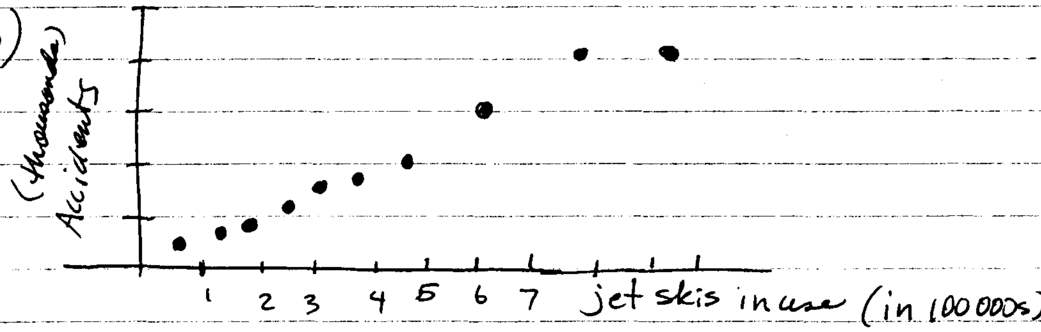
AP Stats 3.5, 7, 8, 10, 11

[3.5] a) explanatory variable: jet skis in use

b)



There is a strong linear relationship between the number of jet skis and the number of accidents.

[3.7] a) The variables are positively associated. As the number of jet skis increases the number of accidents also increases.

b) The association is linear.

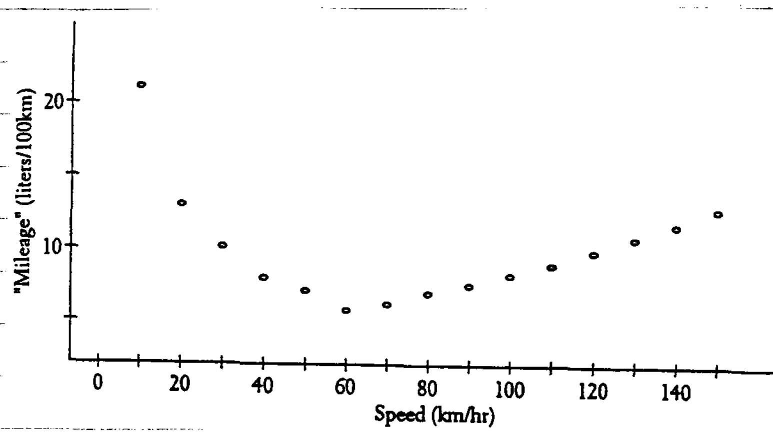
c) The association appears to be strong.

Using some points (0,0) and (600,000, 3002)

$m = \frac{3002}{600{,}000} \approx .005$ so $y \approx .005x$

If $x = 1{,}000{,}000$ $y = .005(1{,}000{,}000) = 5000$ accidents.

[3.8] a) graph ⟶



b) The relationship is curved — low in the middle, high at the ends. Mileage is at its best at moderate speeds — not at low or high speeds.

c) Above avg. mileage numbers are at both low and high speeds.

d) The relationship is strong — the points closely follow the curve pattern. We could use it for prediction.

3.10

a) Two mothers are 57" tall. Fathers are 66" and 67".

b) Tallest father is 74" tall. (3 of them). Mothers are 62, 64, 67.

c) We could choose either variable as explanatory.

d) A weak positive association says that there is some tendency for people of similar (relatively) heights to marry. (Short men with short women...) there is a great deal of scatter in the graph. This indicates a weak association.

3.11 a) lowest: 106 calories (150 mg. sodium)
highest: 195 calories (520 mg sodium)

b) There is a positive association between calories and sodium. High calorie hot dogs tend to have high sodium and low calorie hot dogs tend to have low sodium.

c) The point on the lower left seems to be an outlier. (even though it does fit the overall pattern.) Even without the outlier, the relationship seems to be moderately strong linear.

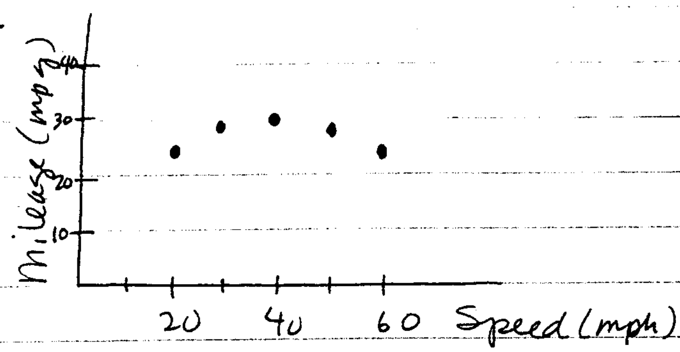
AP Stats 3.20, 21, 22, 24

3.20 There is no obvious linear relationship. The correlation is probably near zero.

3.21 The relationship between calories and sodium content in hot dogs is clearly positive but scattered. The correlation is positive but not near 1.

3.22 If women always married men 2 years older the correlation would be exactly equal to 1.

3.24



$\bar{x} = 40$

$S_x = 15.8$

$\bar{y} = 26.8$

$S_y = 2.68$

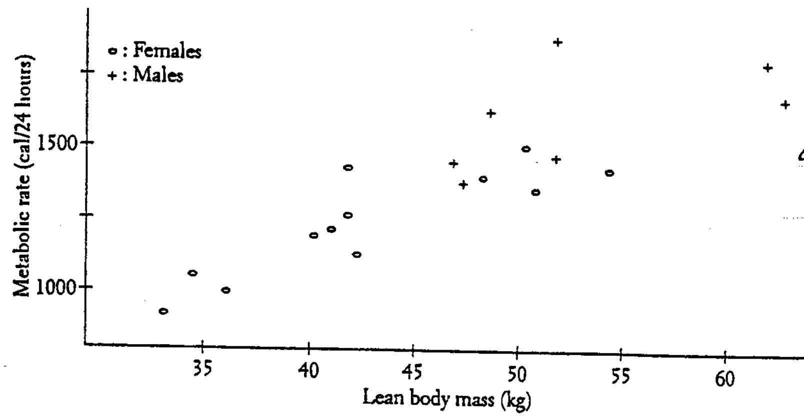
$$r = \frac{1}{5-1} \sum \left(\frac{x_i - 40}{15.8} \cdot \frac{y_i - 26.8}{2.68} \right) = \frac{1}{4} \cdot (0)(-6.146) = 0$$

x	$x-40$	$\frac{x-40}{15.8}$	y	$y-26.8$	$\frac{y-26.8}{2.68}$
20	-20	-1.266	24	-2.8	-4.118
30	-10	-.633	28	1.2	.448
40	0	0	30	3.2	1.194
50	10	.633	28	1.2	.448
60	20	+1.266	24	-2.8	-4.118
		$\Sigma = 0$			$\Sigma = -6.146$

Correlation measures strength of linear relationships. This association is obviously non-linear.

AP Stats 3.25, 28, 29, 30

3.25



a) Both correlations are positive, but the men are more spread out so it will be smaller.

b) women
$r = 0.8765$
men
$r = 0.5921$

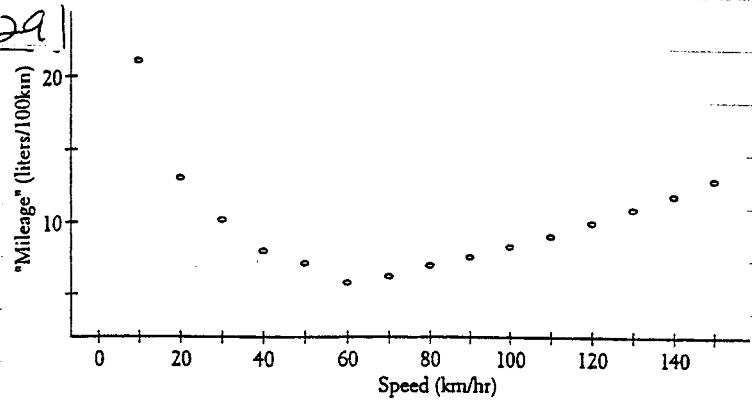
c)
$\bar{x}_w = 43.03$
$\bar{x}_m = 53.10$
The difference in means has no effect on the correlation.

d) No change - standard measures have no dimension.

3.28 The writer interpreted a correlation of zero to mean a negative correlation (close to -1).

The psychologist's finding was that there was little or no linear relationship between research and teaching. For example, knowing that a professor is a good reseacher doesnt tell you whether she is a good teacher.

3.29



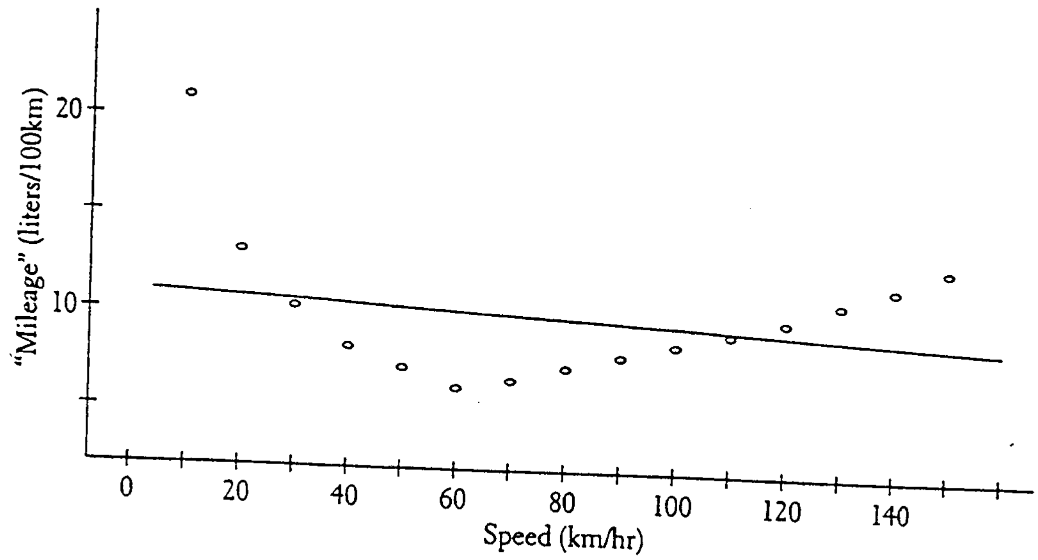
$r = -0.1716$
It is close to zero because the relationship is a (non-linear).

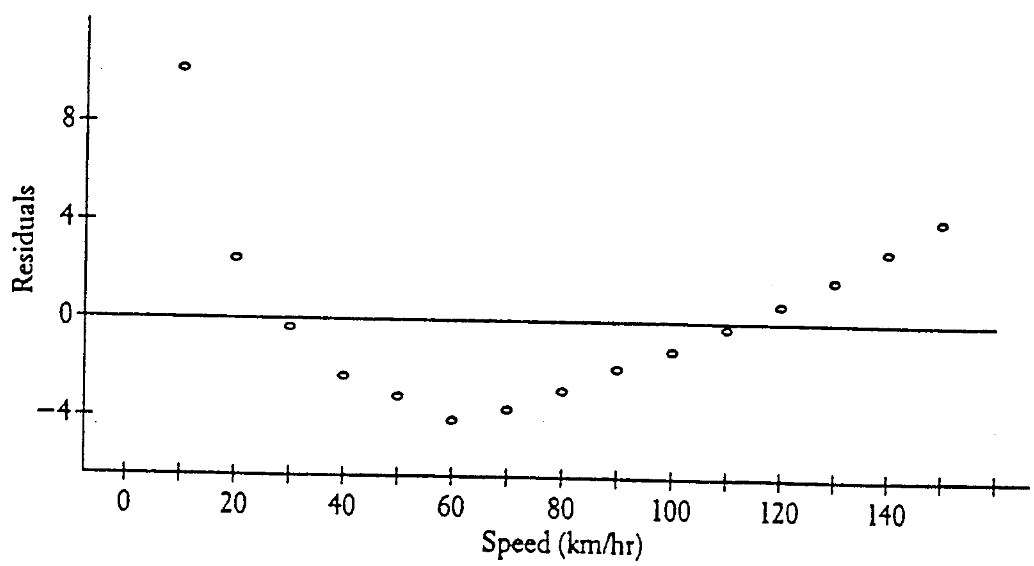
3.30 a) Sex is a categorical variable. We cannot calculate correlation without numbers!

b) The values for r must be between 1 and -1. r = 1.09 is impossible.

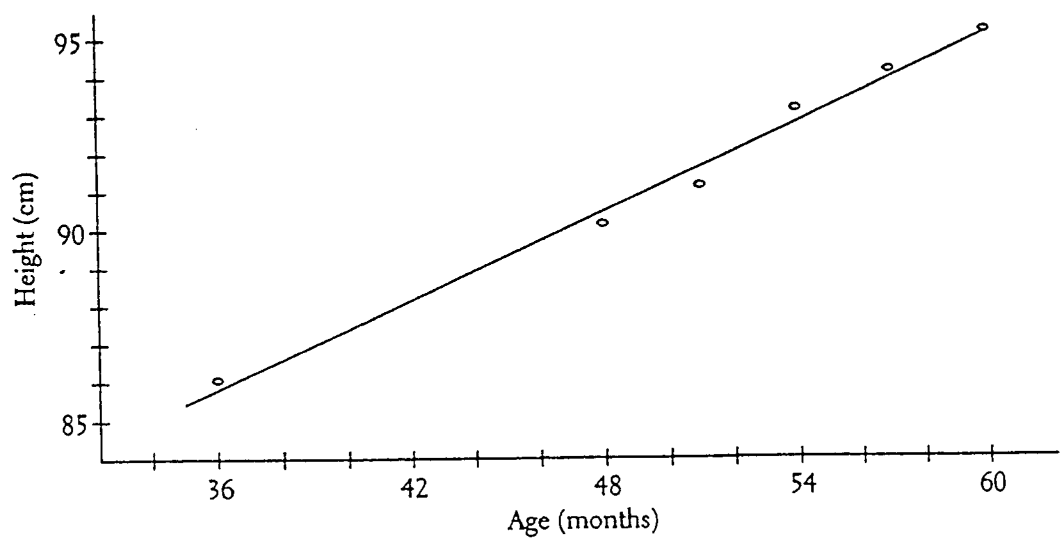
c) r comes from standardized values of x and y and, therefore, has no units.

3.39 (a) Below. (b) The line is clearly *not* a good predictor of the actual data—it is too high in the middle and too low on each end. (c) The sum is -0.01—a reasonable discrepancy allowing for round-off error. (d) A straight line is not the appropriate model for these data.

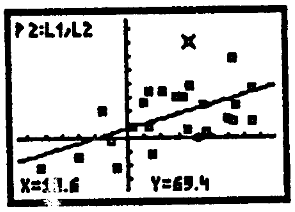




3.45 (a) Below. (b) $\hat{y} = 71.950 + 0.38333x$. (c) When $x = 40$, $\hat{y} = 87.2832$; when $x = 60$, $\hat{y} = 94.9498$. (d) Sarah is growing at about 0.38 cm/month; she should be growing about 0.5 cm each month ($0.5 = \frac{6}{60-48}$).

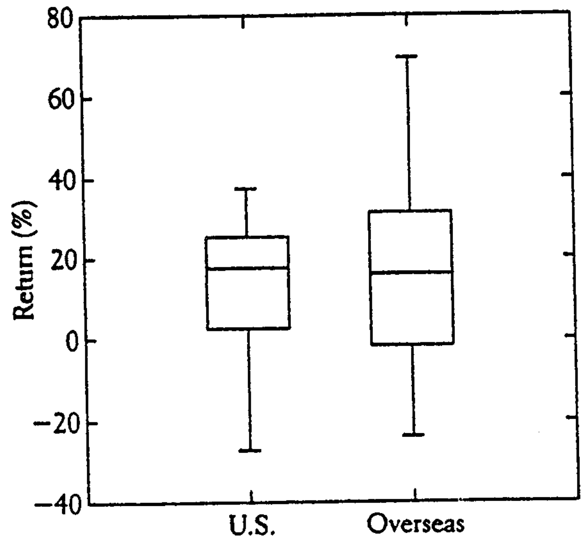


3.46 (a) See plot below. (b) $r = 0.5215$, and $r^2 = 0.2720$. There is a moderate positive association between U.S. and overseas stock returns; this relationship allows us to explain about 27% of the variation in one quantity with the other. (c) The regression line is $\hat{y} = 5.9546 + .7136x$. (d) $\hat{y} = 22.4\%$. We can only explain about one-third of the variation in overseas returns with the U.S. return information, as evidenced by the wide scatter around the line, so we should not expect too much accuracy in our predictions. (e) The outlier point occurred in 1986. The two points on the left end of the graph — from 1973 and 1974 — are potentially influential, especially the far left point.

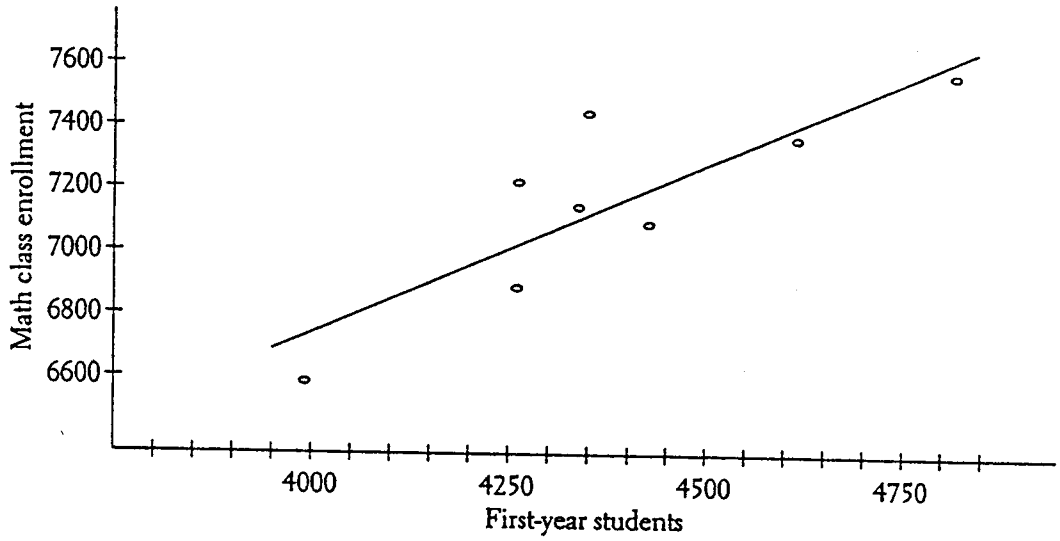


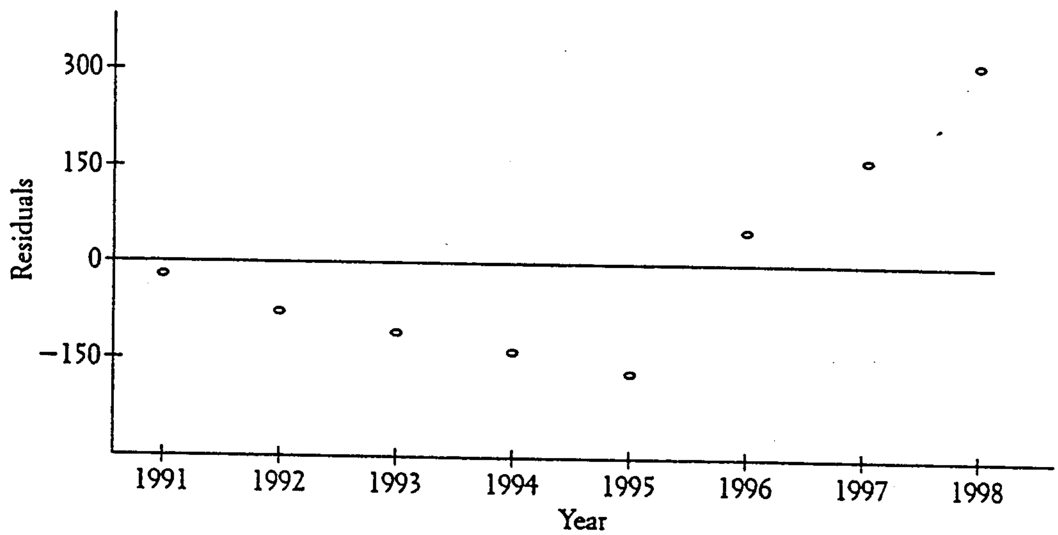
3.47 When $x = 480$, $\hat{y} = 255.95$ cm, or 100.77 in, or about 8.4 feet!

3.48 (a) U.S. stocks: −26.4 3.2 16.8 27 37.6 Overseas stocks: −23.2 .3 12.8 31.1 69.4. (b) Overseas stocks generally had higher returns — as the five-number summaries and the boxplots show, a quarter of the time they did better than 30%. (c) The overseas stocks also fluctuated much more wildly — ($Q_3 - Q_1$) is about 30% larger for overseas stocks, and the boxplot shows a lot more spread.



3.49 (a) About 69.4% of the variation is explained ($r^2 = 0.694$). (b) The sum is zero. (c) The residuals change from negative to positive in 1991 — the year the change was made. In that year, the regression line changes from overestimating to underestimating.

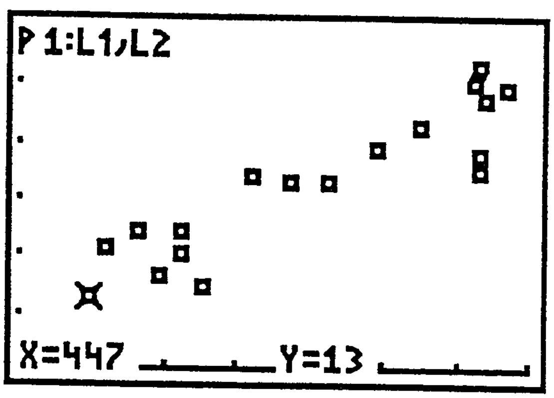




CHAPTER REVIEW

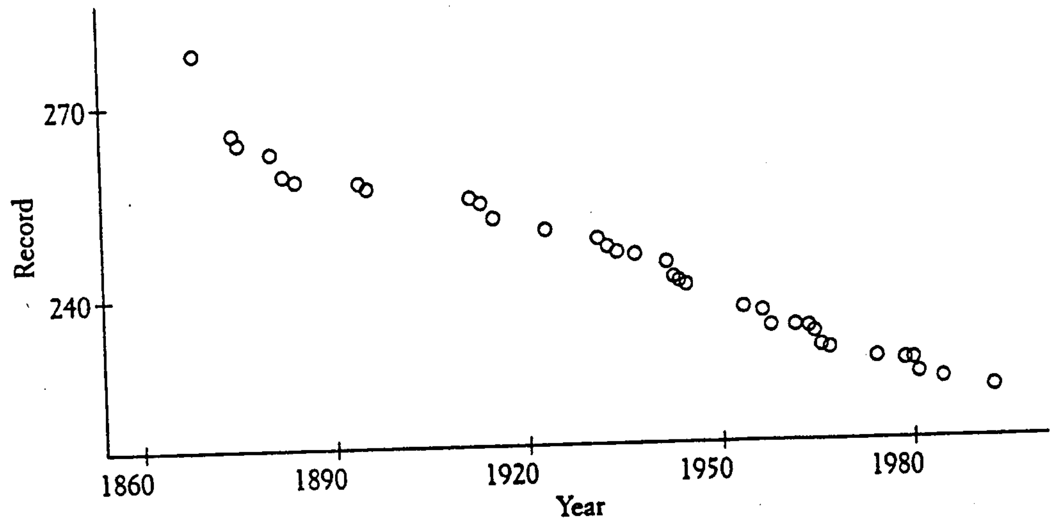
3.50 (a) Powerboat registrations.

(b)



(c) Yes, it appears that there is a strong straight-line pattern. The value of r^2 is 0.833, so 83.3% of the variation in manatees killed is explained by least-squares regression of powerboat registrations on manatees killed.

3.51



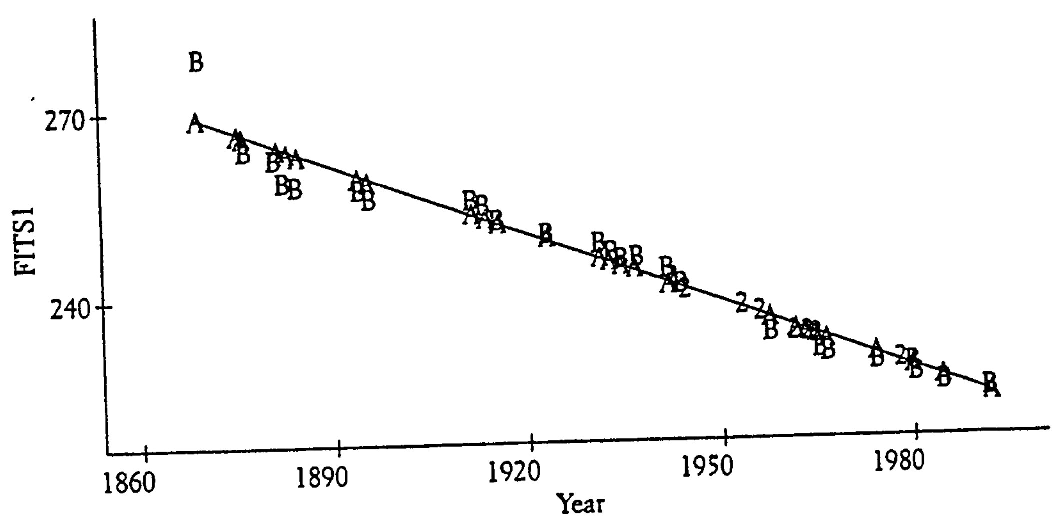
The variables are negatively associated; there is a downward trend. The mile record is decreasing over time. Here is a partial Minitab output:

```
Correlation of Year and Record = -0.983
The regression equation is     Record = 950 - 0.364 Year

Unusual Observations
Obs.    Year     Record        Fit  Stdev.Fit   Residual   St.Resid
  1     1868    278.800    269.266      0.920      9.534      4.03R
  6     1884    258.400    263.436      0.756     -5.036     -2.08R

R denotes an obs. with a large st. resid.
```

Here is the scatterplot with the regression line:



A = FITS1 vs. Year
B = Record vs. Year

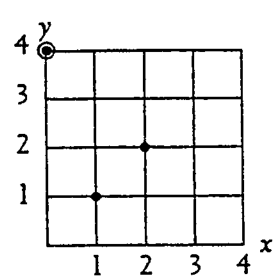
The regression line appears to be an excellent model for the data. The correlation is −0.983, and this indicates a strong negative association. Minitab identifies two regression outliers: a high outlier in 1868 and a low outlier in 1884. There appear to be no influential observations. On average, approximately (226.3−266.0)/(1985−1874) = −.358 or a little more than a third of a second was lopped off the record each year. It should be safe to predict the world record in 2000, but it may be risky to predict the record in 2005.

3.52 (a) 6. (b) 2. (c) 5. (d) 8. (e) 3. (f) 7. (g) 4. (h) 1.
If you want to duplicate these scatterplots, here are the data. Notice that in the first group of four data sets, the first 8 points are always the same. Then we look at the effect of a ninth point. The same is true for the second group of four data sets.

x1	y1	x2	y2	x3	y3	x4	y4
2	2	2	2	2	2	2	2
3	1	3	1	3	1	3	1
4	2	4	2	4	2	4	2
5	3	5	3	5	3	5	3
5	4	5	4	5	4	5	4
6	4	6	4	6	4	6	4
7	3	7	3	7	3	7	3
8	5	8	5	8	5	8	5
		5	10	15	6	15	1

x5	y5	x6	y6	x7	y7	x8	y8
1	5	1	5	1	5	1	5
2	3	2	3	2	3	2	3
3	4	3	4	3	4	3	4
4	4	4	4	4	4	4	4
4	3	4	3	4	3	4	3
5	2	5	2	5	2	5	2
6	1	6	1	6	1	6	1
7	2	7	2	7	2	7	2
		15	7	12	3	1	10

3.53 (a) Start with points (1, 1) and (2, 2). Then add the influential point (0, 4).



(b) Start with the set of points (1, 1), (1, 2), (2, 1.1), and (2, 2). Then add the influential point (10, 10).