Unit 10: Scatterplots

PREREQUISITES

Units 10, 11, and 12 form a natural cluster on describing relationships between two quantitative variables. This unit on scatterplots can be taught at a lower level than Units 11 and 12 on regression and correlation. For this unit, students need to be able to draw axes and plot ordered pairs.

ADDITIONAL TOPIC COVERAGE

Additional coverage on scatterplots can be found in *The Basic Practice of Statistics*, Chapter 4, Scatterplots and Correlation

ACTIVITY DESCRIPTION

For this activity, you will need to collect the following data from one or more classes.

- Have male students record their heights and the heights of their fathers.
- Have female students record their heights and the heights of their mothers.

Using the parent's height as the explanatory variable, students make scatterplots of student height versus parent height for females and the same for males. By combining the data and using different symbols (or colors) for males and females, students can work with multivariate data that consists of two quantitative variables and one categorical variable.

If you decide *not* to have students from your class collect the data, then use the sample data in Table T10.1, which follows this activity description. The sample answers are based on these data, which were collected from several sections of a college introductory statistics course. If your class data are from high school students, don't be surprised if you find a linear pattern for females and you do not find a linear pattern for males. Males finish growing later than females and that may influence the pattern that appears in the scatterplots.

Females		Males	
Student Height	Mother's Height (inches)	Student Height	Father's Height (inches)
62	61	75	74
66	66	72	70
68	64	70	68
60.25	62	72.5	74
68	66	67.7	71
64	61	73	70
62	65	67	61
67	64	67.7	68
66	70	65	67
63	66	71	71
68	66	70	68
64	63	75	76
60	62	73	70
69	65	70	67
66	65	71	74
67	62	69	68
66.5	64	71	72
65	64	72	70
66	65	74	72
62	62	73	68
64	69	69	72
69	70	72	70
64	66	80	70
57	62	69	68
64	60	72	78
63	65	70	68
60	63	70	68
62	59	68	68
66	67	66	68
		69	68
		72	68
		72	68

Table 10.1: Sample Class Data

THE VIDEO SOLUTIONS

1. A manatee is a large, slow-moving sea mammal. They can weigh more than half a ton and are mainly plant eaters.

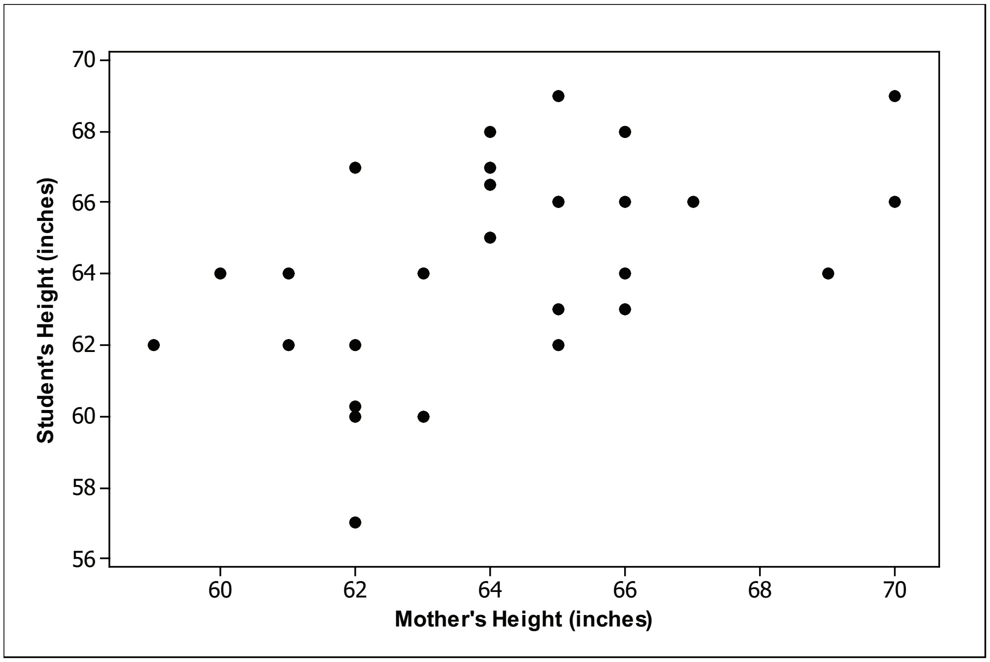
2. There is a positive association between manatees killed by powerboats and the number of powerboat registrations. In other words, as the number of powerboat registrations increases, the number of manatees killed also tends to increase.

3. The number of powerboat registrations is the explanatory variable.

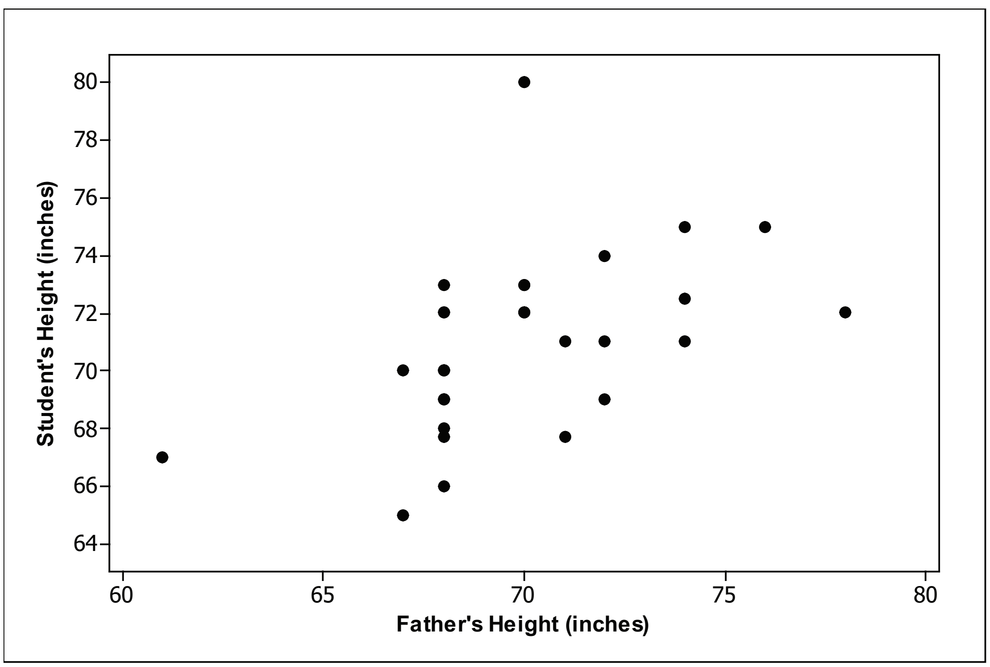
4. As one variable increases, the other tends to decrease. For example, in making pies the time it takes for you to make a particular type of pie decreases with the number of times you have made that type of pie. (The more pies you make the faster you get.)

UNIT ACTIVITY SOLUTIONS

1. Sample answer based on sample class data. (See Activity Description for sample class data.) Students could choose different scaling for axes.

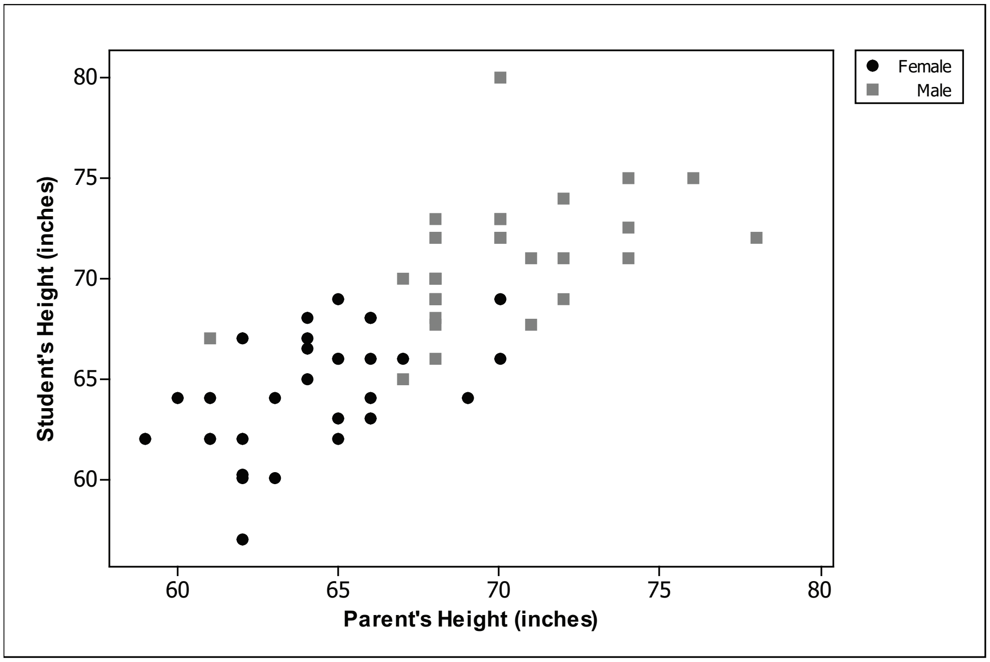


2. Sample answer:



3. Sample answer: Both scatterplots appear to have linear form with positive association. The linear trend for the males appears stronger than the linear trend for the females, but that could be just because different scaling has been used for the two graphs. For the males, there appear to be two outliers. One outlier seems to be consistent with the overall pattern, but that male is the shortest male student (and his father is also on the short side). However, there is one student whose father appears to be average in height, but that student is the tallest student in the class.

4. Sample answer:



5. Sample answer: The scatterplot of the combined data has linear form, which appears to be moderately strong. As might be expected, most of the data from the female students appears in the lower left of the scatterplot, which indicates that female students and their mothers tended to be shorter than male students and their fathers.

EXERCISE SOLUTIONS

1. a. The amount of time spent studying for a statistics exam is the explanatory variable and the grade on the exam is the response variable.

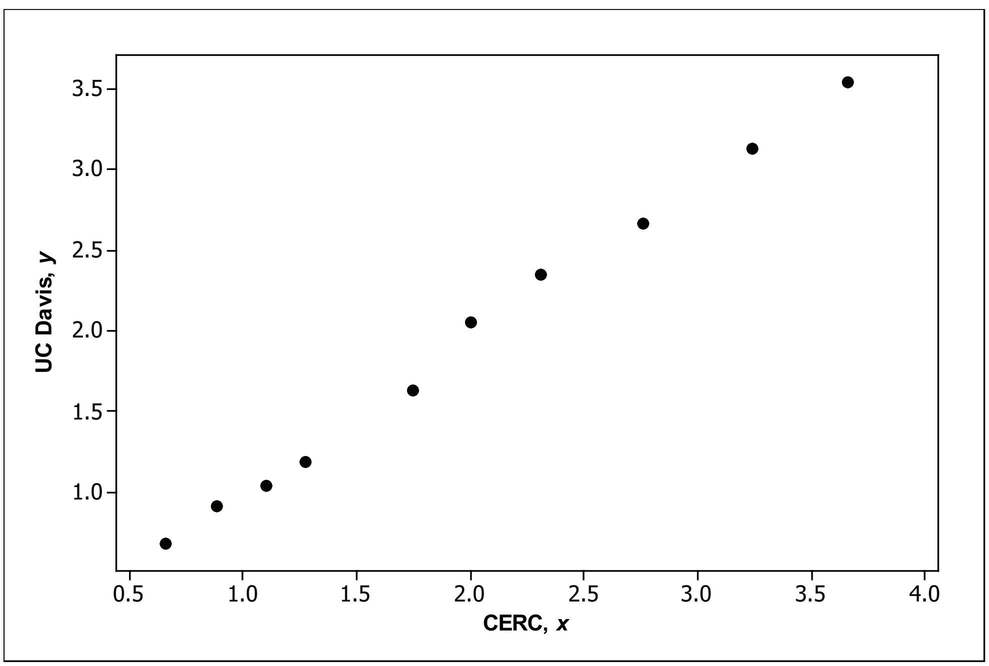
b. Sample answer #1: There is a relationship but neither explains the other.

Sample answer #2: Taller people tend to be heavier and hence height is the explanatory variable and weight is the response variable. You might want to try to predict a person's weight given how tall they are.

c. Rainfall helps explain crop yields, so rainfall is the explanatory variable and crop yields is the response variable.

d. There is a relationship between hand length and foot length. However, neither explains the other.

2. a.

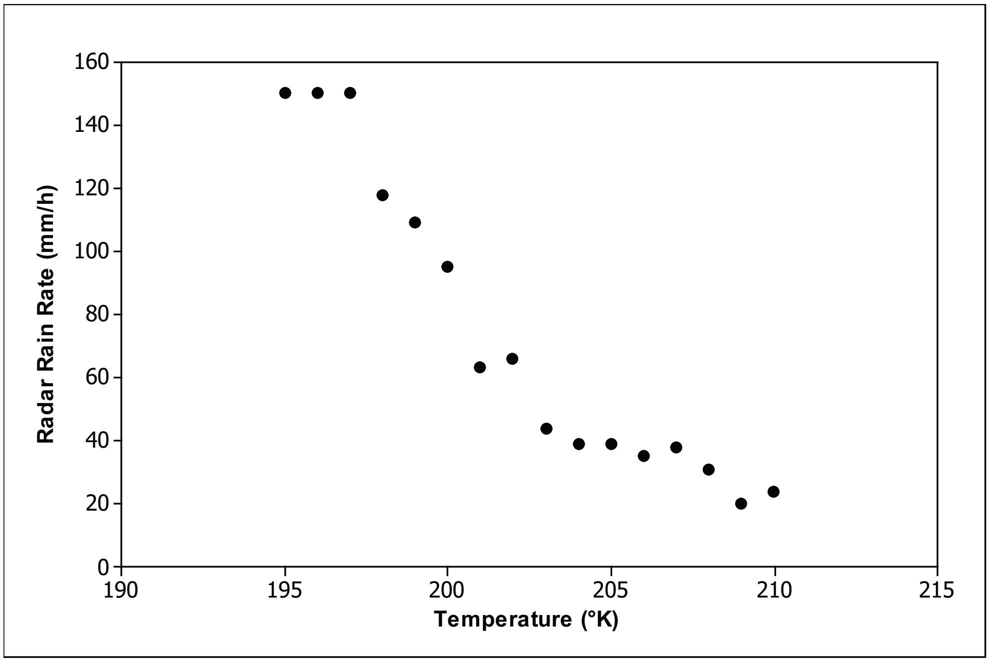


b. This is an example of positive association. Above-average mercury concentrations from CERC are associated with above-average mercury concentrations from UC Davis and below-average mercury concentrations from CERC are associated with below.

c. The pattern is linear because the dots fall close to a straight line.

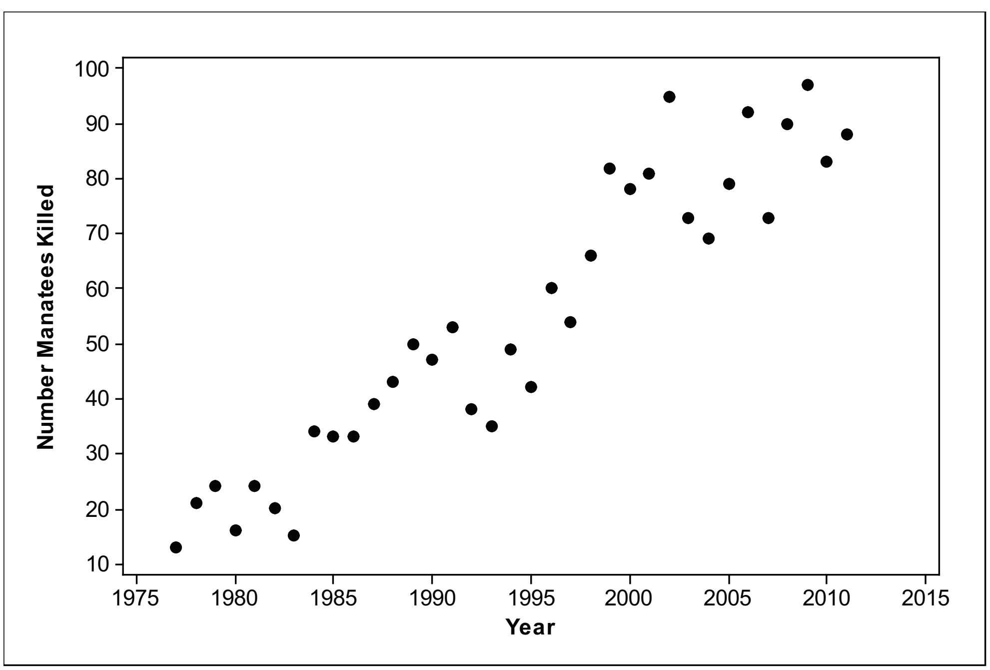
3. a. The temperature is the explanatory variable. We would like to use cloud top brightness temperatures to explain rain rate.

b.

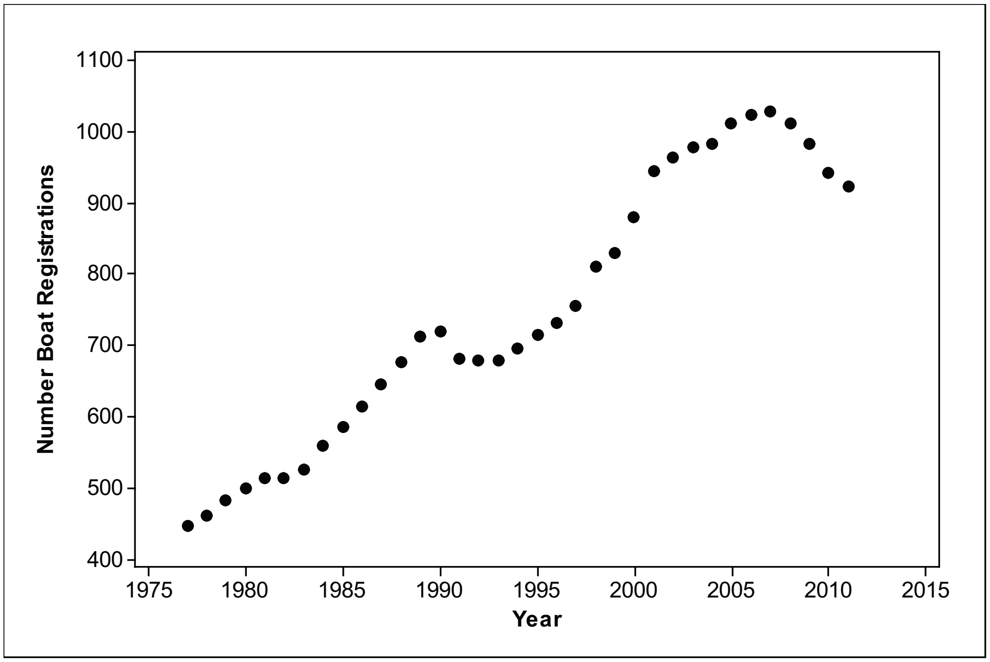


c. The association is negative – as temperature increases, rain rate tends to decrease. The pattern appears roughly linear. However, there does appear to be a slight upward bend to the pattern.

4. a. The pattern appears to be linear. As the years increase, the number of manatees killed by powerboats also tends to increase.



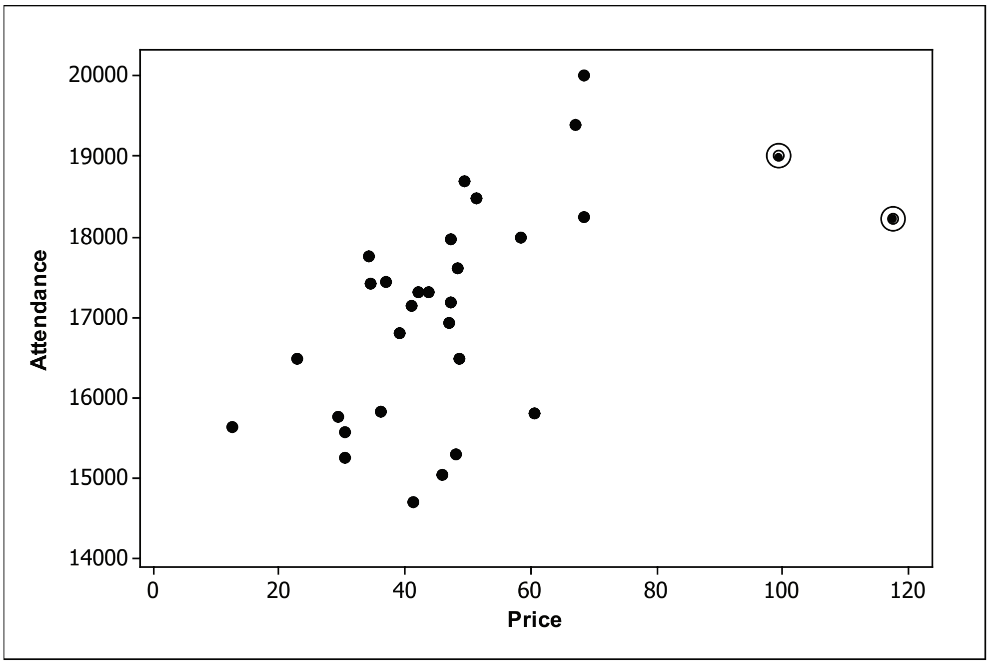
b. The pattern appears to be nonlinear. The association between the two variables appeared positive up until 2007, then it switched to negative.



REVIEW QUESTIONS SOLUTIONS

1. a. Better teams may be more popular. So, a good team can charge higher prices and also have higher attendance than a weak team. That's a *positive association.* However, high prices will drive some fans away, so we might expect higher prices to go with lower attendance. That's a negative association. We need to have the actual data before we know which of the situations described above is correct.

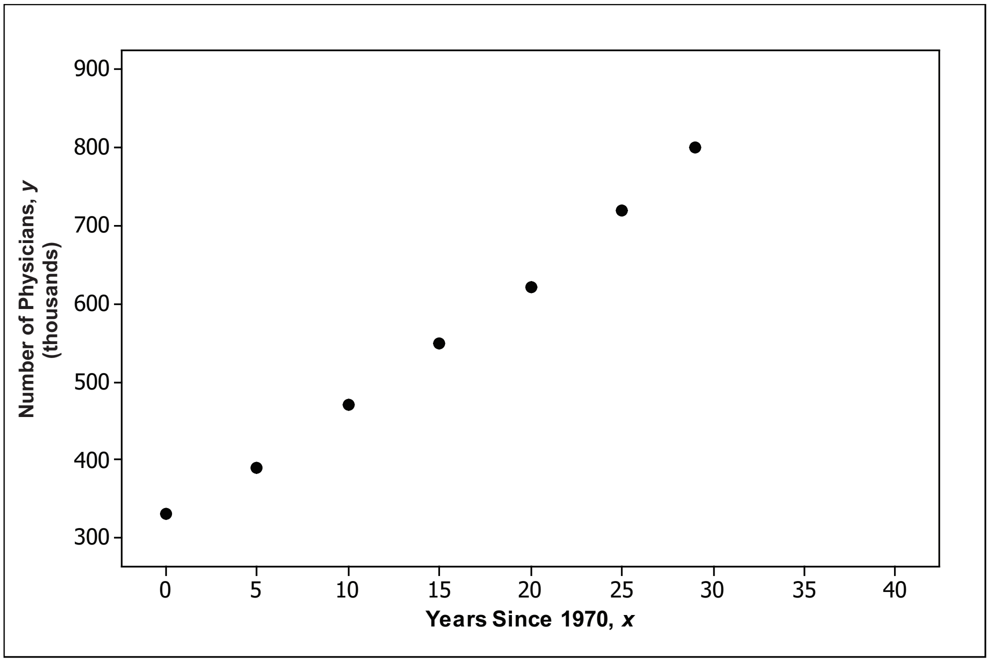
b.



Sample answer: The relationship between price and attendance appears to be a positive association. In this scatterplot, we are using ticket price to explain attendance. As price increases, attendance also tends to increase. There appear to be two outliers from what otherwise might be a linear pattern. The teams associated with these two outliers are the New York Knicks and the Los Angeles Lakers, the two teams with the highest ticket prices. These outliers could be an indication that after a certain point, high prices will begin to adversely affect attendance levels.

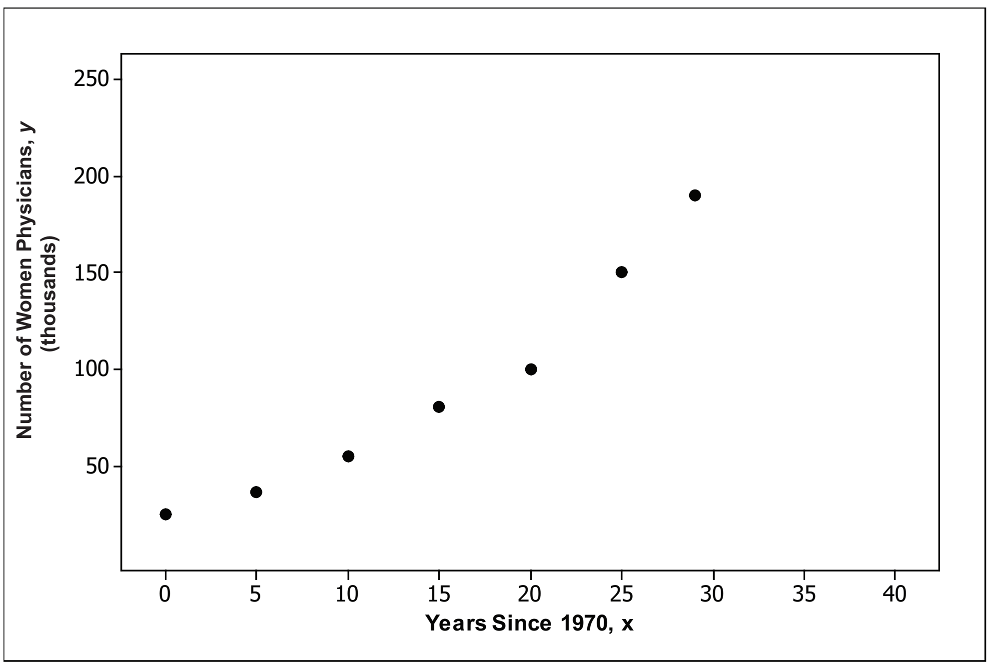
2. a) Values for *x* (second column): 0, 5, 10, 15, 20, 25, 29.

b) Note: Entries in Total Number of Physicians column in Table 10.7 have been rounded to nearest 10,000.



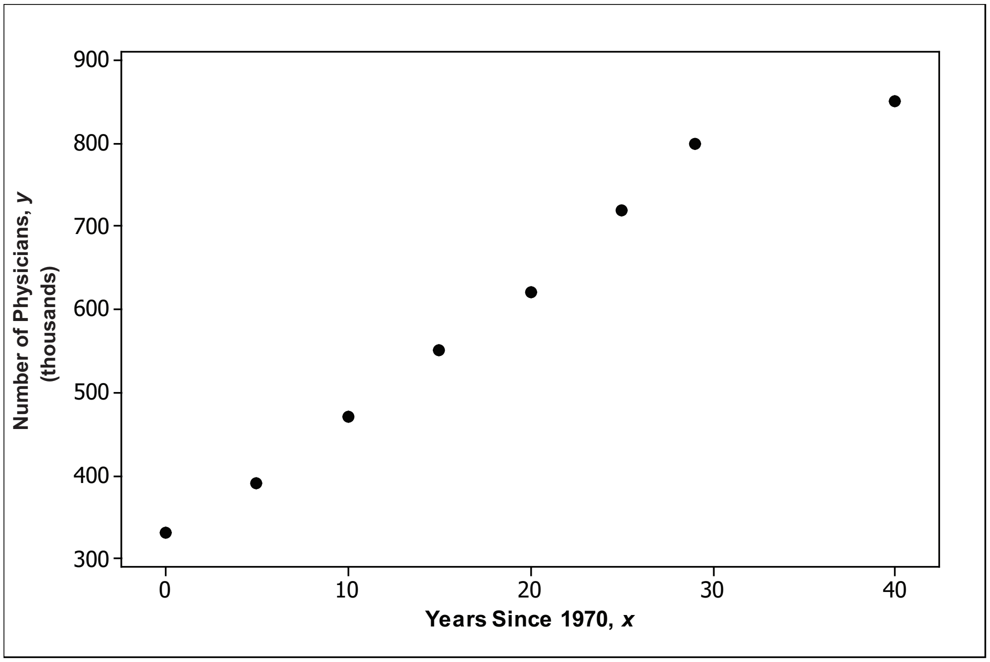
c. The pattern of the dots appears to have linear form.

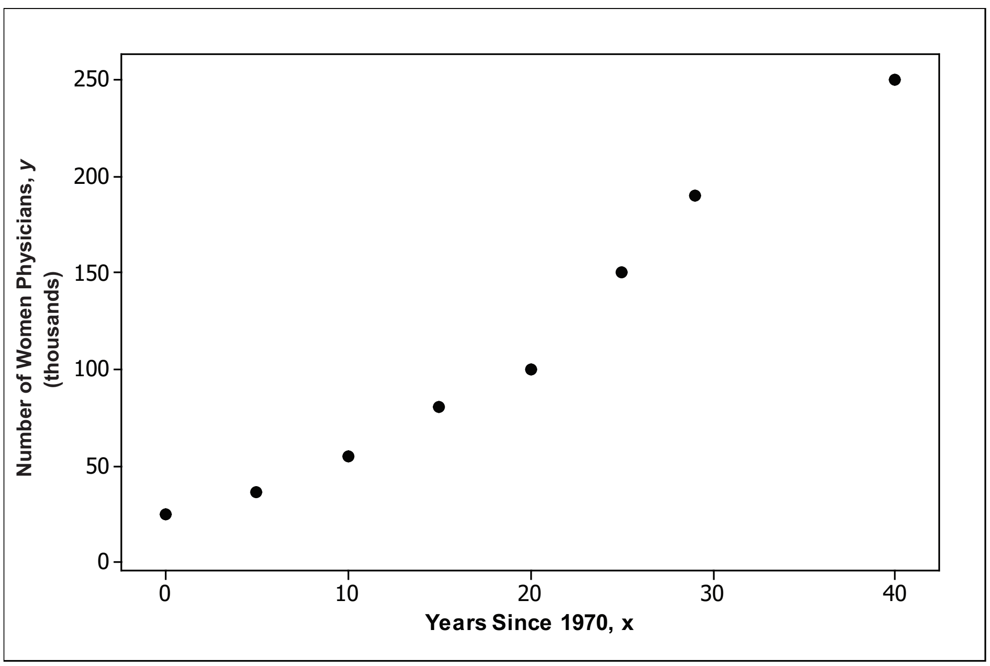
d.



e. Sample: The scatterplot of number of women physicians versus *x* is nonlinear. The pattern of dots appears to curve upward as the values of *x* increase. No matter what line you try to draw, the pattern of points will not appear randomly scattered above and below your line.

f.

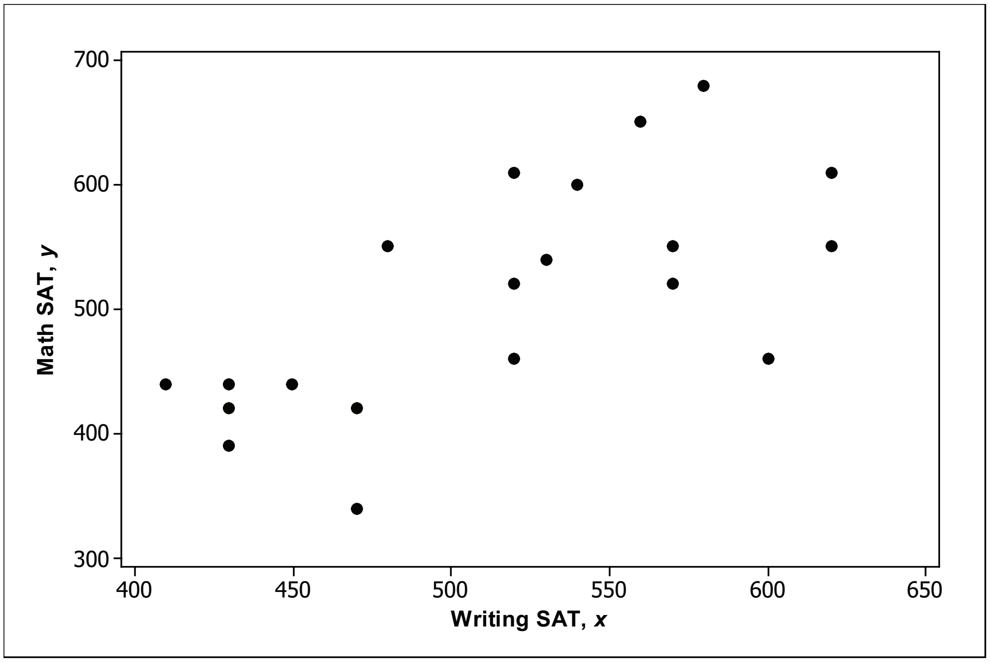




The added data value corresponding to the year 2010 deviates from the overall pattern of the original data. It looks as if the growth pattern slowed down between 1999 and 2010 compared to the growth pattern from 1979 to 1999.

Sample answer: The added data value corresponding to the year 2010 appears to fit fairly well with the overall pattern of the original data. However, it appears slightly lower than what might be expected – this could indicate that the rate of increase is slowing or it could just be normal variability of data about the overall curved pattern.

3. a.



b. There is a positive association between Math and Writing SAT scores. Students with above average Writing SATs also tend to have above average Math SATs and students with below average Writing SATs tend to have below average Math SATs. The relationship appears to have linear form.