Chapter 6
Exploring Data: Relationships

Solutions

Exercises:

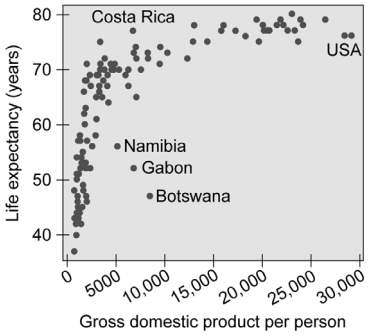
1. (a) It is more reasonable to explore study time as an explanatory variable and the exam grade as the response variable.

(b) It is more reasonable to explore the relationship only.

(c) It is more reasonable to explore rainfall as an explanatory variable and the corn yield as the response variable.

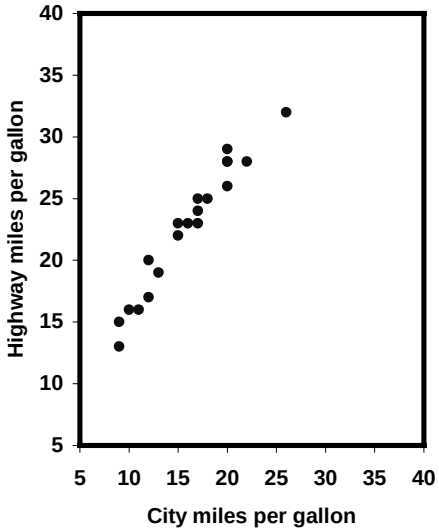
(d) It is more reasonable to explore the relationship only.

3. (a) Life expectancy increases with GDP in a curved pattern. The increase is very rapid at first, but levels off for GDP above roughly $5000 per person.

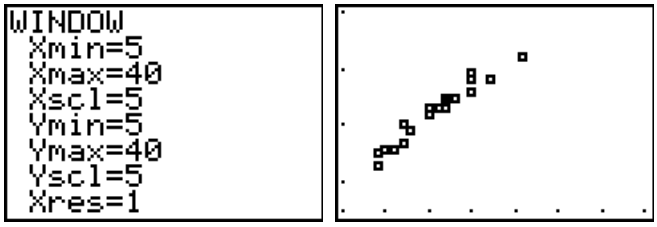


(b) Richer nations have better diets, clean water, and better health care, so we expect life expectancy to increase with wealth. But once food, clean water, and basic medical care are in place, greater wealth has only a small effect on lifespan.

5. (a) The scatterplot is as follows.



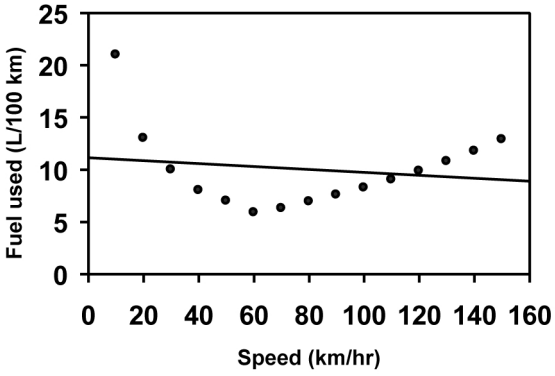
Using a TI-83, we get the following.



Purists should notice that because the variables measure similar quantities the plot is square with the same scales on both axes.

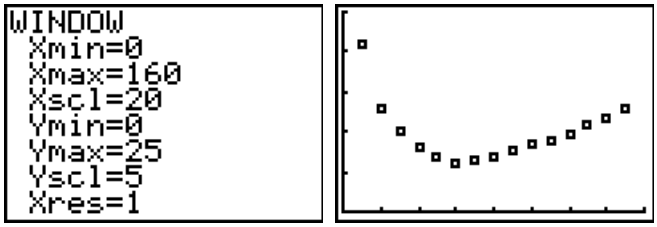
(b) There is a strong positive straight-line relationship.

7. (a) Here is the scatterplot, with speed as the explanatory variable:



Continued on next page

7. (a) continued
Using a TI-83, we get the following.



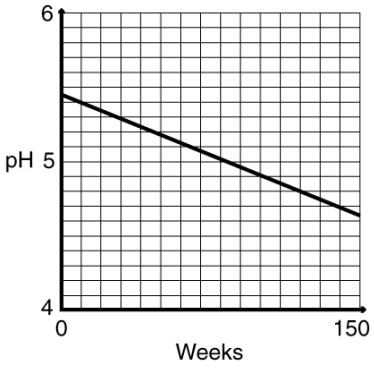
(b) The relationship is curved; fuel usage first decreases as speed increases (higher gears cover more distance per motor revolution) then increases as speed is further increased (air resistance builds at higher speeds).

(c) There is no overall direction.

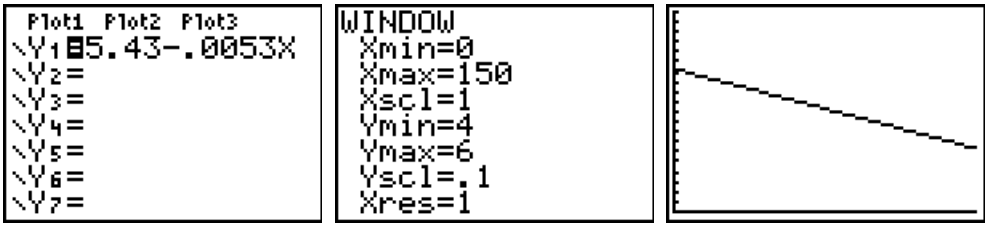
(d) The relationship is quite strong. There is little scatter about the overall curved pattern.

9. The estimated slope would be $\dfrac{506-386}{191-139}=\dfrac{120}{52}\approx 2.31$.

11. (a) Choose two values of weeks, preferably near 1 and 150. Find pH from the equation given, plot the two points (weeks horizontal) and draw the line through them.



Using a TI-83, we get the following.



(b) Week 1: predicted pH $=5.43-(0.0053\times 1)=5.43-0.0053=5.4247\approx 5.42$

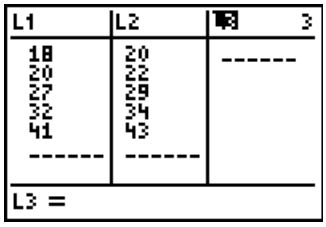
Week 150: predicted pH $=5.43-(0.0053\times 150)=5.43-0.795=4.635\approx 4.64$

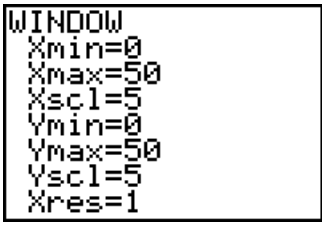
(c) The slope -0.0053 says that on average pH declined by 0.0053 per week during the study period.

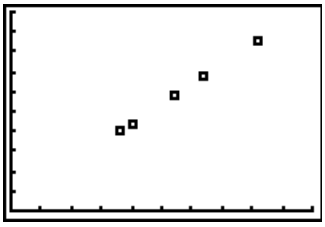
13. Some sample ages would be as follows.

Age of Woman	Age of Man
18	20
20	22
27	29
32	34
41	43

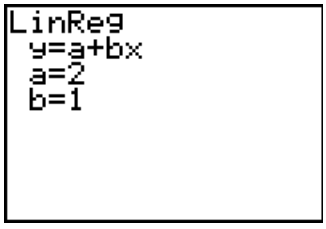
Using a TI-83, we get the following.





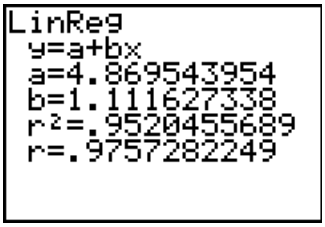


Using the linear regression feature, we obtain the following.

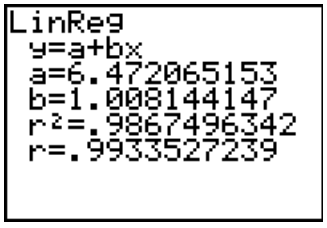


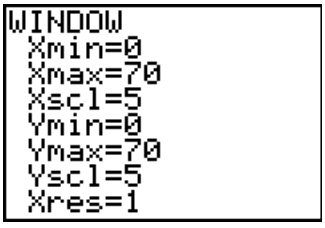
The line is $y = x + 2$, and the slope would be 1.

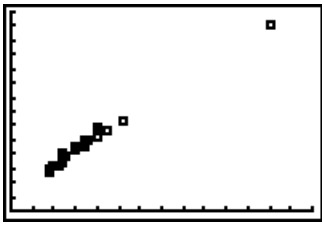
15. With a correlation of 0.9757, the indication is a very strong straight-line pattern.



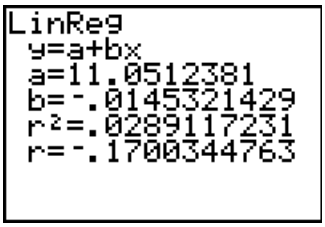
17. The correlation is 0.9934. The correlation is stronger when the Insight is added, because that point extends (strengthens) the straight-line pattern.



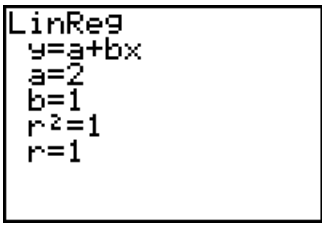




19. See the answer to Exercise 7 for the scatterplot. The correlation is –0.1700. Correlation measures the strength of only linear (straight-line) relationship. This relationship is strong but curved.



21. The correlation would be 1 because there is a perfect straight-line relationship, $y = x + 2$. This is verified by using the sample data from Exercise 13.



23. (a) Negative: older cars will in general sell for lower prices.

(b) Negative: heavier cars will (other things being equal) get fewer miles per gallon.

(c) Positive: taller people are on average heavier than shorter people.

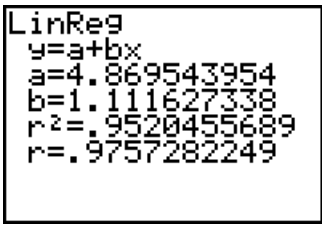
(d) Small: there is no reason to expect that height and IQ are related.

25. Ask how similar the market sector of each fund is to large U.S. stocks and arrange in order.

(a) Dividend Growth, $r = 0.98$; Small Cap Stock, $r = 0.81$; Emerging Markets, $r = 0.35$.

(b) No: it just says that they tend to move in the same direction, whether up or down.

27. (a) Predicted MPG $= 4.87 + 1.11x$.



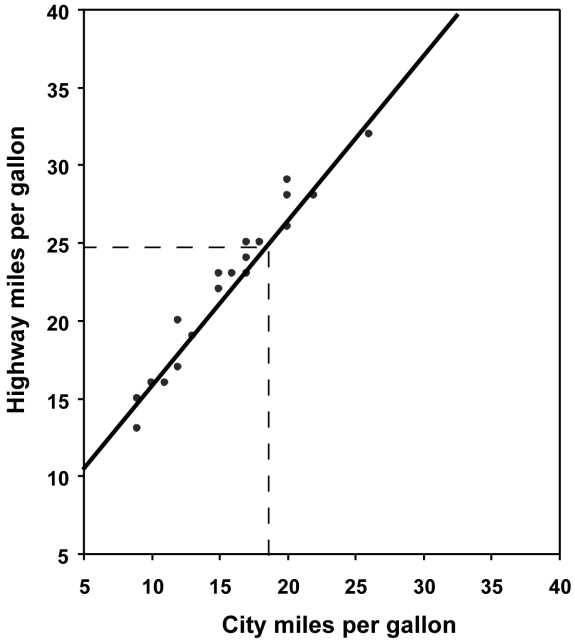
(b) Predicted MPG $= 4.87 + 1.11(18) = 4.87 + 19.98 = 24.85$ MPG.

(c) We assess accuracy from how closely the points in the plot follow a straight line. Looking at the plot in Exercise 5, we expect quite accurate predictions.

29. Choose two city MPG values, such as $x = 10$ and $x = 30$, and use the equation of the line to find each value of y. Plot the two points and draw the line between them. Here is the plot.

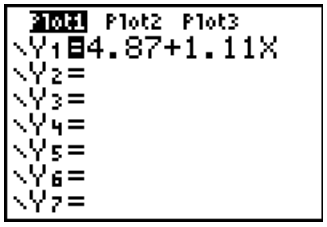
$x = 10$: Predicted MPG $= 4.87 + 1.11(10) = 4.87 + 11.1 = 15.97$ MPG.

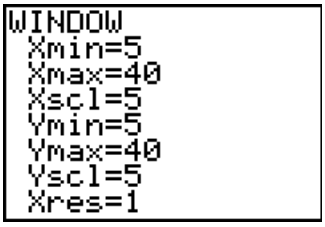
$x = 30$: Predicted MPG $= 4.87 + 1.11(30) = 4.87 + 33.3 = 38.17$ MPG.

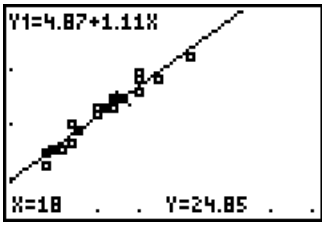


Predicted highway mileage of a car that gets 18 mpg in the city is approximately 25 mpg (24.85 mph) from exercise 27.

Using a TI-83, we get the following.







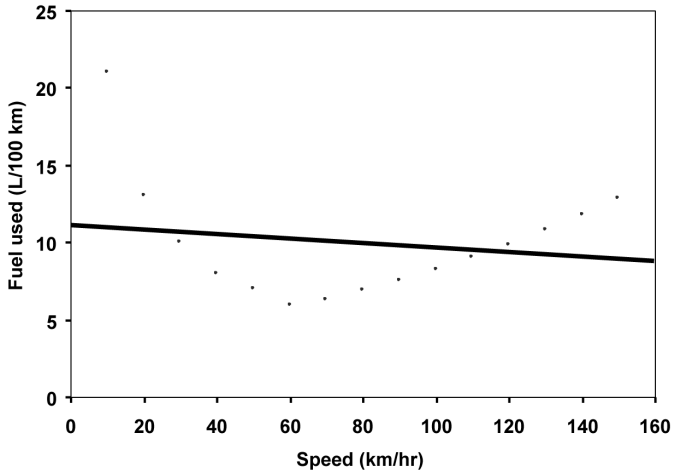
31. Since predicted fuel $= 11.058 - 0.0147 \times \text{speed}$, we have the following.

Speed = 10 kph: predicted fuel $= 11.058 - 0.0147 \times 10 = 11.058 - 0.147 = 10.911$ kpg

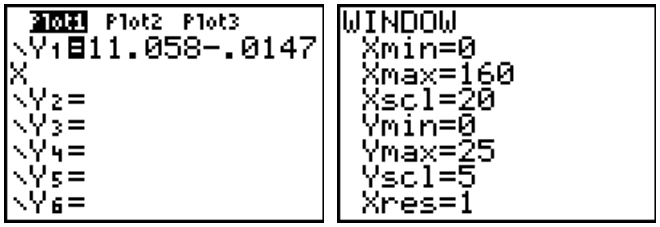
Speed = 70 kph: predicted fuel $= 11.058 - 0.0147 \times 70 = 11.058 - 1.029 = 10.029$ kpg

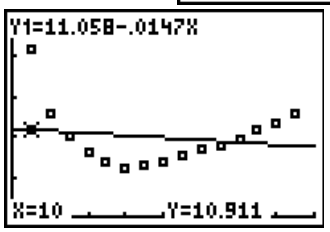
Speed = 150 kph: predicted fuel $= 11.058 - 0.0147 \times 150 = 11.058 - 2.205 = 8.853$ kpg

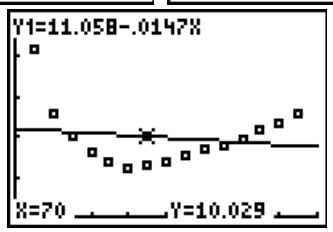
The predicted values from the equation given are approximately 10.9, 10.0, and 8.85, respectively. The observed values are 21.00, 6.30, and 12.88, respectively. The least-squares line gives the best straight-line fit, which is of little value here.

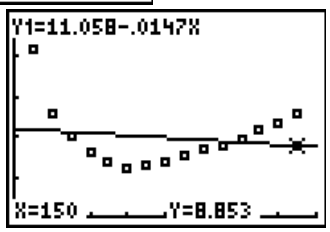


Using a TI-83, we get the following.









33. The slope of the least-squares line is $b = r\dfrac{s_y}{s_x} = 0.5\left(\dfrac{2.7}{2.5}\right) = 0.54$. The intercept is as follows.

$$a = \bar{y} - b\bar{x} = 68.5 - (0.54)(64.5) = 68.5 - 34.83 = 33.67$$

For $x = 67$, we predict $33.67 + (0.54)(67) = 33.67 + 36.18 = 69.85$ inches.

35. The predicted y for $x = \bar{x}$ is as follows.

$$\text{predicted } y = a + b\bar{x} = \left(\bar{y} - b\bar{x}\right) + \left(r\frac{s_y}{s_x}\right)\bar{x} = \bar{y} - \left(r\frac{s_y}{s_x}\right)\bar{x} + \left(r\frac{s_y}{s_x}\right)\bar{x} = \bar{y}$$

37. First compare the distributions for the two years. To make the boxplots, we need the five-number summary for each data set.

Putting the 2002 data in order, we have the following.

$$-50.5,\ -49.5,\ -47.8,\ -42,\ -37.8,\ -26.9,\ -23.4,\ -21.1,\ -18.9,\ -17.2,\ -17.1,$$
$$-12.8,\ -11.7,\ -11.5,\ -11.4,\ -9.6,\ -7.7,\ -6.7,\ -5.6,\ -2.3,\ -0.7,\ -0.7,\ 64.3$$

The minimum is -50.5 and the maximum is 64.3. The median is the $\frac{23+1}{2} = \frac{24}{2} = 12^{\text{th}}$ piece of data, namely -12.8. Since there are 11 observations to the left of the median, Q_1 is the $\frac{11+1}{2} = \frac{12}{2} = 6^{\text{th}}$ piece of data, namely -26.9. Since there are 11 observations to the right of the median, Q_3 is the $12 + 6 = 18^{\text{th}}$ piece of data, namely -6.7.

Thus, the five - number summary is $-50.5,\ -26.9,\ -12.8,\ -6.7,\ 64.3$.

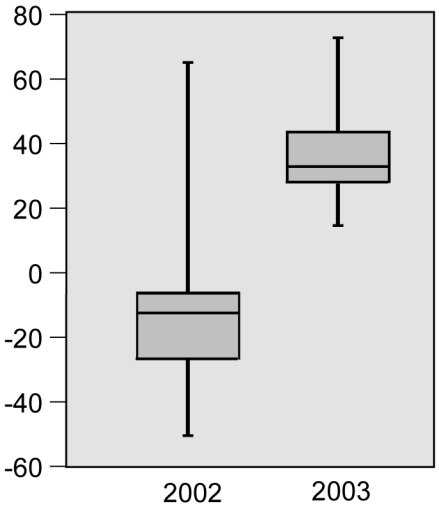
Putting the 2003 data in order, we have the following.

$$14.1,\ 19.1,\ 22.9,\ 23.9,\ 26.1,\ 27.5,\ 28.7,\ 29.5,\ 30.6,\ 31.1,\ 32.1,$$
$$32.3,\ 35.0,\ 36.5,\ 36.9,\ 36.9,\ 41.8,\ 43.9,\ 57.0,\ 59.4,\ 62.7,\ 68.1,\ 71.9$$

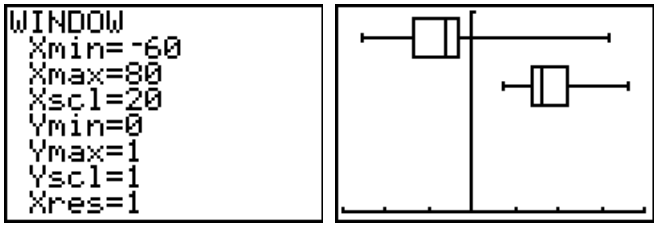
The minimum is 14.1 and the maximum is 71.9. The median is the $\frac{23+1}{2} = \frac{24}{2} = 12^{\text{th}}$ piece of data, namely 32.3. Since there are 11 observations to the left of the median, Q_1 is the $\frac{11+1}{2} = \frac{12}{2} = 6^{\text{th}}$ piece of data, namely 27.5. Since there are 11 observations to the right of the median, Q_3 is the $12 + 6 = 18^{\text{th}}$ piece of data, namely 43.9.

Thus, the five - number summary is 14.1, 27.5, 32.3, 43.9, 71.9.

Here are boxplots.



Using a TI-83, we get the following. The top boxplot is for 2002 and the bottom one is for 2003.

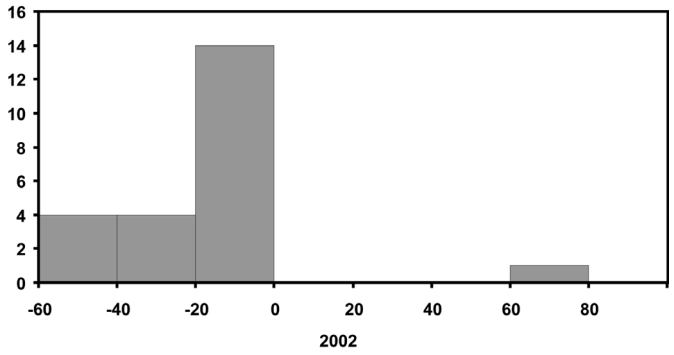


Continued on next page

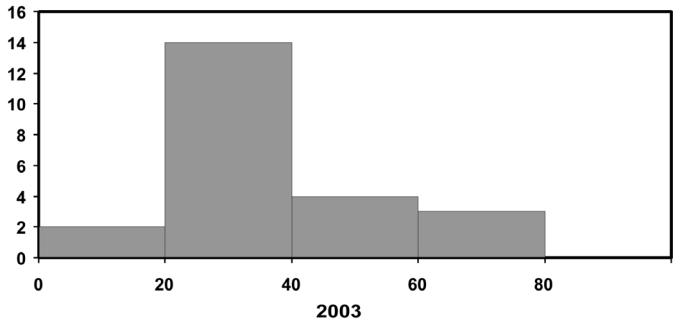
37. continued

Histograms (or stemplots) show that the 2003 returns are roughly single-peaked and symmetric, and that the 2002 returns are left-skewed with an extreme high outlier. Below are the histograms.

Class	Count
– 60 – (– 40.1)	4
– 40 – (– 20.1)	4
– 20 – (– 0.1)	14
0 – 19.9	0
20 – 39.9	0
40 – 59.9	0
60 – 79.9	1

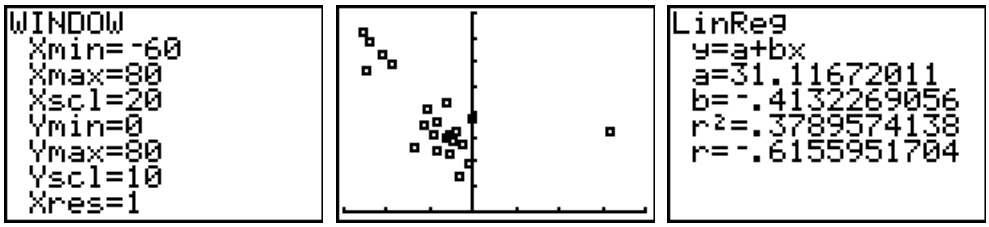


Class	Count
0 – 19.9	2
20 – 39.9	14
40 – 59.9	4
60 – 79.9	3

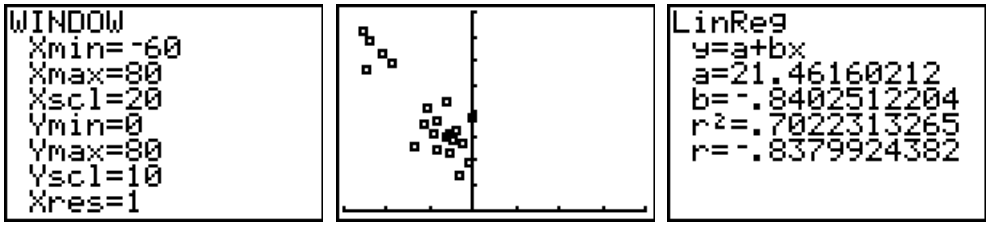


The median returns are – 12.8% in 2002 and 32.1% in 2003 (from the five - number summary). The correlation is $r = -0.616$; because of the influence of outliers on correlation, it is better to report the correlation without the outlier, $r = -0.838$.

With Outlier

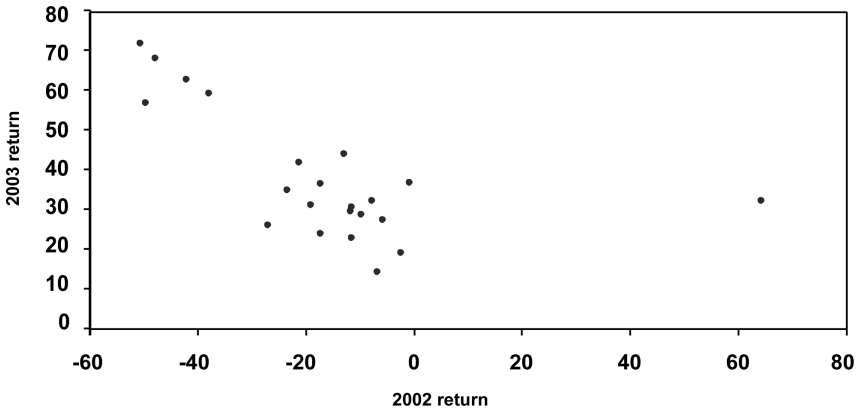


Without Outlier (64.3, 32.1)



Continued on next page

37. continued
That is, the funds that went down most in 2002 tended to go up most in 2003. The scatterplot confirms this:



39. (a) All four sets of data have r = 0.816 and regression line $y = 3.0 + 0.5x$ to a close approximation.

Set A

```
LinReg
 y=a+bx
 a=3.000090909
 b=.5000909091
 r²=.6665424595
 r=.8164205163
```

Set B

```
LinReg
 y=a+bx
 a=3.000909091
 b=.5
 r²=.6662420337
 r=.816236506
```

Set C

```
LinReg
 y=a+bx
 a=3.002454545
 b=.4997272727
 r²=.6663240411
 r=.8162867395
```

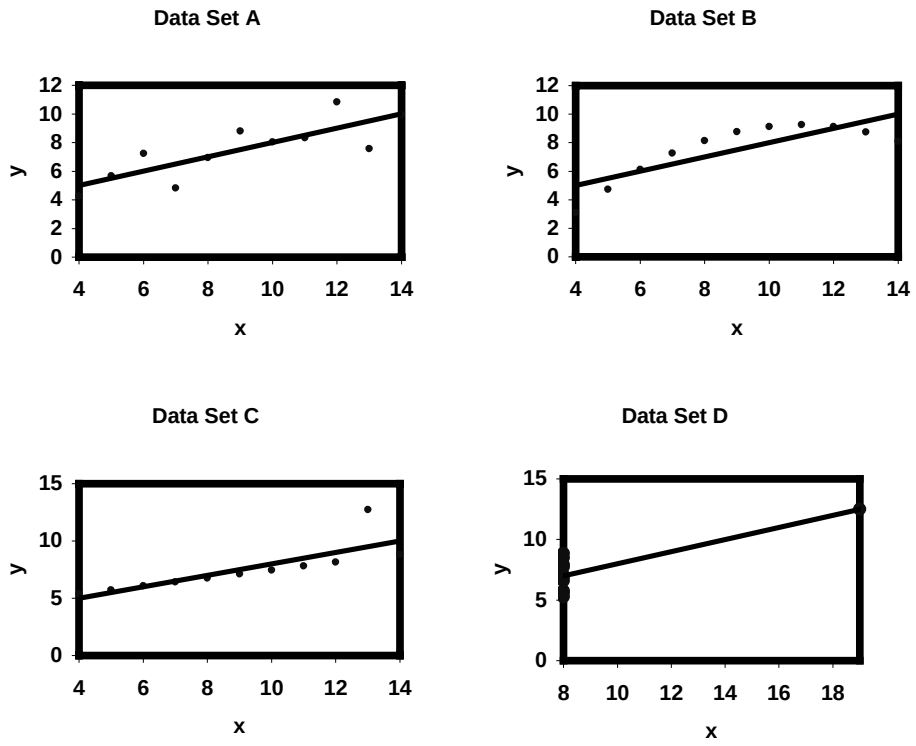
Set D

```
LinReg
 y=a+bx
 a=3.001727273
 b=.4999090909
 r²=.6667072569
 r=.8165214369
```

Continued on next page

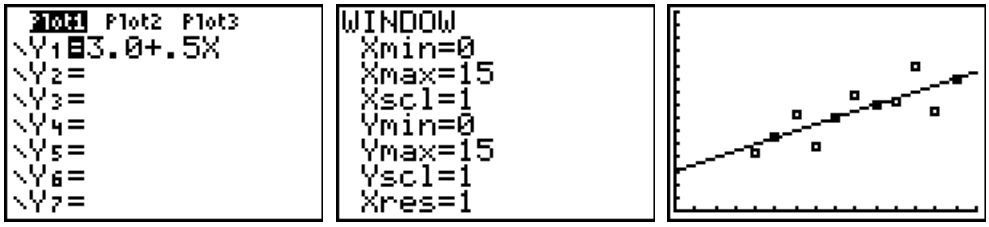
39. continued

(b) Here are the plots:

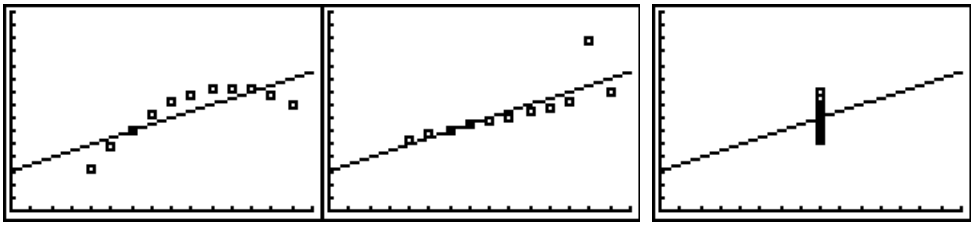


Using a TI-83, we get the following.

Set A



Set B Set C Set D



(c) Only A is a normal" regression setting in which the line is useful for prediction. Plot B is curved, C has an extreme outlier in *y*, and D has all but one *x* identical. The lesson: Plot your data before calculating.

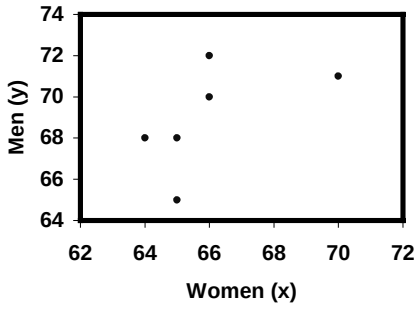
In Exercises 41 & 43, answers will vary.

41. Heavier people who are concerned about their weight may be more likely than lighter people to choose artificial sweeteners in place of sugar.

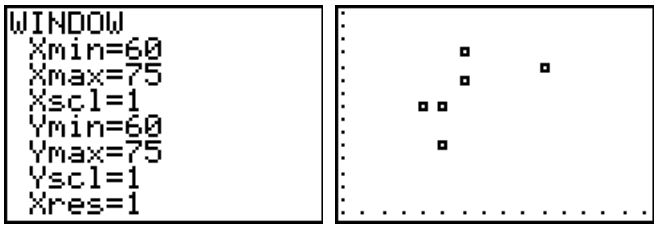
43. Higher income generally means better water and sewage utilities, better diet, and better medical care, which will produce better health. But better health means more children can go to school and more workers are able to work and can stay on the job, which raises national income. For example, AIDS is having a direct negative effect on the economies of African nations.

45. Explanatory: parents' income. Response: amount of college debt.
We expect a negative association: children of richer parents do not need to borrow as much to pay for college.

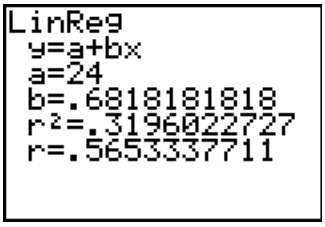
47. (a) There is a positive association, so r will be positive. The pattern is a bit irregular, so r won't be close to 1.



Using a TI-83, we get the following.



(b) $r = 0.5653$.



49. (a) The slope is $b = 0.68$. For each additional inch of women's height, the height of the next man dated goes up by 0.68 inch on average.

(b) The prediction is as follows.

predicted male height = $24 + 0.68 \times$ female height
$= 24+(0.68)(67) = 24 + 45.56 = 69.56$ inches