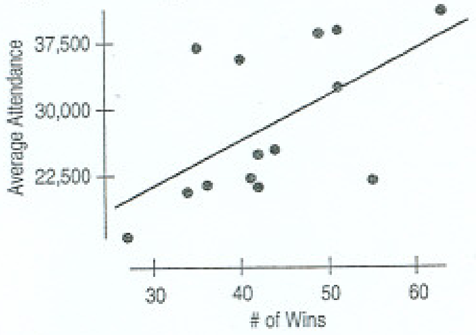
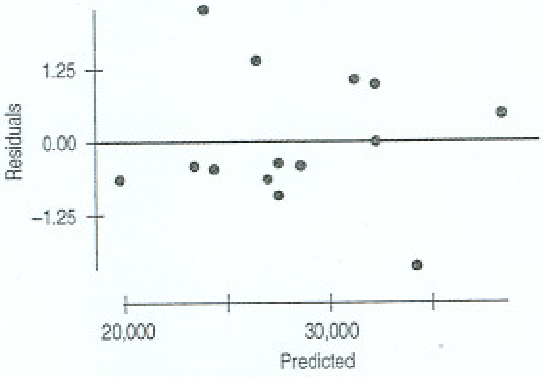
16. Baseball. An exercise in the last chapter looked at the relationship between the number of *games won* by American League baseball teams and the *average attendance* at their home games for the first half of the 2001 season. Here are the scatterplot, the residuals plot, and part of the regression analysis.





Dependent variable is: Attendance
R^2 = 33.3%

Variable	Coefficient
Intercept	5773.27
Wins	517.609

a) Do you think a linear model is appropriate here? Explain.
b) Interpret the meaning of R^2 in this context.

a) Yes, a linear model is appropriate because the residual plot shows no obvious pattern.

b) The r^2 value of 33.3% demonstrates that 33.3% of the variation in the average attendance can be explained by the variation in the number of wins.

18. Second inning. Consider again the regression of *average attendance* on *wins* for the baseball teams examined in Exercise 16.

a) What is the correlation between *wins* and *average attendance?*
b) What would you predict about the *average attendance* for a team that is 2 standard deviations above average in games won?
c) If a team is 1 standard deviation below average in attendance, what would you predict about the number of games the team has won?

a) $r = \sqrt{r^2} = \sqrt{.333} = .58$
Moderately strong positive correlation

b) The S_x and S_y are found in the slope with the r.

$$r\frac{S_y}{S_x} = \frac{r \cdot S_y}{S_x}$$

If you think of it this way, then r is just a scale factor. If it says 2 SD's above in games won we have:

$\frac{r\,S_y}{2S_x}$ and the S_y would have to go up

$.58 \bullet 2 = 1.16$ SD's

c) Similarly we have

$\frac{r(-1\,S_y)}{S_x}$ → divide this time:

$\frac{1}{.58} = 1.74$ SD's below

20. Last inning. Refer again to the regression analysis for average attendance and games won by American League baseball teams, seen in Exercise 16.

a) Write the equation of the regression line.

b) Estimate the *average attendance* for a team with 50 *wins.*

c) Interpret the meaning of the slope of the regression line in this context.

d) In general, what would a negative residual mean in this context?

e) The St. Louis Cardinals are not included in these data because they are a National League team. At the time these data were collected, the Cardinals had won 43 games and averaged 38,988 fans at their home games. Calculate the residual for this team, and explain what it means. (Yes, we're taking a risk applying this equation to the other league.)

a) $\hat{y} = 5773.27 + 517.609x$

in context:

$\widehat{\text{Attendance}} = 5773.27 + 517.609(\text{wins})$

b) $x = 50$ wins

$\hat{y} = 5773.27 + 517.609(50)$

$\boxed{\hat{y} = 31{,}653.72 \text{ people}}$

c) $b_1 = 517.609$

For every additional win, there is an increase of approximately 518 people in attendance.

d) A negative residuals implies that despite the number of wins, the average attendance fell short of the predicted number of people.

d) To calculate the residual (the error) we need $y - \hat{y}$ or the actual minus the predicted amount.

Actual = 38,988 } At $x = 43$ wins

Predicted = ?

Find the predicted # in attendance by plugging in the wins to the LSRL equation.

$\hat{y} = 5773.27 + 517.609(43)$

$\hat{y} = 28030.457$

Now $y - \hat{y} =$

$38{,}988 - 28{,}030.457 =$

$\boxed{10{,}957.543}$

This means the actual # of fans in attendance exceeded the predicted amount when the Cardinals had 43 wins.