Homework #11 - Stat 202

2.14 Make some sketches. For each of the following situations, make a scatterplot that illustrates the given relationship between two variables.

(a) A strong negative linear relationship.

(b) No apparent relationship.

(c) A weak positive relationship.

(d) A more complicated relationship. Give the sketch and explain the relationship.

2.20 Will you live longer if you use the Internet? The World Bank collects data on many variables related to world development for countries throughout the world. Two of these are Internet use, in number of users per 100 people, and life expectancy, in years.[12] Figure 2.11 is a scatterplot of life expectancy versus Internet use.
INTERNETANDLIFE ← plot with Statcrunch

(a) Describe the relationship between these two variables.

(b) A friend looks at this plot and concludes that using the Internet will increase the length of your life. Write a short paragraph explaining why the association seen in the scatterplot does not provide a reason to draw this conclusion.

2.27 Explanatory and response variables. In each of the following situations, is it more reasonable to simply explore the relationship between the two variables or to view one of the variables as an explanatory variable and the other as a response variable? In the latter case, which is the explanatory variable and which is the response variable?

(a) The weight of a child and the age of the child from birth to 10 years.

(b) High school English grades and high school math grades.

(c) The rental price of apartments and the number of bedrooms in the apartment.

(d) The amount of sugar added to a cup of coffee and how sweet the coffee tastes.

(e) The student evaluation scores for an instructor and the student evaluation scores for the course.

2.32 Biological clocks. Many plants and animals have "biological clocks" that coordinate activities with the time of day. When researchers looked at the length of the biological cycles in the plant *Arabidopsis* by measuring leaf movements, they found that the length of the cycle is not always 24 hours. The researchers suspected that the plants adapt their clocks to their north-south position. Plants don't know geography, but they do respond to light, so the researchers looked at the relationship between the plants' cycle lengths and the length of the day on June 21 at their locations. The data file includes data on cycle length and day length, both in hours, for 146 plants.[15] Plot cycle length as the response variable against day length as the explanatory variable. Does there appear to be a positive association? Is it a strong association? Explain your answers.
BIOCLOCKS

2.41 Brand names and generic products.

(a) If a store always prices its generic "store brand" products at 90% of the brand name products' prices, what would be the correlation between the prices of the brand name products and the store brand products? (*Hint:* Draw a scatterplot for several prices.)

2.42 Strong association but no correlation. Here is a data set that illustrates an important point about correlation: CORRELATION

X	20	30	40	50	60
Y	10	30	50	30	10

(a) Make a scatterplot of Y versus X.

(b) Describe the relationship between Y and X. Is it weak or strong? Is it linear?

(c) Find that the correlation between Y and X.

(d) What important point about correlation does this exercise illustrate?

2.47 Second test and final exam. In Exercise 2.25 you looked at the relationship between the score on the second test and the score on the final exam in an elementary statistics course. Here are the data: STATCOURSE8

Second-test score	158	162	144	162	136	158	175	153
Final-exam score	145	140	145	170	145	175	170	160

(a) Find the correlation between these two variables.

2.49 The effect of a different point. Examine the data in the Exercise 2.47 and add a ninth student who has low scores on the second test and the final exam and fits the overall pattern of the other scores in the data set. Calculate the correlation and compare it with the correlation that you calculated in Exercise 2.47. Write a short summary of your findings.

2.50 The effect of an outlier. Refer to the Exercise 2.47. Add a ninth student whose scores on the second test and final exam would lead you to classify the additional data point as an outlier. Recalculate the correlation with this additional case and summarize the effect it as on the value of the correlation.

2.54 An interesting set of data. Make a scatterplot of the following data:

x	1	2	3	4	10	10
y	1	3	3	5	1	11

Use your calculator to show that the correlation is about 0.5. What feature of the data is responsible for reducing the correlation to this value despite a strong straight-line association between x and y in most of the observations?
INTERESTING

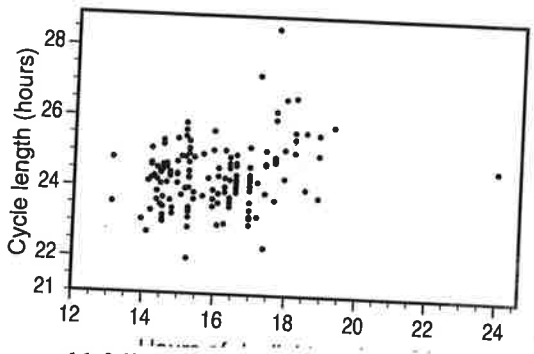
Solutions

2.14. **(a)** The points should all fall close to a negatively sloped line. **(b)** Look for a "cloud" of points with no discernible pattern. Watch for students who mistakenly consider "no relationship" as meaning "no linear relationship." For example, points that suggest a circle, triangle, or curve may indicate a non-linear relationship. **(c)** The points should be widely scattered around a positively sloped line. **(d)** Sketches might be curved, angular, or something more imaginative.

2.20. **(a)** Figure 2.11 shows a positive curved relationship. More specifically, in countries with fewer than 10 Internet users per 100 people, life expectancy ranges between 40 and about 75 years. For countries with more than 10 Internet users per 100 people, life expectancy increases (slowly) with increasing Internet usage. **(b)** A more likely explanation for the association is that countries with higher Internet usage are typically more developed and affluent, which comes with benefits such as better access to medical care, etc.

2.27. **(a)** Age is explanatory; weight is the response variable. **(b)** Explore the relationship; there is no reason to view one or the other as explanatory. **(c)** Number of bedrooms is explanatory; price is the response variable. **(d)** Amount of sugar is explanatory; sweetness is the response variable. **(e)** Explore the relationship.

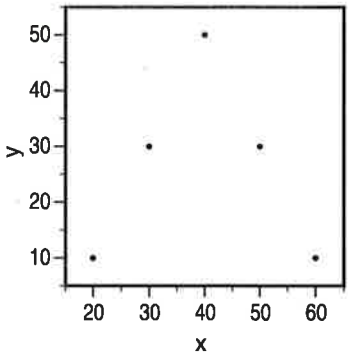
2.32. There appears to be a positive association between cycle length and day length, but it is quite weak: The points of the scatterplot are generally located along a positively sloped line but with a lot of spread around that line. (Ideally, both axes should have the same scale.)



2.41. In both these cases, the points in a scatterplot would fall exactly on a positively sloped line, so both have correlation $r = 1$. **(a)** With $x =$ the price of a brand-name product, and $y =$ the store-brand price, the prices satisfy $y = 0.9x$. **(b)** The prices satisfy $y = x - 1$.

2.42. **(a)** Scatterplot on the right. **(b)** The relationship is very strong (assuming these five points are truly representative of the overall pattern), but it is not linear. **(c)** The correlation is $r = 0$. **(d)** Correlation only measures the strength of *linear* relationships.

2.42 →



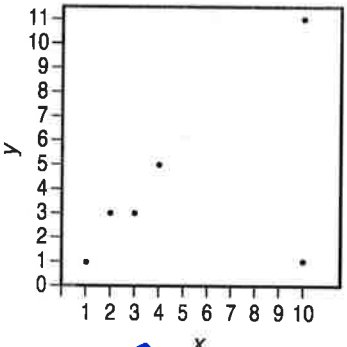
2.47. **(a)** $r \doteq 0.5194$. **(b)** The first-test/final-exam correlation will be lower, because the relationship is weaker. (See the next solution for confirmation.)

2.49. Such a point should be at the lower left part of the scatterplot. Because it tends to strengthen the relationship, the correlation increases.

Note: *In this case, r was positive, so strengthening the relationship means r gets* larger. *If r had been negative, strengthening the relationship would have* decreased *r (toward −1).*

2.50. Any outlier should make r closer to 0, because it weakens the relationship. To be considered an outlier, the point for the ninth student should be in either the upper left or lower right portion of the scatterplot. The former would correspond to a student who had a below-average second-test score but an above-average final-exam score. The latter would be a student who did well on the second test but poorly on the final.

Note: *In this case, because r > 0, this means r gets* smaller. *If r had been negative, getting closer to 0 would mean that r gets* larger *(but gets smaller in absolute value).*



2.34 ↑ Outlier especially for low correlation