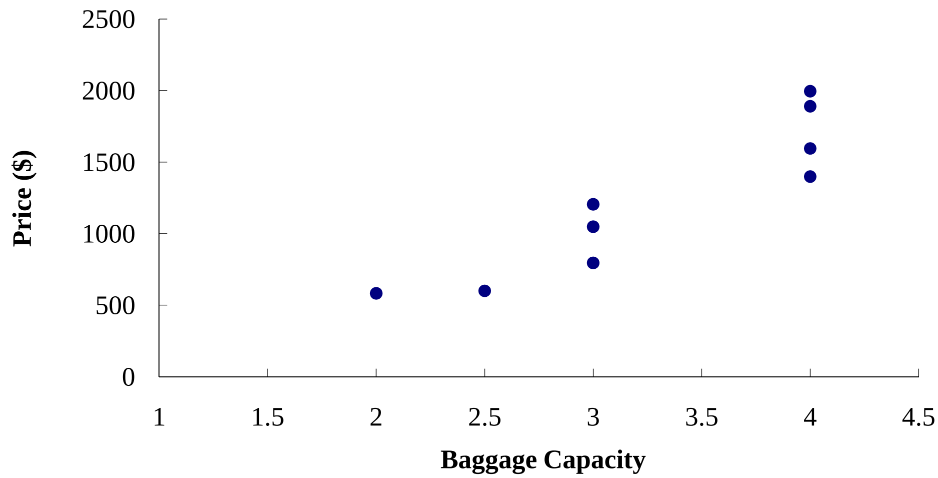
STAT 244BU   Hw 6 Answer key

5. a.



b. Let *x* = baggage capacity and *y* = price ($).

There appears to be a positive linear relationship between *x* and *y*.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between *x* and *y*; in part (d) we will determine the equation of a straight line that "best" represents the relationship according to the least squares criterion.

d. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{29.5}{9} = 3.277778 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{11,110}{9} = 123.444444$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 2909.888891 \quad \Sigma(x_i - \bar{x})^2 = 4.555559$

$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2909.888891}{4.555559} = 638.755615$

$b_0 = \bar{y} - b_1\bar{x} = 1234.4444 - (638.7561)(3.2778) = -859.254658$

$\hat{y} = -859.26 + 638.76x$

e. A one point increase in the baggage capacity rating will increase the price by approximately $639.

f. $\hat{y} = -859.26 + 638.76x = -859.26 + 638.76(3) = \$1057$

19. The estimated regression equation and the mean for the dependent variable are:

$\hat{y}_i = 40{,}639 - 1301.2x \qquad \bar{y} = 36{,}562.27$

The sum of squares due to error and the total sum of squares are

$\text{SSE} = \sum(y_i - \hat{y}_i)^2 = 47{,}116{,}828 \qquad \text{SST} = \sum(y_i - \bar{y})^2 = 94{,}072{,}519$

Thus, SSR = SST - SSE = 94,072,519 – 47,116,828 = 46,955,691

r^2 = SSR/SST = 46,955,691/94,072,519 = .4991

We see that 49.91% of the variability in *y* has been explained by the least squares line.

$r_{xy} = \sqrt{.4991} = -.71$

27. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{37}{10} = 3.7 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{1654}{10} = 165.4$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 315.2 \qquad \Sigma(x_i - \bar{x})^2 = 10.1$

$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{315.2}{10.1} = 31.2079$

$b_0 = \bar{y} - b_1\bar{x} = 165.4 - (31.2079)(3.7) = 49.9308$

$\hat{y} = 49.9308 + 31.2079x$

b. SSE = $\Sigma(y_i - \hat{y}_i)^2 = 2487.66$ SST = $\Sigma(y_i - \bar{y})^2$ = 12,324.4

Thus, SSR = SST - SSE = 12,324.4 - 2487.66 = 9836.74

MSR = SSR/1 = 9836.74

MSE = SSE/(*n* - 2) = 2487.66/8 = 310.96

F = MSR / MSE = 9836.74/310.96 = 31.63

Using *F* table (1 degree of freedom numerator and 8 denominator), *p*-value is less than .01

Using Excel or Minitab, the *p*-value corresponding to $F = 31.63$ is .001.

Because *p*-value $\leq \alpha$, we reject H_0: $_1 = 0$

Upper support and price are related.

c. r^2 = SSR/SST = 9,836.74/12,324.4 = .80

The estimated regression equation provided a good fit; we should feel comfortable using the estimated regression equation to estimate the price given the upper support rating.

d. $\hat{y}$ = 49.93 + 31.21(4) = 174.77

39. a. Let x = miles of track and y = weekday ridership in thousands.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{203}{7} = 29 \qquad \bar{y} = \frac{\Sigma y_i}{n} = \frac{309}{7} = 44.1429$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 1471 \qquad \Sigma(x_i - \bar{x})^2 = 838$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{1471}{838} = 1.7554$$

$$b_0 = \bar{y} - b_1\bar{x} = 44.1429 - (1.7554)(29) = -6.76$$

$$\hat{y} = -6.76 + 1.755x$$

b. SST =3620.9   SSE = 1038.7   SSR = 2582.1

r^2 = SSR/SST = 2582.1/3620.9 = .713

The estimated regression equation explained 71.3% of the variability in *y*; a good fit.

c. s^2 = MSE = 1038.7/5 = 207.7

$$s = \sqrt{207.7} = 14.41$$

$$s_{\hat{y}_p} = s\sqrt{\frac{1}{n}+\frac{(x_p-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}} = 14.41\sqrt{\frac{1}{7}+\frac{(30-29)^2}{838}} = 5.47$$

$$\hat{y} = -6.76 + 1.755x = -6.76 + 1.755(30) = 45.9$$

$$\hat{y}_p \pm t_{\alpha/2} s_{\hat{y}_p}$$

45.9 2.571(5.47) = 45.9 14.1

or 31.8 to 60

d. $$s_{\text{ind}} = s\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}} = 14.41\sqrt{1+\frac{1}{7}+\frac{(30-29)^2}{838}} = 15.41$$

$$\hat{y}_p \pm t_{\alpha/2} s_{\text{ind}}$$

45.9 2.571(15.41) = 45.9 39.6

or 6.3 to 85.5

The prediction interval is so wide that it would not be of much value in the planning process. A larger data set would be beneficial.

43. a. The Minitab output is shown below:

The regression equation is
Price = 4.98 + 2.94 Weight

Predictor	Coef	SE Coef	T	P
Constant	4.979	3.380	1.47	0.154
Weight	2.9370	0.2934	10.01	0.000

S = 8.457 R-Sq = 80.7% R-Sq(adj) = 79.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7167.9	7167.9	100.22	0.000
Residual Error	24	1716.6	71.5		
Total	25	8884.5			

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	34.35	1.66	(30.93, 37.77)	(16.56, 52.14)

b. The p-value = .000 < α = .05 (t or F); significant relationship

c. r^2 = .807. The least squares line provided a very good fit.

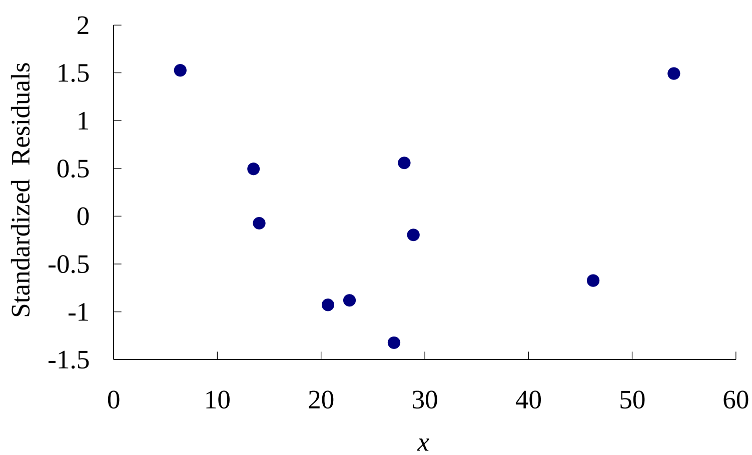
d. The 95% confidence interval is 30.93 to 37.77.

e. The 95% prediction interval is 16.56 to 52.14.

49. a. Let x = return on investment (ROE) and y = price/earnings (P/E) ratio.

$$\hat{y} = -32.13 + 3.22x$$

b.



c. There is an unusual trend in the residuals. The assumptions concerning the error term appear questionable.

53. a. The Minitab output is shown below:

The regression equation is
Price = 28.0 + 0.173 Volume

Predictor	Coef	SE Coef	T	P
Constant	27.958	4.521	6.18	0.000
Volume	0.17289	0.07804	2.22	0.036

S = 17.53 R-Sq = 17.0% R-Sq(adj) = 13.5%

Analysis of Variance

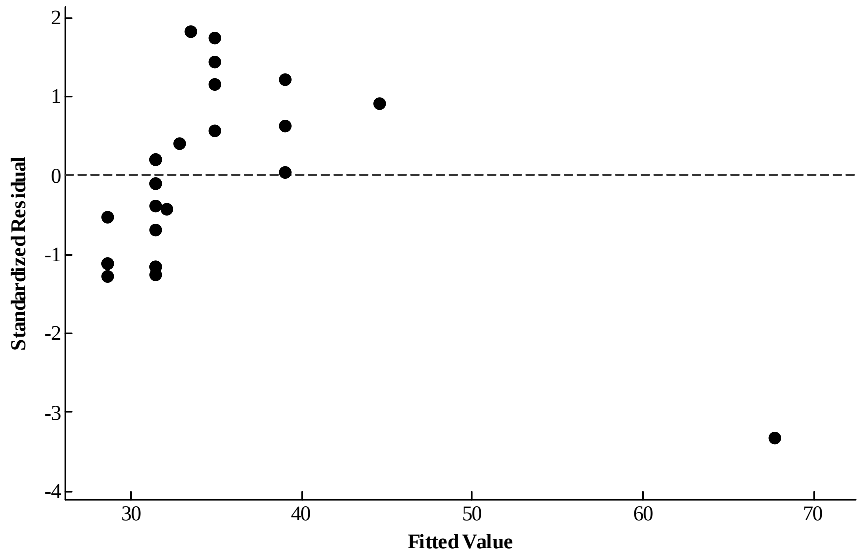
Source	DF	SS	MS	F	P
Regression	1	1508.4	1508.4	4.91	0.036
Residual Error	24	7376.1	307.3		
Total	25	8884.5			

Unusual Observations

Obs	Volume	Price	Fit	SE Fit	Residual	St Resid
22	230	40.00	67.72	15.40	-27.72	-3.31RX

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

b. The Minitab output identifies observation 22 as having a large standardized residual and is an observation whose *x* value gives it large influence. The following residual plot verifies these observations.


Standardized Residual
2
1
0
-1
-2
-3
-4
30
40
50
60
70
Fitted Value