why the value of the intercept is not statistically significant.

c) Find the predicted reading scores for two children with IQ scores of 90 and 130, respectively.

#47 (Taken from SYM book, p. 192): Husbands and wives. The mean height of American women in their early twenties is 64.5 inches and the standard deviation is 2.5 inches. The mean height of men the same age is 68.5 inches, with standard deviation 2.7 inches. The correlation between the heights of husbands and wives is about r = 0.5.

a) Find the equation of the least-squares regression line for predicting husband's height from wife's height. Show your work.

b) Use your regression line to predict the height of the husband of a woman who is 67 inches tall. Explain why you could have given this result without doing the calculation.

#54 (Taken from SYM book, p. 193): Do heavier people burn more energy? The table below presents data on the lean body mass and the resting metabolic rate for 12 women who were subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all the fat. Metabolic rate, in calories burned per 24 hours, is the rate at which the body consumes energy.

Mass:	36.1	54.6	48.5	42.0	50.6	42.0	40.3	33.1	42.4	34.5	51.1	41.2
Rate:	995	1425	1396	1418	1502	1256	1189	913	1124	1052	1347	1204

a) Enter the data into your calculator and make a scatterplot.

b) Use your calculator's regression function to find the equation of the least-squares regression line. Add this line to your scatterplot from part a.

c) Explain in words what the slope of the regression line tells us.

d) Another woman has a lean body mass of 45 kilograms. What is her predicted metabolic rate?

#56 (Taken from SYM book, p. 193): Do heavier people burn more energy? Refer to the data from Exercise 54.

a) Use your calculator to make a residual plot. Describe what this graph tells you about how well the line fits the data.

b) Which has the largest residual? Explain what the value of that residual means in context.

Thursday, September 26:

#58 (Taken from SYM book, p. 193): Do heavier people burn more energy? Refer to exercises 54 and 56. For the regression you performed earlier, $r^2 = 0.768$ and $s = 95.08$. Explain what each of these values means in this setting.

#63 (Taken from SYM book, p. 194): Merlins breeding. Below is summary data of the number of breeding pairs of merlins in an isolated area in each of nine years and the percent of males who returned the next year. The data show that the percent returning is lower after successful breeding seasons and that the relationship is roughly linear.

Predictor	***Coef***	***SE Coef***	***T***	***P***
Constant	157.68	27.68	5.70	0.001
Pairs	-2.9935	0.8655	-3.46	0.011
S = 9.46334	R-sq = 63.1%		R-sq (adj) = 75.8%	

a) What is the equation of the least-squares regression line for predicting the percent of males that return from the number of breeding pairs? Use the equation to predict the percent of returning males after a season with 30 breeding pairs.

b) What percent of the year-to-year variation in percent of returning males is explained by the straight-line relationship with number of breeding pairs the previous year?

c) Use the information to find the correlation r between percent of males that return and the number of breeding pairs. How do you know whether the sign of r is + or – a?

d) Interpret the value of s in this setting.

#64 (Taken from SYM book, p. 195): Does social rejection hurt? Below is summary data from a study that shows that social exclusion causes "real pain." That is, activity in an area of the brain that responds to the physical pain goes up as distress for social exclusion goes up. A scatterplot shows a moderately strong, linear relationship.

Predictor	***Coef***	***SE Coef***	***T***	***P***
Constant	-0.12608	0.02465	-5.12	0.000
Distress	0.060782	0.009979	6.09	0.000
S = 0.0250896	R-sq = 77.18%		R-sq (adj) = 75.1%	

a) What is the equation of the least-squares regression line for predicting brain activity from social distress score? Use the equation to predict brain activity for social distress score 2.0.

b) What percent of the variation in brain activity among these subjects is explained by the straight-line relationship with social distress score?

c) Use the information in the figure to find the correlation r between social distress score and brain activity. How do you know whether the sign of r is + or -?

Chapter 9:

#61 (Taken from SYM book, p. 194): Nahya infant weights. A study of nutrition in developing countries collected data from the Egyptian village of Nahya. Here are the mean weights (in Kilograms) for 170 infants in Nahya who were weighed each month during their first year of life:

Age (months):	1	2	3	4	5	6	7	8	9	10	11	12
Weight (kg):	4.3	5.1	5.7	6.3	6.8	7.1	7.2	7.2	7.2	7.2	7.5	7.8

A hasty user of statistics enters the data into software and computes the least-squares line without plotting the data. The result is $\widehat{weight} = 4.88 + 0.267\ (age)$. A residual plot is shown below. Would it be appropriate to use this regression line to predict y from x? Justify your answer.

#62 (Taken from SYM book, p. 194): Driving speed and fuel consumption. A study was conducted on fuel consumption (y) of a car at various speeds (x). Fuel consumption is measured in liters of gasoline per 100 kilometers driven and speed is measured in kilometers per hour. A statistical software package gives the least-squares regression line and the residual plot shown below. The regression line is $\hat{y} - 11.058 - 0.01466x$. Would it be appropriate to use the regression line to predict y from x? Justify your answer.

#65 (Taken from SYM book, p. 195): Outsourcing by airlines. Below is data on 14 airlines and the percent of major maintenance outsourced and the percent of flight delays blamed on the airline.

Airline	**Outsource percent**	**Delay Percent**
AirTran	66	14
Alaska	92	42
American	46	26
America West	76	39
ATA	18	19
Continental	69	20
Delta	48	26
Frontier	65	31
Hawaiian	80	70
JetBlue	68	18
Northwest	76	43
Southwest	68	20
United	63	27
US Airways	77	24

a) Make a scatterplot with outsourcing percent as x and delay percent as y. Hawaiian Airlines is a high outlier in the y direction. Because several other airlines have similar values of x, the influence of this outlier is unclear without actual calculation.

b) Find the correlation r with and without Hawaiian Airlines. How influential is the outlier for correlation?

c) Find the least-squares line for predicting y from x with and without Hawaiian Airlines. Draw both lines on your scatter plot. Use both lines to predict the percent of delays blamed on an airline that has outsourced 76% of its major maintenance. How influential is the outlier for the least-squares regression line?

Chapter 10: Re-expressing Data

#15 (Taken from Bock, Velleman, and DeVeaux, p. 241): Bakes. The table below shows stopping distances in feet for a car tested 3 times at each of 5 speeds. We hope to create a model that predicts *Stopping Distance* from the *Speed* of the car.

Speed (mph)	Stopping Distances (ft)
20	64, 62, 59
30	114, 118, 105
40	153, 171, 165
50	231, 203, 238
60	317, 321, 276

a) Explain why a linear model is not appropriate.

b) Re-express the data to straighten the scatterplot.

c) Create an appropriate model.

d) Estimate the stopping distance for a car traveling 55 mph.

e) Estimate the stopping distance for a car traveling 70 mph.

f) How much confidence do you place in these predictions? Why?

#17 (p. 242): Baseball salaries 2005. Ballplayers have been signing even larger contracts. The highest salaries (in millions of dollars per season) for some notable players are given in the following table.

Player	Year	Salary (million $)
Nolan Ryan	1980	1.0
George Foster	1982	2.0
Kirby Puckett	1990	3.0
Jose Canseco	1990	4.7
Roger Clemens	1991	5.3
Ken Griffey, Jr.	1996	8.5
Albert Belle	1997	11.0
Pedro Martinez	1998	12.5
Mike Piazza	1999	12.5
Mo Vaughn	1999	13.3
Kevin Brown	1999	15.0
Carlos Delgado	2001	17.0
Alex Rodriguez	2001	22.0
Manny Ramirez	2004	22.5
Alex Rodriguez	2005	26.0

a) Examine a scatterplot of the data. Does it look straight?

b) Find the regression of *Salary* vs. *Year* and plot the residuals. Do they look straight?

c) Re-express the data, if necessary, to straighten the relationship.

d) What model would you report for the trend in salaries?

18. Planet distances and year 2006. At a meeting of the International Astronomical Union (IAU) in Prague in 2006, Pluto was determine not to be a planet, but rather the largest member of the Kuiper belt of icy objects. Let's examine some facts. Here is a table of the 9 sun-orbiting objects formerly known as planets:

Planet	Position #	Dist. From sun (Millions of miles)	Length of year (Earth years)
Mercury	1	36	0.24
Venus	2	67	0.61
Earth	3	93	1.00
Mars	4	142	1.88
Jupiter	5	484	11.86
Saturn	6	887	29.46
Uranus	7	1784	84.07
Neptune	8	2796	164.82
Pluto	9	3707	247.68

a) Plot the *length* of the year against the *Distance* from the sun. Describe the shape of your plot.

b) Re-express one or both variables to straighten the plot. Use the re-expressed data to create a model describing the length of a planet's year based on its distance from the sun.

c) Comment on how well your model fits the data.