where π_j is probability of a rat in Group j having One or More Tumors.
$\chi^2 = 3.312$ with df = (2-1)(3-1) = 2 and p-value = 0.191.

c. No

10.81 a. The results are summarized in the following table:

Question	$\hat{\pi}$	$\hat{\sigma}_{\hat{\pi}}$	95% C.I.
Did Not Explain?	0.254	0.01947	(0.216, 0.292)
Might Bother?	0.916	0.0124	(0.892, 0.940)
Did Not Ask?	0.471	0.02232	(0.427, 0.515)
Drug Not Changed?	0.877	0.0147	(0.848, 0.906)

10.85 $\bar{y} = \sum_i (No.Mites)(frequency)/500 = 1.146$
we obtain the following using a Poisson distribution with $\mu = 1.2$:

Mites/Leaf (k_i)	0	1	2	3	4	≥ 5	Total
$\pi_i = P(y = k_i)$	0.3012	0.3614	0.2169	0.0867	0.0260	0.0078	1.0
$E_i = 500\pi_i$	150.6	180.7	108.5	43.3	13	3.9	500
n_i	233	127	57	33	30	20	500

$\chi^2 = 176.6$ with df =6-1 = 5, $\Rightarrow p - value = Pr(\chi^2 > 176.6) < 0.001 \Rightarrow$
Reject H_o

Chapter 11: Linear Regression and Correlation

11.3 The calculations are give here:

i	x_i	y_i	$(x_i - 3)^2$	$(x_i - 3)(y_i - 5.6)$
1	1	2	4	7.2
2	2	4	1	1.6
3	3	6	0	0
4	4	7	1	1.4
5	5	9	4	6.8
Total	15	28	10	17

$\bar{x} = 15/5 = 3$ $\qquad$ $\bar{y} = 28/5 = 5.6$

$S_{xx} = \sum_i (x_i - 3)^2 = 10$

$S_{xy} = \sum_i (x_i - 3)(y_i - 5.6) = 17$

$\hat{\beta}_1 = S_{xy}/S_{xx} = 17/10 = 1.7$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 5.6 - (1.7)(3) = 0.5$

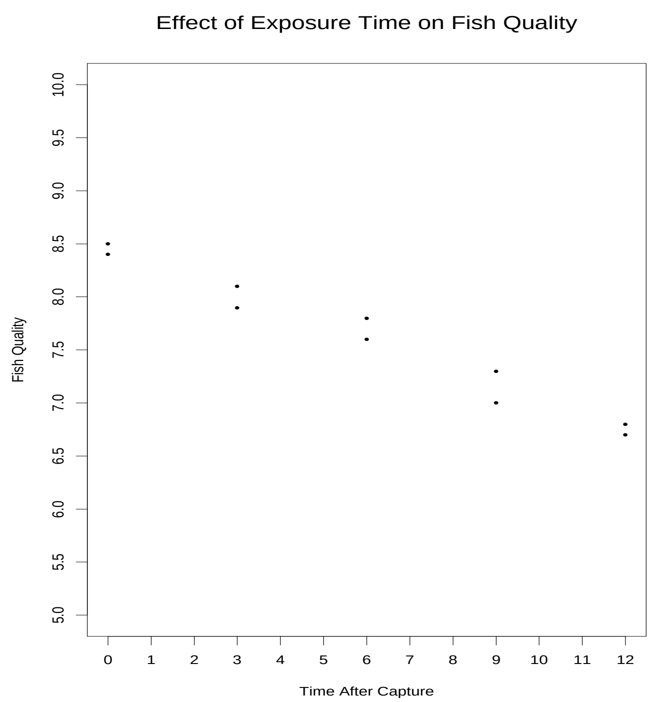
$\hat{y} = 0.5 + 1.7x$

11.5 $\hat{\beta}_1 = 1.633$

$\hat{\beta}_0 = 2.468$

$\hat{y} = 2.468 + 1.633x$

11.9 a. A scatterplot of the data is given here:



b. Minitab output is given here:

```
Regression Analysis: QUALITY versus STORAGE


The regression equation is
QUALITY = 8.46 - 0.142 STORAGE

Predictor        Coef     SE Coef          T        P
Constant      8.46000     0.06610     127.99    0.000
STORAGE     -0.141667    0.008995     -15.75    0.000

S = 0.1207      R-Sq = 96.9%     R-Sq(adj) = 96.5%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      3.6125      3.6125    248.07     0.000
Residual Error     8      0.1165      0.0146
```

```
Total              9      3.7290
```

The least squares estimates are $\hat{\beta}_0 = 8.46$ and $\hat{\beta}_1 = -0.142$

c. The estimated slope value $\hat{\beta}_1 = -0.142$ indicates that for each 1 hour increase in the time between the capture of the fish and their placement in storage there is approximately a 0.142 decrease in the average quality of the fish.

11.10 $\hat{y} = 8.46 - 0.142x \quad \Rightarrow \quad \hat{y} = 8.46 - 0.142(10) = 7.04$

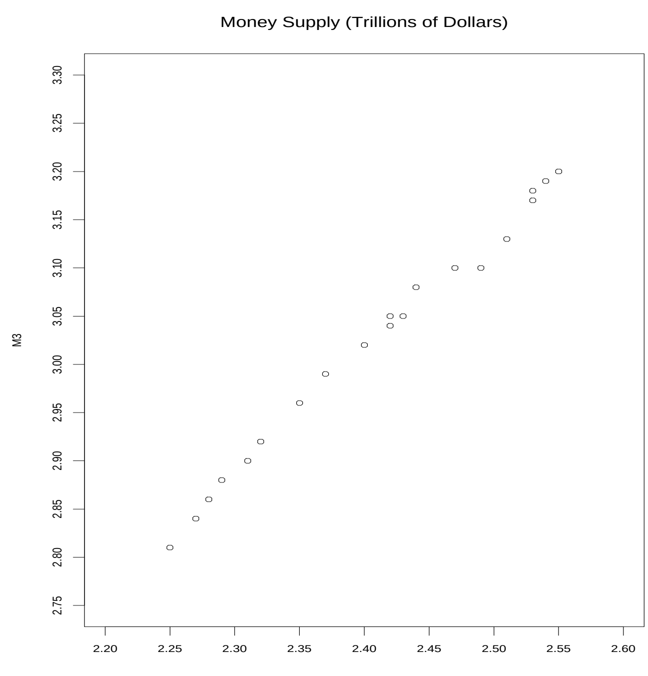
It would not be a good idea to make predictions at $x = 18$

11.15 logarithmic transformation is suggested from the plot.

11.16 a. For the transformed data, the plotted points appear to be reasonably linear.

b. The least squares line is $\hat{y} = 3.10 + 2.76\sqrt{x}$
(Using rounded values $3.097869 \approx 3.10$ and $2.7633138 \approx 2.76$).

11.21 a. A scatterplot of the data is given here:



b. The lifetime of one of the bits at a speed 100 is quite a bit smaller than the other three lifetimes at this speed.

11.22 a. The estimated intercept is 6.03 and the estimated slope is -0.017.

b. The slope has a negative sign which indicates a decreasing relation between lifetime and speed.

c. 0.6324.

11.23 a. The prediction equation is $\hat{y} = 6.03 - 0.017x$. Thus, we just replace x with the speeds of 60, 80, 100, 120, and 140 to obtain the values:

$$x = 60 \Rightarrow \hat{y} = 5.01$$

$$x = 80 \Rightarrow \hat{y} = 4.67$$

$$x = 100 \Rightarrow \hat{y} = 4.33$$

$$x = 120 \Rightarrow \hat{y} = 3.99$$

$$x = 140 \Rightarrow \hat{y} = 3.65$$

11.29 Minitab output is given here:

```
Regression Analysis: FIRMNESS versus CONCENTRATION


The regression equation is
FIRMNESS = 48.9 + 10.3 CONCENTRATION

Predictor        Coef     SE Coef          T        P
Constant       48.933       1.541      31.76    0.000
CONCENTR      10.3333      0.7957      12.99    0.000

S = 2.387       R-Sq = 97.7%     R-Sq(adj) = 97.1%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      961.00      961.00    168.65     0.000
Residual Error     4       22.79        5.70
Total              5      983.79
```

a. $\hat{y} = 48.9 + 10.3x$

b. $S^2 = 5.70$

c. $SE(\hat{\beta}_1) = 0.7957$

11.30 $H_o : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$

Test Statistic: $t = \frac{\hat{\beta}_1}{s_e/S_{xx}} = \frac{10.33}{2.387/9} = 12.99$

$p - value = 2P(t_4 > 12.99) < 0.0001 \Rightarrow$ Reject H_o

11.35 a. Yes

b. $\hat{y} = 12.51 + 35.83x$

11.36 a. 1.069

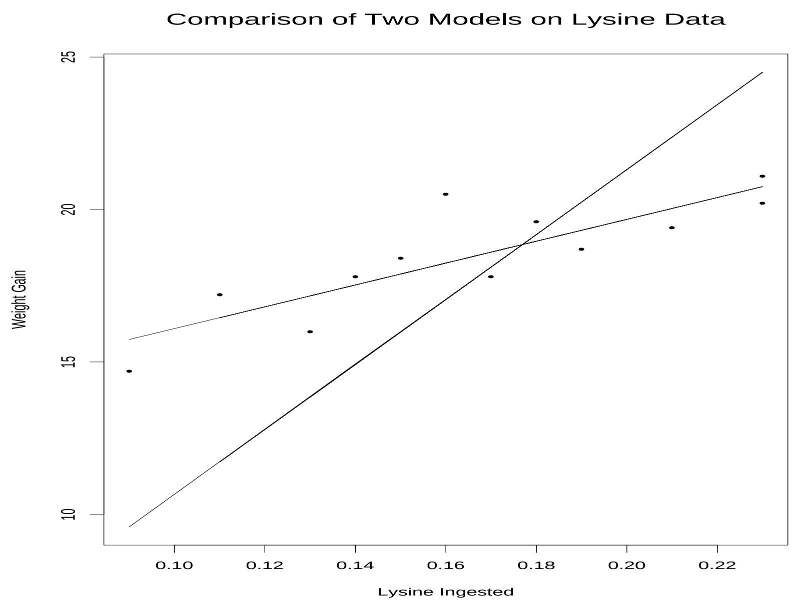
b. 6.957

c. $H_o : \beta_1 \leq 0$ versus $H_a : \beta_1 > 0$
Test Statistic: $t = 5.15$
$p - value = P(t_{10} > 5.15) = 0.0002 \Rightarrow$ Reject H_o and conclude there is significant evidence that there is a positive linear relationship.

11.37 a. No

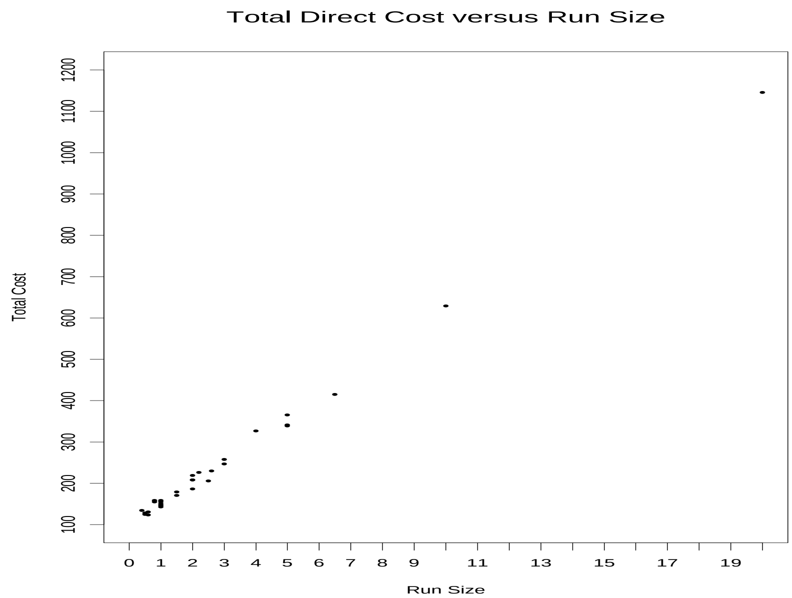
11.38 a. $\hat{\beta}_1 = 106.52$

b. Scatterplot of the data and the two fitted lines are given here:



The model with an intercept term produced a better fit to the data values.

11.39 a. Scatterplot of the data is given here:



b. $\hat{y} = 99.777 + 5.1918x$

The residual standard deviation is $s = 12.2065$

c. A 95% C.I. for the slope is given by $(5.072, 5.312)$

11.43 95% prediction interval for log biological recovery percentage at x=30 is given by (0.941, 1.449)

11.44 a. $\hat{y} = -1.733333 + 1.316667x$

b. The p-value for testing $H_o : \beta_1 \leq 0$ versus $H_a : \beta_1 > 0$ is

$p - value = P(t_{10} \geq 6.342) < 0.0005 \Rightarrow$ Reject H_o

11.45 a. 95% Confidence Intervals for $E(y)$ at selected values for x:

$x = 4 \Rightarrow (2.6679, 4.3987)$

$x = 5 \Rightarrow (4.2835, 5.4165)$

$x = 6 \Rightarrow (5.6001, 6.7332)$

$x = 7 \Rightarrow (6.6179, 8.3487)$

b. 95% Prediction Intervals for y at selected values for x :

$x = 4 \Rightarrow (1.5437, 5.5229)$

$x = 5 \Rightarrow (2.9710, 6.7290)$

$x = 6 \Rightarrow (4.2877, 8.0456)$

$x = 7 \Rightarrow (5.4937, 9.4729)$

11.46 a. $\hat{y} = 99.77704 + 51.9179x \Rightarrow$ When x=2.0, $\hat{E}(y) = 203.613$

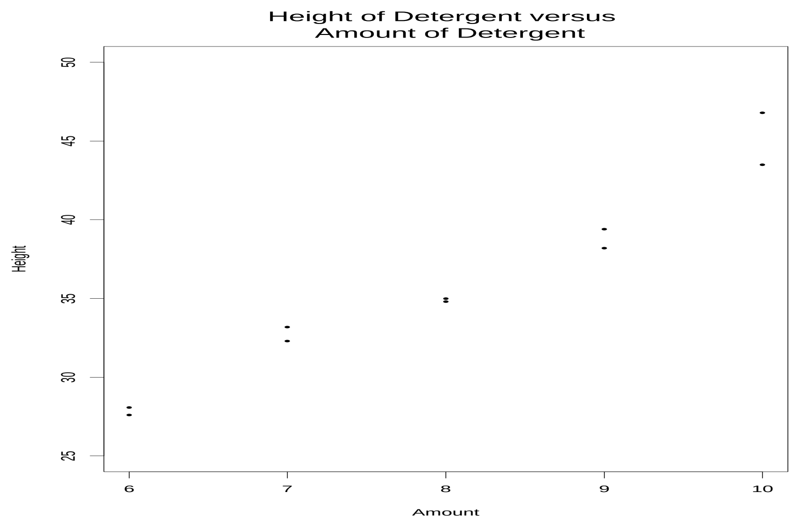
b. The 95% C.I. is given in the output as $(198.902, 208.323)$

11.47 No

11.48 a. The 95% P.I. is given in the output as $(178.169, 229.057)$

b. Yes

11.53 a. Scatterplot of the data is given here:

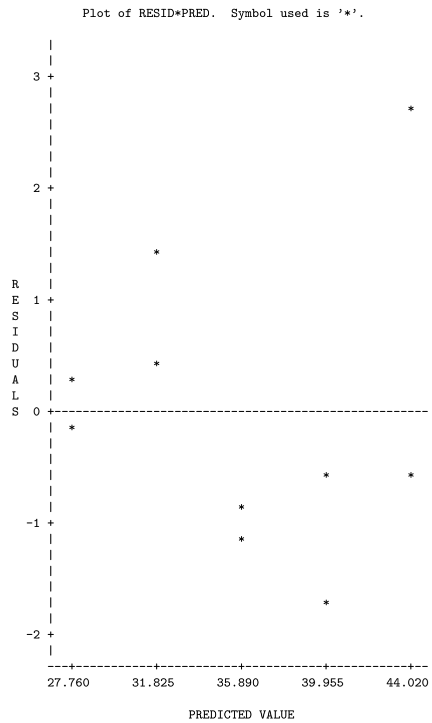


b. The SAS output is given here:

```
Model: MODEL1
Dependent Variable: Y

                           Analysis of Variance

                          Sum of         Mean
Source          DF       Squares        Square      F Value       Prob>F

Model            1     330.48450     330.48450      169.213       0.0001
Error            8      15.62450       1.95306
C Total          9     346.10900

    Root MSE       1.39752     R-square       0.9549
    Dep Mean      35.89000     Adj R-sq       0.9492
    C.V.           3.89390

                           Parameter Estimates

                   Parameter       Standard    T for H0:
 Variable  DF       Estimate          Error   Parameter=0    Prob > |T|

 INTERCEP   1       3.370000     2.53872138         1.327        0.2210
 X          1       4.065000     0.31249500        13.008        0.0001
```

$\hat{y} = 3.37 + 4.065x$

c. The residual plot is given here:



The residual plot indicates that higher order terms in x may be needed in the model.

11.54 a. A test of lack of fit is given here:

$SSP_{exp} = 6.715$

From the output of exercise 11.53, SS(Residuals) = 15.6245. Thus,

$SS_{Lack} = 15.6245 - 6.715$

$df_{Lack} = 3$

$df_{exp} = 5$

$F = 2.21 < 5.41 = F_{.05,3,5}$

There is not sufficient evidence that the linear model is inadequate.

b. $S_{xx} = \sum_i (x_i - \bar{x}) = 20$. Prediction interval for y at x =6,7,8,9,10

x	$\hat{y}$	95% P.I.
6	27.760	(24.086, 31.434)
7	31.825	(28.369, 35.281)
8	35.890	(32.510, 39.270)
9	39.955	(36.499, 43.411)
10	44.020	(40.346, 47.694)

11.57 a. $\hat{y} = 0.113 + 0.118x$

Dependent variable is Transformed CUMVOL and Independent variable is Log(Dose).

b. $\hat{x} = (y - 0.11277)/0.11847$

$\hat{x}_L = 2.40 + \frac{1}{(1-.2426)}(\hat{x} - 2.40 - d)$

$\hat{x}_U = 2.40 + \frac{1}{(1-.2426)}(\hat{x} - 2.40 + d) \Rightarrow$

y	TRANS(y)	$\hat{LOG}(x)$	$\hat{x}$	d	$\hat{LOG}(x)_L$	$\hat{LOG}(x)_U$	$\hat{x}_L$	$\hat{x}_U$
10	.322	1.764	5.84	.242289	1.24038	1.88018	3.46	6.55
14	.383	2.285	9.83	.198634	1.98616	2.51068	7.29	12.31
19	.451	2.855	17.38	.221359	2.70876	3.29328	15.01	26.93

11.58 The four values of CUMVOL are 10, 20, 30, 12 yielding $\bar{y} = 0.42969$ and $s_{\bar{y}} = .05849$, where y is the arcsine of the square root of CUMVOL.

For 50%: $\hat{x} = 2.367$ and 95% confidence limits are (0.79, 4.45)

For 75%: $\hat{x} = 5.861$ and 95% confidence limits are (2.20, 12.88)

11.59 $r = 0.9722$.

11.60 a. Test $H_o : \rho_{yx} \leq 0$ versus $H_a : \rho_{yx} > 0$

Test Statistic: $t13.13$

$p - value < 0.0005$

Conclusion: There is significant evidence that the correlation is positive.

11.61 a. $r^2_{yx} = 99.64\%$ (R-squared on output).

b. r_{yx} should be positive.

11.63 a. There appears to be a general increase in salary as the level of experience increases.

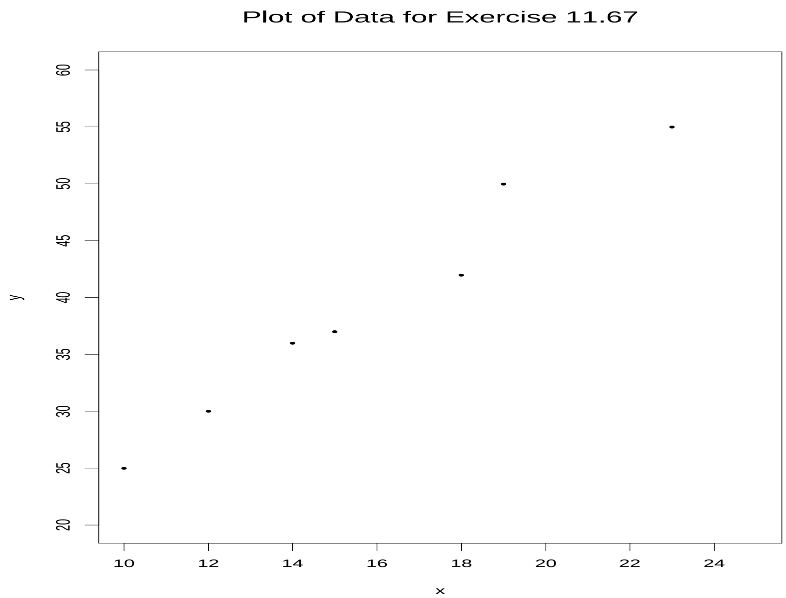
11.65 a. $\hat{y} = 40.507 + 1.470x$

b. The residual standard deviation is $s_\epsilon = 5.402$.

c. From the output $t = 6.916$ with a p-value of 0.000. This would imply overwhelming evidence that there is a relation between starting salary and experience.

d. $R^2 = 0.494$

11.67 a. Scatterplot of the data is given here:



b. $\hat{\beta}_1 = 2.3498 \Rightarrow$

$\beta_0 = 2.0247 \Rightarrow \hat{y} = 2.2047 + 2.3498x$

c. When $x_{n+1} = 21, \hat{y} = 51.55$

11.68 a. $s_\epsilon = 1.9583$

All the residuals fall within $0 \pm 2s_\epsilon = (-3.9166, 3.9166)$

11.77 a. $t = 2.14$, with a p-value of 0.0396.

11.79 a. $\hat{y} = 9.7709 + 0.0501x$ and $s_\epsilon = 2.2021$

b. 90% C.I. on $(0.0486, 0.0516)$

11.81 There is an error in the Excel output given in the book. Output from Minitab is given here:

```
Regression Analysis: Rate: versus Mileage:


The regression equation is
Rate: = 9.79 + 0.0503 Mileage:

Predictor        Coef     SE Coef          T        P
Constant       9.7932      0.4391      22.30    0.000
Mileage:    0.0503229   0.0008230      61.14    0.000

S = 2.042       R-Sq = 98.8%     R-Sq(adj) = 98.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       15590       15590   3738.63    0.000
Residual Error    46         192           4
Total             47       15782


Predicted Values for New Observations

New Obs     Fit     SE Fit         95.0% CI             95.0% PI
1        26.903      0.298   (  26.303,  27.503)  (  22.749,  31.057)

Values of Predictors for New Observations

New Obs  Mileage:
1             340
```

$\hat{y} = 26.903$

The 95% P.I. when x=340 is (22.749, 31.057)

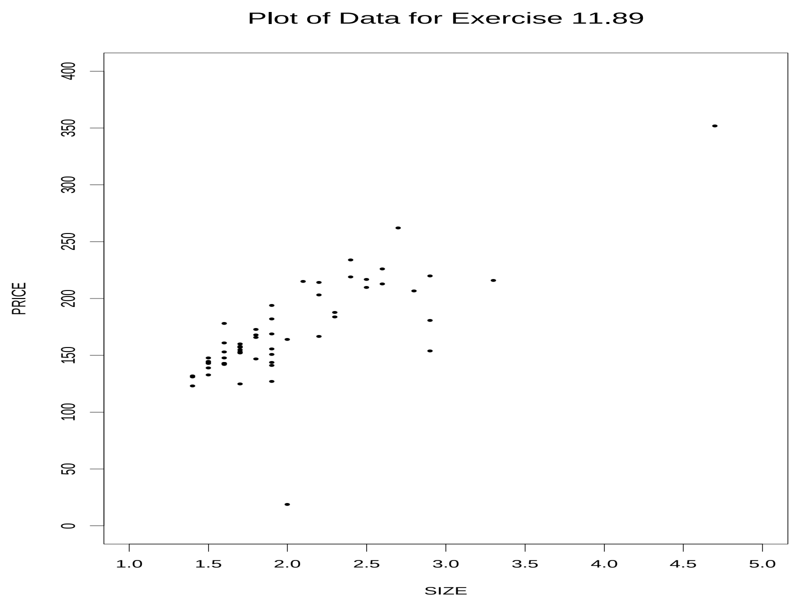
11.87 The Minitab output is given here:

```
Regression Analysis: Yield versus Nitrogen


The regression equation is
Yield = 11.8 + 6.50 Nitrogen

Predictor        Coef     SE Coef          T        P
Constant       11.778       1.887       6.24    0.000
Nitrogen       6.5000      0.8736       7.44    0.000

S = 2.140              R-Sq = 88.8%         R-Sq(adj) = 87.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      253.50      253.50     55.36    0.000
Residual Error     7       32.06        4.58
  Lack of Fit      1        2.72        2.72      0.56    0.484
  Pure Error       6       29.33        4.89
Total              8      285.56
```

$\hat{y} = 11.778 + 6.5x$

From the output, $MS_{Lack} = 2.72$, $MS_{exp} = 4.89 \Rightarrow F = \frac{2.72}{4.89} = 0.56$ with df= 1, 6. This yields $p - value = 0.484$ which implies there is no indication of lack of fit in the model.

11.89 a. Scatterplot of the data is given here:



c. The minitab output is given here:

```
Regression Analysis: PRICE versus SIZE


The regression equation is
PRICE = 51.1 + 59.2 SIZE

Predictor        Coef     SE Coef          T        P
Constant        51.08       13.97       3.66    0.001
SIZE           59.152       6.666       8.87    0.000

S = 29.14              R-Sq = 58.9%         R-Sq(adj) = 58.1%
PRESS = 50352.0        R-Sq(pred) = 55.66%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       66862       66862     78.73    0.000
Residual Error    55       46707         849
Total             56      113569

Unusual Observations
Obs       SIZE      PRICE         Fit      SE Fit    Residual    St Resid
  4       4.70     352.00      329.09       18.32       22.91        1.01 X
 27       2.00      19.10      169.38        3.86     -150.28       -5.20R
 47       2.90     154.00      222.62        7.06      -68.62       -2.43R

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

Durbin-Watson statistic = 2.10
```

The least squares line is $\hat{y} = 51.08 + 59.152x$

d. Minitab output for the data set without the outlier is given here:

```
Regression Analysis: PRICE versus SIZE


The regression equation is
PRICE = 54.0 + 59.0 SIZE

Predictor        Coef     SE Coef          T        P
Constant        53.99       10.06       5.37    0.000
SIZE           59.040       4.794      12.31    0.000

S = 20.96              R-Sq = 73.7%         R-Sq(adj) = 73.3%
PRESS = 26440.0        R-Sq(pred) = 70.73%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       66607       66607    151.64    0.000
Residual Error    54       23719         439
Total             55       90327

Unusual Observations
Obs       SIZE      PRICE         Fit      SE Fit    Residual    St Resid
  4       4.70     352.00      331.48       13.18       20.52        1.26 X
```

```
 13        2.90     181.00       225.20          5.09       -44.20         -2.17R
 26        2.70     262.00       213.40          4.32        48.60          2.37R
 46        2.90     154.00       225.20          5.09       -71.20         -3.50R

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

Durbin-Watson statistic = 1.83

Predicted Values for New Observations

New Obs     Fit     SE Fit         95.0% CI             95.0% PI
1       349.19      14.59   (  319.94,  378.43)  (  297.99,  400.38) XX
X  denotes a row with X values away from the center
XX denotes a row with very extreme X values

Values of Predictors for New Observations

New Obs      SIZE
1            5.00
```

$\hat{y} = 53.99 + 59.04x$.

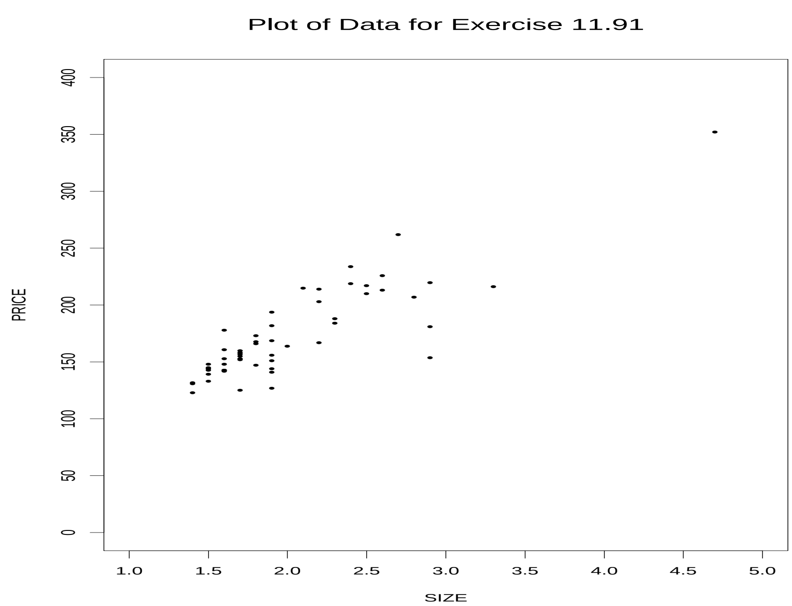
e. The residual standard deviations are $s_\epsilon = 29.14$ with outlier and $s_\epsilon = 20.96$ without outlier.

11.90 a. The estimated intercept is $\hat{\beta}_o = 53.99$.

c. Using the estimates from the Minitab output, a 95% C.I. for β_1 is $(49.428, 68.652)$.

11.91 a. From the Minitab output, a 95% P.I. for the selling price when the size is 5: $(297.99, 400.38)$.

b. Scatterplot of the data is given here:



c. The P.I. may not be valid since the statistical procedures depend on the condition that the variance remain constant across the range of x-values.

11.96 Minitab output is given here:

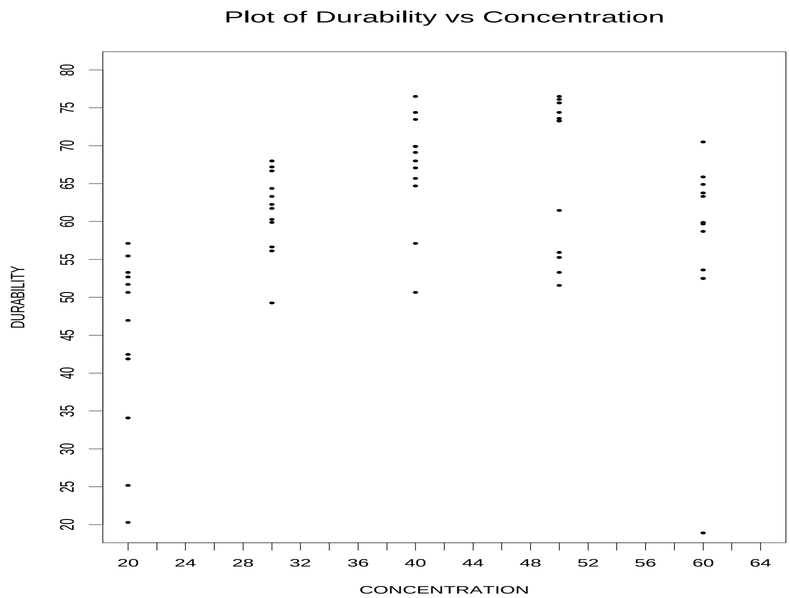
```
Regression Analysis: DURABIL versus CONCENTR


The regression equation is
DURABIL = 47.0 + 0.308 CONCENTR

Predictor        Coef     SE Coef          T        P
Constant       47.020       4.728       9.94    0.000
CONCENTR       0.3075      0.1114       2.76    0.008

S = 12.21              R-Sq = 11.6%         R-Sq(adj) = 10.1%
PRESS = 9368.62        R-Sq(pred) = 4.20%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      1134.7      1134.7      7.61    0.008
Residual Error    58      8644.5       149.0
Total             59      9779.2

Unusual Observations
Obs   CONCENTR    DURABIL         Fit      SE Fit    Residual    St Resid
  2       20.0      25.20       53.17        2.73      -27.97      -2.35R
  4       20.0      20.30       53.17        2.73      -32.87      -2.76R
 60       60.0      18.90       65.47        2.73      -46.57      -3.91R

R denotes an observation with a large standardized residual

Durbin-Watson statistic = 1.35
```

a. $\hat{y} = 47.020 + 0.3075x$. The estimated slope $\hat{\beta}_1 = 0.3075$ can be interpreted as follows: there is a 0.3075 increase in average durability for a 1 unit increase in the concentration.

b. The coefficient of determination, $R^2 = 11.6\%$.

11.97 For testing $H_o : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$, $t = 2.76$ with $p - value = 0.008$. The p-value is less than $\alpha = 0.01$ thus we reject H_o

11.98 Scatterplot of the data is given here:



a. From the scatterplot, there is a definite curvature in the relation between Durability and Concentration. A straight-line model would not appear to be appropriate.

b. The coefficient of determination, R^2, measures the strength of the linear (straight-line) relation only. A straight-line model does not adequately describe the relation between Durability and Concentrtation. This is indicated by the small percentage of the variation, 11.6%, in the values of Durability explained by the model containing just a linear relation with Concentration.

Chapter 12: Multiple Regression and the General Linear Model

12.1 a. $y_j = \mu_{i_j} + \epsilon_j$, with $i_j = 1, 2, \cdots, 12$

b. $x_1 = \begin{cases} 1 & \text{if} \quad V_1 = \text{1st Level} \\ 0 & \text{if} \quad \text{Otherwise} \end{cases}$ $\quad x_2 = \begin{cases} 1 & \text{if} \quad V_2 = \text{1st Level} \\ 0 & \text{if} \quad \text{Otherwise} \end{cases}$

$x_3 = \begin{cases} 1 & \text{if} \quad V_2 = \text{2nd Level} \\ 0 & \text{if} \quad \text{Otherwise} \end{cases}$ $\quad x_4 = \begin{cases} 1 & \text{if} \quad V_3 = \text{1st Level} \\ 0 & \text{if} \quad \text{Otherwise} \end{cases}$