12.1.1 (a) $4.2 + (1.7 \times 10) = 21.2$

(b) $3 \times 1.7 = 5.1$

(c) $P(N(4.2 + (1.7 \times 5), 3.2^2) \geq 12) = 0.587$

(d) $P(N(4.2 + (1.7 \times 8), 3.2^2) \leq 17) = 0.401$

(e) $P(N(4.2 + (1.7 \times 6), 3.2^2) \geq N(4.2 + (1.7 \times 7), 3.2^2)) = 0.354$

12.1.2 (a) $123.0 + (-2.16 \times 20) = 79.8$

(b) $-2.16 \times 10 = -21.6$

(c) $P(N(123.0 + (-2.16 \times 25), 4.1^2) \leq 60) = 0.014$

(d) $P(30 \leq N(123.0 + (-2.16 \times 40), 4.1^2) \leq 40) = 0.743$

(e) $P(N(123.0 + (-2.16 \times 30), 4.1^2) \leq N(123.0 + (-2.16 \times 27.5), 4.1^2)) = 0.824$

12.1.3 (a) $y = 5 + (0.9 \times 20) = 23.0$

(b) The expected value of the porosity decreases by $5 \times 0.9 = 4.5$.

(c) $P(N(5 + (0.9 \times 25), 1.4^2) \leq 30) = 0.963$

(d) $P\left(17 \leq N\left(5 + (0.9 \times 15), \frac{1.4^2}{4}\right) \leq 20\right) = 0.968$

12.1.4 Since the model minimizes the sum of squares of residuals (vertical differences) the model will change if the x and y variables are exchanged.

If there is a reason to consider that one variable can be naturally thought of as being "determined" by the choice of the other variable, then that indicates the appropriate choice of the x and y variables (the y variable should be the "determined" variable).

In addition, if the model is to be used to predict one variable for given values of the other variable, then that also indicates the appropriate choice of the x and y variables (the y variable should be the variable that is being predicted).

12.1.5 $P(N(675.30 - (5.87 \times 80), 7.32^2) \leq 220)$

$= P\left(N(0,1) \leq \frac{220-205.7}{7.32}\right)$

$= \Phi(1.954) = 0.975$

12.1.6 C

12.1.7 B

12.2.2 $n = 20$

$\sum_{i=1}^{20} x_i = 8.552$

$\sum_{i=1}^{20} y_i = 398.2$

$\sum_{i=1}^{20} x_i^2 = 5.196$

$\sum_{i=1}^{20} y_i^2 = 9356$

$\sum_{i=1}^{20} x_i y_i = 216.6$

$\bar{x} = \frac{8.552}{20} = 0.4276$

$\bar{y} = \frac{398.2}{20} = 19.91$

$S_{XX} = 5.196 - (20 \times 0.4276^2) = 1.539$

$S_{XY} = 216.6 - (20 \times 0.4276 \times 19.91) = 46.330$

Using these values

$\hat{\beta}_1 = \frac{46.330}{1.539} = 30.101$

$\hat{\beta}_0 = 19.91 - (30.10 \times 0.4276) = 7.039$

and

$SSE = 9356 - (7.039 \times 398.2) - (30.101 \times 216.6) = 33.291$

so that

$\hat{\sigma}^2 = \frac{33.291}{20-2} = 1.85.$

When $x = 0.5$ the fitted value is
$7.039 + (30.101 \times 0.5) = 22.09.$

12.2.3 $\hat{\beta}_0 = 39.5$

$\hat{\beta}_1 = -2.04$

$\hat{\sigma}^2 = 17.3$

$39.5 + (-2.04 \times (-2.0)) = 43.6$

12.2.4 (a) $\hat{\beta}_0 = -2,277$

$\hat{\beta}_1 = 1.003$

(b) $1.003 \times 1000 = \$1003$

(c) $-2277 + (1.003 \times 10000) = 7753$

The predicted cost is \$7,753,000.

(d) $\hat{\sigma}^2 = 774211$

(e) If the model is used then it would be extrapolation,
so the prediction may be inaccurate.

12.2.5 (a) $\hat{\beta}_0 = 36.19$

$\hat{\beta}_1 = 0.2659$

(b) $\hat{\sigma}^2 = 70.33$

(c) Yes, since $\hat{\beta}_1 > 0$.

(d) $36.19 + (0.2659 \times 72) = 55.33$

12.2.6 (a) $\hat{\beta}_0 = 54.218$

$\hat{\beta}_1 = -0.3377$

(b) No, $\hat{\beta}_1 < 0$ suggests that aerobic fitness deteriorates with age.

The predicted change in VO2-max for an additional 5 years of age is
$-0.3377 \times 5 = -1.6885$.

(c) $54.218 + (-0.3377 \times 50) = 37.33$

(d) If the model is used then it would be extrapolation,
so the prediction may be inaccurate.

(e) $\hat{\sigma}^2 = 57.30$

12.2.7 (a) $\hat{\beta}_0 = -29.59$

$\hat{\beta}_1 = 0.07794$

(b) $-29.59 + (0.07794 \times 2,600) = 173.1$

(c) $0.07794 \times 100 = 7.794$

(d) $\hat{\sigma}^2 = 286$

12.2.8 (a) $\hat{\beta}_0 = -1.911$

$\hat{\beta}_1 = 1.6191$

(b) $1.6191 \times 1 = 1.6191$

The expert is underestimating the times.

$-1.911 + (1.6191 \times 7) = 9.42$

(c) If the model is used then it would be extrapolation, so the prediction may be inaccurate.

(d) $\hat{\sigma}^2 = 12.56$

12.2.9 (a) $\hat{\beta}_0 = 12.864$

$\hat{\beta}_1 = 0.8051$

(b) $12.864 + (0.8051 \times 69) = 68.42$

(c) $0.8051 \times 5 = 4.03$

(d) $\hat{\sigma}^2 = 3.98$

12.2.10 B

12.2.11 C

12.3.1 (a) $(0.522 - (2.921 \times 0.142), 0.522 + (2.921 \times 0.142)) = (0.107, 0.937)$

(b) The t-statistic is

$\frac{0.522}{0.142} = 3.68$

and the p-value is 0.002.

12.3.2 (a) $(56.33 - (2.086 \times 3.78), 56.33 + (2.086 \times 3.78)) = (48.44, 64.22)$

(b) The t-statistic is

$\frac{56.33-50.0}{3.78} = 1.67$

and the p-value is 0.110.

12.3.3 (a) $s.e.(\hat{\beta}_1) = 0.08532$

(b) $(1.003 - (2.145 \times 0.08532), 1.003 + (2.145 \times 0.08532)) = (0.820, 1.186)$

(c) The t-statistic is

$\frac{1.003}{0.08532} = 11.76$

and the p-value is 0.000.

12.3.4 (a) $s.e.(\hat{\beta}_1) = 0.2383$

(b) $(0.2659 - (1.812 \times 0.2383), 0.2659 + (1.812 \times 0.2383)) = (-0.166, 0.698)$

(c) The t-statistic is

$\frac{0.2659}{0.2383} = 1.12$

and the p-value is 0.291.

(d) There is *not* sufficient evidence to conclude that on average trucks take longer to unload when the temperature is higher.

12.3.5 (a) $s.e.(\hat{\beta}_1) = 0.1282$

(b) $(-\infty, -0.3377 + (1.734 \times 0.1282)) = (-\infty, -0.115)$

(c) The t-statistic is

$\frac{-0.3377}{0.1282} = -2.63$

and the (two-sided) p-value is 0.017.

12.3.6 (a) $s.e.(\hat{\beta}_1) = 0.00437$

(b) $(0.0779 - (3.012 \times 0.00437), 0.0779 + (3.012 \times 0.00437)) = (0.0647, 0.0911)$

(c) The t-statistic is

$\frac{0.0779}{0.00437} = 17.83$

and the p-value is 0.000.

There is sufficient evidence to conclude that the house price depends upon the size of the house.

12.3.7 (a) $s.e.(\hat{\beta}_1) = 0.2829$

(b) $(1.619 - (2.042 \times 0.2829), 1.619 + (2.042 \times 0.2829)) = (1.041, 2.197)$

(c) If $\beta_1 = 1$ then the actual times are equal to the estimated times except for a constant difference of β_0.

The t-statistic is

$\frac{1.619-1.000}{0.2829} = 2.19$

and the p-value is 0.036.

12.3.8 (a) $s.e.(\hat{\beta}_1) = 0.06427$

(b) $(0.8051 - (2.819 \times 0.06427), 0.8051 + (2.819 \times 0.06427)) = (0.624, 0.986)$

(c) The t-statistic is

$\frac{0.8051}{0.06427} = 12.53$

and the p-value is 0.000.

There is sufficient evidence to conclude that resistance increases with temperature.

12.3.9 For the hypotheses

$H_0 : \beta_1 = 0$ versus $H_A : \beta_1 \neq 0$

the t-statistic is

$t = \frac{54.87}{21.20} = 2.588$

so that the p-value is $2 \times P(t_{18} \geq 2.588) = 0.019$.

12.3.10 The model is $y = \beta_0 + \beta_1 x$,

where y is the density of the ceramic and x is the baking time.

$n = 10$

$\sum_{i=1}^{10} x_i = 2600$

$\sum_{i=1}^{10} y_i = 31.98$

$\sum_{i=1}^{10} x_i^2 = 679850$

$\sum_{i=1}^{10} y_i^2 = 102.3284$

$\sum_{i=1}^{10} x_i y_i = 8321.15$

$\bar{x} = \frac{2600}{10} = 260$

$\bar{y} = \frac{31.98}{10} = 3.198$

$S_{XX} = 679850 - (10 \times 260^2) = 3850$

$S_{YY} = 102.3284 - (10 \times 3.198^2) = 0.05636$

$S_{XY} = 8321.15 - (10 \times 260 \times 3.198) = 6.35$

Using these values

$\hat{\beta}_1 = \frac{6.35}{3850} = 0.00165$

$\hat{\beta}_0 = 3.198 - (0.00165 \times 260) = 2.769$

and

$SSE = 102.3284 - (2.769 \times 31.98) - (0.00165 \times 8321.15) = 0.04589$

so that

$\hat{\sigma}^2 = \frac{0.04589}{10-2} = 0.00574.$

For the hypotheses

$H_0 : \beta_1 = 0$ versus $H_A : \beta_1 \neq 0$

the t-statistic is

$t = \frac{0.00165}{\sqrt{0.00574/3850}} = 1.35$

so that the p-value is $2 \times P(t_9 \geq 1.35) = 0.214$.

Therefore, the regression is not significant and there is not sufficient evidence to establish that the baking time has an effect on the density of the ceramic.

12.3.11 A

12.4.2 $(1392, 1400)$

12.4.3 $(21.9, 23.2)$

12.4.4 $(6754, 7755)$

12.4.5 $(33.65, 41.02)$

12.4.6 $(201.4, 238.2)$

12.4.7 $(-\infty, 10.63)$

12.4.8 $(68.07, 70.37)$

12.4.9 $\sum_{i=1}^{8} x_i = 122.6$

$\sum_{i=1}^{8} x_i^2 = 1939.24$

$\bar{x} = \frac{122.6}{8} = 15.325$

$S_{XX} = 1939.24 - \frac{122.6^2}{8} = 60.395$

With $t_{0.025,6} = 2.447$ the confidence interval is

$\beta_0 + (\beta_1 \times 15) \in 75.32 + (0.0674 \times 15) \pm 2.447 \times 0.0543 \times \sqrt{\frac{1}{8} + \frac{(15-15.325)^2}{60.395}}$

which is $(76.284, 76.378)$.

12.4.10 $n = 7$

$\sum_{i=1}^{7} x_i = 240.8$

$\sum_{i=1}^{7} y_i = 501.8$

$\sum_{i=1}^{7} x_i^2 = 8310.44$

$\sum_{i=1}^{7} y_i^2 = 36097.88$

$\sum_{i=1}^{7} x_i y_i = 17204.87$

$\bar{x} = \frac{240.8}{7} = 34.400$

$\bar{y} = \frac{501.8}{7} = 71.686$

$S_{XX} = 8310.44 - \frac{240.8^2}{7} = 26.920$

$S_{YY} = 36097.88 - \frac{501.8^2}{7} = 125.989$

$S_{XY} = 17204.87 - \frac{240.8 \times 501.8}{7} = -57.050$

Using these values

$\hat{\beta}_1 = \frac{-57.050}{26.920} = -2.119$

$\hat{\beta}_0 = 71.686 - (-2.119 \times 34.400) = 144.588$

and

$SSE = 36097.88 - (144.588 \times 501.8) - (-2.119 \times 17204.87) = 5.086$

so that

$\hat{\sigma}^2 = \frac{5.086}{7-2} = 1.017.$

With $t_{0.005,5} = 4.032$ the confidence interval is

$\beta_0 + (\beta_1 \times 35) \in 144.588 + (-2.119 \times 35) \pm 4.032 \times \sqrt{1.017} \times \sqrt{\frac{1}{7} + \frac{(35-34.400)^2}{26.920}}$

which is

$70.414 \pm 1.608 = (68.807, 72.022).$

12.5.1 $(1386, 1406)$

12.5.2 $(19.7, 25.4)$

12.5.3 $(5302, 9207)$

12.5.4 $(21.01, 53.66)$

12.5.5 $(165.7, 274.0)$

12.5.6 $(-\infty, 15.59)$

12.5.7 $(63.48, 74.96)$

12.5.8 $\bar{x} = \frac{603.36}{30} = 20.112$

$S_{XX} = 12578.22 - \frac{603.36^2}{30} = 443.44$

$\hat{\sigma}^2 = \frac{329.77}{30-2} = 11.778$

With $t_{0.025,28} = 2.048$ the prediction interval is

$51.98 + (3.44 \times 22) \pm 2.048 \times \sqrt{11.778} \times \sqrt{\frac{31}{30} + \frac{(22-20.112)^2}{443.44}}$

$= 127.66 \pm 7.17 = (120.49, 134.83)$.

12.5.9 $n = 7$

$\sum_{i=1}^{7} x_i = 142.8$

$\sum_{i=1}^{7} y_i = 361.5$

$\sum_{i=1}^{7} x_i^2 = 2942.32$

$\sum_{i=1}^{7} y_i^2 = 18771.5$

$\sum_{i=1}^{7} x_i y_i = 7428.66$

$\bar{x} = \frac{142.8}{7} = 20.400$

$\bar{y} = \frac{361.5}{7} = 51.643$

$S_{XX} = 2942.32 - \frac{142.8^2}{7} = 29.200$

$S_{YY} = 18771.5 - \frac{361.5^2}{7} = 102.607$

$S_{XY} = 7428.66 - \frac{142.8 \times 361.5}{7} = 54.060$

Using these values

$\hat{\beta}_1 = \frac{54.060}{29.200} = 1.851$

$\hat{\beta}_0 = 51.643 - (1.851 \times 20.400) = 13.875$

and

$SSE = 18771.5 - (13.875 \times 361.5) - (1.851 \times 7428.66) = 2.472$

so that

$\hat{\sigma}^2 = \frac{2.472}{7-2} = 0.494.$

With $t_{0.005,5} = 4.032$ the prediction interval is

$13.875 + (1.851 \times 20) \pm 4.032 \times \sqrt{0.494} \times \sqrt{\frac{8}{7} + \frac{(20-20.400)^2}{29.200}}$

which is

$50.902 \pm 3.039 = (47.864, 53.941).$

12.5.10 A

12.6.1

Source	df	SS	MS	F	p-value
Regression	1	40.53	40.53	2.32	0.137
Error	33	576.51	17.47		
Total	34	617.04			

$R^2 = \frac{40.53}{617.04} = 0.066$

12.6.2

Source	df	SS	MS	F	p-value
Regression	1	120.61	120.61	6.47	0.020
Error	19	354.19	18.64		
Total	20	474.80			

$R^2 = \frac{120.61}{474.80} = 0.254$

12.6.3

Source	df	SS	MS	F	p-value
Regression	1	870.43	870.43	889.92	0.000
Error	8	7.82	0.9781		
Total	9	878.26			

$R^2 = \frac{870.43}{878.26} = 0.991$

12.6.4

Source	df	SS	MS	F	p-value
Regression	1	6.82×10^6	6.82×10^6	1.64	0.213
Error	23	95.77×10^6	4.16×10^6		
Total	24	102.59×10^6			

$R^2 = \frac{6.82\times10^6}{102.59\times10^6} = 0.06$

12.6.5

Source	df	SS	MS	F	p-value
Regression	1	10.71×10^7	10.71×10^7	138.29	0.000
Error	14	1.08×10^7	774,211		
Total	15	11.79×10^7			

$R^2 = \frac{10.71\times10^7}{11.79\times10^7} = 0.908$

12.6.6

Source	df	SS	MS	F	p-value
Regression	1	87.59	87.59	1.25	0.291
Error	10	703.33	70.33		
Total	11	790.92			

$R^2 = \frac{87.59}{790.92} = 0.111$

The large p-value implies that there is not sufficient evidence to conclude that on average trucks take longer to unload when the temperature is higher.

12.6.7

Source	df	SS	MS	F	p-value
Regression	1	397.58	397.58	6.94	0.017
Error	18	1031.37	57.30		
Total	19	1428.95			

$R^2 = \frac{397.58}{1428.95} = 0.278$

The R^2 value implies that about 28% of the variability in VO2-max can be accounted for by changes in age.

12.6.8

Source	df	SS	MS	F	p-value
Regression	1	90907	90907	318.05	0.000
Error	13	3716	286		
Total	14	94622			

$R^2 = \frac{90907}{94622} = 0.961.$

The high R^2 value indicates that there is almost a perfect linear relationship between appraisal value and house size.

12.6.9

Source	df	SS	MS	F	p-value
Regression	1	411.26	411.26	32.75	0.000
Error	30	376.74	12.56		
Total	31	788.00			

$R^2 = \frac{411.26}{788.00} = 0.522$

The p-value is not very meaningful because it tests the null hypothesis that the actual times are unrelated to the estimated times.

12.6.10

Source	df	SS	MS	F	p-value
Regression	1	624.70	624.70	156.91	0.000
Error	22	87.59	3.98		
Total	23	712.29			

$R^2 = \frac{624.70}{712.29} = 0.877$

12.6.11 A

12.6.12 A

12.7.1 There is no suggestion that the fitted regression model is not appropriate.

12.7.2 There is no suggestion that the fitted regression model is not appropriate.

12.7.3 There is a possible suggestion of a slight reduction in the variability of the VO2-max values as age increases.

12.7.4 The observation with an area of 1,390 square feet appears to be an outlier.

There is no suggestion that the fitted regression model is not appropriate.

12.7.5 The variability of the actual times increases as the estimated time increases.

12.7.6 There is a possible suggestion of a slight increase in the variability of the resistances at higher temperatures.

12.7.7 D

12.8.1 The model

$y = \gamma_0 \, e^{\gamma_1 x}$

is appropriate.

A linear regression can be performed with $\ln(y)$ as the dependent variable and with x as the input variable.

$\hat{\gamma}_0 = 9.12$

$\hat{\gamma}_1 = 0.28$

$\hat{\gamma}_0 \, e^{\hat{\gamma}_1 \times 2.0} = 16.0$

12.8.2 The model

$y = \frac{x}{\gamma_0 + \gamma_1 x}$

is appropriate.

A linear regression can be performed with $\frac{1}{y}$ as the dependent variable and with $\frac{1}{x}$ as the input variable.

$\hat{\gamma}_0 = 1.067$

$\hat{\gamma}_1 = 0.974$

$\frac{2.0}{\hat{\gamma}_0 + (\hat{\gamma}_1 \times 2.0)} = 0.66$

12.8.3 $\hat{\gamma}_0 = 8.81$

$\hat{\gamma}_1 = 0.523$

$\gamma_0 \in (6.84, 11.35)$

$\gamma_1 \in (0.473, 0.573)$

12.8.4 (b) $\hat{\gamma}_0 = 89.7$

$\hat{\gamma}_1 = 4.99$

(c) $\gamma_0 \in (68.4, 117.7)$

$\gamma_1 \in (4.33, 5.65)$

12.8.5 $\hat{\gamma}_0 = e^{\hat{\beta}_0} = e^{2.628} = 13.85$

$\hat{\gamma}_1 = \hat{\beta}_1 = 0.341$

With $t_{0.025,23} = 2.069$ the confidence interval for γ_1 (and β_1) is $0.341 \pm (2.069 \times 0.025) = (0.289, 0.393)$.

12.8.6 The model can be rewritten

$y = \gamma_0 \ln(\gamma_1) - 2\gamma_0 \ln(x)$.

If a simple linear regression is performed with $\ln(x)$ as the input variable and y as the output variable, then

$\hat{\beta}_0 = \hat{\gamma}_0 \ln(\hat{\gamma}_1)$

and

$\hat{\beta}_1 = -2\hat{\gamma}_0$.

Therefore,

$\hat{\gamma}_0 = \frac{-\hat{\beta}_1}{2}$

and

$\hat{\gamma}_1 = e^{-2\hat{\beta}_0/\hat{\beta}_1}$.

12.8.7 $\hat{\gamma}_0 = 12.775$

$\hat{\gamma}_1 = -0.5279$

When the crack length is 2.1 the expected breaking load is

$12.775 \times e^{-0.5279 \times 2.1} = 4.22$.

12.9.3 The sample correlation coefficient is $r = 0.95$.

12.9.4 The sample correlation coefficient is $r = 0.33$.

12.9.5 The sample correlation coefficient is $r = -0.53$.

12.9.6 The sample correlation coefficient is $r = 0.98$.

12.9.7 The sample correlation coefficient is $r = 0.72$.

12.9.8 The sample correlation coefficient is $r = 0.94$.

12.9.9 The sample correlation coefficient is $r = 0.431$.

12.9.10 It is known that $\hat{\beta}_1 > 0$ but nothing is known about the p-value.

12.9.11 The variables A and B may both be related to a third surrogate variable C. It is possible that the variables A and C have a causal relationship, and that the variables B and C have a causal relationship, without there being a causal relationship between the variables A and B.

12.9.12 A

12.9.13 A

12.9.14 B

12.12.1 (a) $\hat{\beta}_0 = 95.77$

$\hat{\beta}_1 = -0.1003$

$\hat{\sigma}^2 = 67.41$

(b) The sample correlation coefficient is $r = -0.69$.

(c) $(-0.1003 - (2.179 \times 0.0300), -0.1003 + (2.179 \times 0.0300))$

$= (-0.1657, -0.0349)$

(d) The t-statistic is

$\frac{-0.1003}{0.0300} = -3.34$

and the p-value is 0.006.

There is sufficient evidence to conclude that the time taken to finish the test depends upon the SAT score.

(e) $-0.1003 \times 10 = -1.003$

(f) $95.77 + (-0.1003 \times 550) = 40.6$

The confidence interval is $(35.81, 45.43)$.

The prediction interval is $(22.09, 59.15)$.

(g) There is no suggestion that the fitted regression model is not appropriate.

12.12.2 (a) $\hat{\beta}_0 = 18.35$

$\hat{\beta}_1 = 6.72$

$\hat{\sigma}^2 = 93.95$

(b) The sample correlation coefficient is $r = 0.84$.

(c) The t-statistic is 23.91 and the p-value is 0.000.

There is sufficient evidence to conclude that the amount of scrap material depends upon the number of passes.

(d) $(6.72 - (1.960 \times 0.2811), 6.72 + (1.960 \times 0.2811)) = (6.17, 7.27)$

(e) It increases by $6.72 \times 1 = 6.72$.

(f) $18.35 + (6.72 \times 7) = 65.4$

The prediction interval is $(46.1, 84.7)$.

(g) Observations $x = 2$, $y = 67.71$ and $x = 9$, $y = 48.17$ have standardized residuals with absolute values larger than three.

The linear model is reasonable, but a non-linear model with a decreasing slope may be more appropriate.

12.12.3 $\hat{\beta}_0 = 29.97$

$\hat{\beta}_1 = 0.0923$

$\hat{\sigma} = 0.09124$

The t-statistic for the null hypothesis $H_0 : \beta_1 = 0$ is

$\frac{0.09234}{0.01026} = 9.00$

and the p-value is 0.000.

There is a significant association between power loss and bearing diameter.

The sample correlation coefficient is $r = 0.878$.

The fitted value for the power loss of a new engine with a bearing diameter of 25.0 is 32.28 and a 95% prediction interval is $(32.09, 32.47)$.

There are no data points with values $\frac{e_i}{\hat{\sigma}}$ larger than three in absolute value.

12.12.4 $\hat{\beta}_0 = 182.61$

$\hat{\beta}_1 = 0.8623$

$\hat{\sigma} = 32.08$

The sample correlation coefficient is $r = 0.976$.

When the energy lost by the hot liquid is 500, the fitted value for the energy gained by the cold liquid is 613.8 and a 95% prediction interval is $(547.1, 680.3)$.

12.12.5 $\hat{\beta}_0 = 3.252$

$\hat{\beta}_1 = 0.01249$

$\hat{\sigma} = 2.997$

The t-statistic for the null hypothesis $H_0 : \beta_1 = 0$ is

$\frac{0.01249}{0.003088} = 4.04$

and the p-value is 0.001.

There is a significant association between the pulse time and the capacitance value.

The sample correlation coefficient is $r = 0.690$.

For a capacitance of 1700 microfarads, the fitted value for the pulse time is 24.48 milliseconds and a 95% prediction interval is $(17.98, 30.99)$.

The data point with a pulse time of 28.52 milliseconds for a capacitance of 1400 microfarads has a residual $e_i = 7.784$ so that

$\frac{e_i}{\sigma} = \frac{7.784}{2.997} = 2.60$.

12.12.6 (b) $\hat{\gamma}_0 = 0.199$

$\hat{\gamma}_1 = 0.537$

(c) $\gamma_0 \in (0.179, 0.221)$

$\gamma_1 \in (0.490, 0.584)$

(d) $0.199 + \frac{0.537}{10.0} = 0.253$

12.12.7 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the strength of the chemical solution and x is the amount of the catalyst.

$n = 8$

$\sum_{i=1}^{8} x_i = 197$

$\sum_{i=1}^{8} y_i = 225$

$\sum_{i=1}^{8} x_i^2 = 4951$

$\sum_{i=1}^{8} y_i^2 = 7443$

$\sum_{i=1}^{8} x_i y_i = 5210$

$\bar{x} = \frac{197}{8} = 24.625$

$\bar{y} = \frac{225}{8} = 28.125$

$S_{XX} = 4951 - (8 \times 24.625^2) = 99.874$

$S_{YY} = 7443 - (8 \times 28.125^2) = 1114.875$

$S_{XY} = 5210 - (8 \times 24.625 \times 28.125) = -330.625$

Using these values

$\hat{\beta}_1 = \frac{-330.625}{99.874} = -3.310$

$\hat{\beta}_0 = 28.125 - (-3.310 \times 24.625) = 109.643$

and

$SSE = 7443 - (109.643 \times 225) - (-3.310 \times 5210) = 20.378$

so that

$\hat{\sigma}^2 = \frac{20.378}{8-2} = 3.396.$

(b) The t-statistic is

$$\frac{-3.310}{\sqrt{3.396/99.874}} = -17.95$$

so that the p-value is $2 \times P(t_6 > 17.95) = 0.000$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and the regression is significant.

There is sufficient evidence to establish that the amount of the catalyst does effect the strength of the chemical solution.

(c) With $t_{0.025,6} = 2.447$ the prediction interval is

$$109.643 + (-3.310 \times 21.0) \pm 2.447 \times \sqrt{3.396} \times \sqrt{1 + \frac{1}{8} + \frac{(21.0-24.625)^2}{99.874}}$$

$$= 40.125 \pm 5.055$$

$$= (35.070, 45.180).$$

(d) $e_2 = 17 - (109.643 + (-3.310 \times 28)) = 0.05$

12.12.8 (a) $\bar{x} = \frac{856}{20} = 42.8$

$$S_{XX} = 37636 - (20 \times 42.8^2) = 999.2$$

With $t_{0.025,18} = 2.101$ the prediction interval is

$$123.57 - (3.90 \times 40) \pm 2.101 \times 11.52 \times \sqrt{\frac{21}{20} + \frac{(40-42.8)^2}{999.2}}$$

$$= (-57.32, -7.54)$$

(b) $SST = 55230 - \frac{(-869)^2}{20} = 17472$

Source	df	SS	MS	F	p-value
Regression	1	15083	15083	114	0.000
Error	18	2389	133		
Total	19	17472			

$R^2 = \frac{15083}{17472} = 86\%$

12.12.9 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the bacteria yield and x is the temperature.

$n = 8$

$\sum_{i=1}^{8} x_i = 197$

$\sum_{i=1}^{8} y_i = 429$

$\sum_{i=1}^{8} x_i^2 = 4943$

$\sum_{i=1}^{8} y_i^2 = 23805$

$\sum_{i=1}^{8} x_i y_i = 10660$

$\bar{x} = \frac{197}{8} = 24.625$

$\bar{y} = \frac{429}{8} = 53.625$

$S_{XX} = 4943 - (8 \times 24.625^2) = 91.875$

$S_{YY} = 23805 - (8 \times 53.625^2) = 799.875$

$S_{XY} = 10660 - (8 \times 24.625 \times 53.625) = 95.875$

Using these values

$\hat{\beta}_1 = \frac{95.875}{91.875} = 1.044$

$\hat{\beta}_0 = 53.625 - (1.044 \times 24.625) = 27.93$

and

$SSE = 23805 - (27.93 \times 429) - (1.044 \times 10660) = 699.8$

so that

$\hat{\sigma}^2 = \frac{699.8}{8-2} = 116.6.$

(b) The t-statistic is

$\frac{1.044}{\sqrt{116.6/91.875}} = 0.93$

so that the p-value is $2 \times P(t_6 > 0.93) = 0.390$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ cannot be rejected and the regression is not significant.
There is not sufficient evidence to establish that the bacteria yield does depend on temperature.

(c) $e_1 = 54 - (27.93 + 1.044 \times 22) = 3.1$

12.12.10 The F-statistic from the analysis of variance table is

$$F = \frac{MSR}{MSE} = \frac{(n-2)SSR}{SSE} = \frac{(n-2)R^2}{1-R^2} = \frac{18\times0.853}{1-0.853} = 104.4$$

The p-value is $P(F_{1,18} \geq 104.4) = 0.000$.

12.12.11 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the amount of gold obtained and x is the amount of ore processed.

$n = 7$

$\sum_{i=1}^{7} x_i = 85.8$

$\sum_{i=1}^{7} y_i = 87.9$

$\sum_{i=1}^{7} x_i^2 = 1144.40$

$\sum_{i=1}^{7} y_i^2 = 1158.91$

$\sum_{i=1}^{7} x_i y_i = 1146.97$

$\bar{x} = \frac{85.8}{7} = 12.257$

$\bar{y} = \frac{87.9}{7} = 12.557$

$S_{XX} = 1144.40 - (7 \times 12.257^2) = 92.737$

$S_{YY} = 1158.91 - (7 \times 12.557^2) = 55.137$

$S_{XY} = 1146.97 - (7 \times 12.257 \times 12.557) = 69.567$

Using these values

$\hat{\beta}_1 = \frac{69.567}{92.737} = 0.750$

and

$\hat{\beta}_0 = 12.557 - (0.750 \times 12.257) = 3.362.$

(b) Since

$SSE = 1158.91 - (3.362 \times 87.9) - (0.750 \times 69.567) = 2.9511$

it follows that

$\hat{\sigma}^2 = \frac{2.9511}{7-2} = 0.5902.$

The t-statistic is

$\frac{0.750}{\sqrt{0.5902/92.737}} = 9.40$

so that the p-value is $2 \times P(t_5 > 9.40) \simeq 0.$

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(c) $r = \frac{69.567}{\sqrt{92.737}\sqrt{55.137}} = 0.973$

(d) $R^2 = r^2 = 0.973^2 = 0.946$

(e) With $t_{0.025,5} = 2.571$ the prediction interval is

$3.362 + (0.750 \times 15) \pm 2.571 \times \sqrt{0.5902} \times \sqrt{1 + \frac{1}{7} + \frac{(15-12.257)^2}{92.737}}$

which is

$14.615 \pm 2.185 = (12.430, 16.800)$.

(f) $e_1 = 8.9 - (3.362 + 0.750 \times 7.3) = 0.06$
$e_2 = 11.3 - (3.362 + 0.750 \times 9.1) = 1.11$
$e_3 = 10.6 - (3.362 + 0.750 \times 10.2) = -0.41$
$e_4 = 11.6 - (3.362 + 0.750 \times 11.5) = -0.38$
$e_5 = 12.2 - (3.362 + 0.750 \times 13.2) = -1.06$
$e_6 = 15.7 - (3.362 + 0.750 \times 16.1) = 0.26$
$e_7 = 17.6 - (3.362 + 0.750 \times 18.4) = 0.43$

12.12.12 (a) False

(b) True

(c) True

(d) False

(e) True

(f) False

(g) True

(h) False

(i) True

(j) True

(k) False

(l) True

(m) False

12.12.13 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the downloading time and x is the file size.

$n = 9$

$\sum_{i=1}^{9} x_i = 50.06$

$\sum_{i=1}^{9} y_i = 1156$

$\sum_{i=1}^{9} x_i^2 = 319.3822$

$\sum_{i=1}^{9} y_i^2 = 154520$

$\sum_{i=1}^{9} x_i y_i = 6894.34$

$\bar{x} = \frac{50.06}{9} = 5.562$

$\bar{y} = \frac{1156}{9} = 128.444$

$S_{XX} = 319.3822 - (9 \times 5.562^2) = 40.9374$

$S_{YY} = 154520 - (9 \times 128.444^2) = 6038.2223$

$S_{XY} = 6894.34 - (9 \times 5.562 \times 128.444) = 464.4111$

Using these values

$\hat{\beta}_1 = \frac{464.4111}{40.9374} = 11.344$

$\hat{\beta}_0 = 128.444 - (11.344 \times 5.562) = 65.344$

and

$SSE = 154520 - (65.344 \times 1156) - (11.344 \times 6894.34) = 769.737$

so that

$\hat{\sigma}^2 = \frac{769.737}{9-2} = 109.962.$

(b) The t-statistic is

$\frac{11.344}{\sqrt{109.962/40.9374}} = 6.92$

so that the p-value is $2 \times P(t_7 > 6.92) \simeq 0.$

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(c) $65.344 + (11.344 \times 6) = 133.41$

(d) Since

$SSR = SST - SSE = 6038.2223 - 769.737 = 5268.485$

it follows that

$R^2 = \frac{5268.485}{6038.2223} = 87.25\%$.

(e) With $t_{0.025,7} = 2.365$ the prediction interval is

$65.344 + (11.344 \times 6) \pm 2.365 \times \sqrt{109.962} \times \sqrt{1 + \frac{1}{9} + \frac{(6-5.562)^2}{40.9374}}$

which is

$133.41 \pm 26.19 = (107.22, 159.60)$.

(f) $103 - (65.344 + (4.56 \times 6)) = -14.07$

(g) $r = \sqrt{R^2} = \sqrt{0.8725} = 0.934$

(h) It may be quite unreliable to extrapolate the model to predict the downloading time of a file of size 0.40.

12.12.14 (a) The model is $y = \beta_0 + \beta_1 x$ where y is the speed and x is the depth.

$n = 18$

$\sum_{i=1}^{18} x_i = 56.988$

$\sum_{i=1}^{18} y_i = 27343.03$

$\sum_{i=1}^{18} x_i^2 = 234.255$

$\sum_{i=1}^{18} y_i^2 = 41535625$

$\sum_{i=1}^{18} x_i y_i = 86560.46$

$\bar{x} = \frac{56.988}{18} = 3.166$

$\bar{y} = \frac{27343.03}{18} = 1519.06$

$S_{XX} = 234.255 - (18 \times 3.166^2) = 53.8307$

$S_{YY} = 41535625 - (18 \times 1519.06^2) = 5.2843$

$S_{XY} = 86560.46 - (18 \times 3.166 \times 1519.06) = -16.666$

Using these values

$\hat{\beta}_1 = \frac{-16.666}{53.8307} = -0.3096$

and

$\hat{\beta}_0 = 1519.06 - (-0.31 \times 3.16) = 1520.04.$

(b) With

$SSE = 41535625 - (1520.04 \times 27343.03) - (-0.3096 \times 86560.46) = 0.1232$

it follows that

$\hat{\sigma}^2 = \frac{0.1232}{18-2} = 0.00770.$

(c) The t-statistic is

$$\frac{-0.3096}{\sqrt{0.00770/53.8307}} = -25.89$$

so that the p-value is $2 \times P(t_{16} > 25.84) \simeq 0$.

Therefore, the null hypothesis $H_0 : \beta_1 = 0$ can be rejected and it can be concluded that the regression is significant.

(d) With $t_{0.025,16} = 2.120$ the confidence interval is

$$\beta_0 + (\beta_1 \times 4) \in 1520.04 + (-0.3096 \times 4) \pm 2.120 \times \sqrt{0.00770} \times \sqrt{\frac{1}{18} + \frac{(4-3.166)^2}{53.8307}}$$

which is

$$1518.80 \pm 0.05 = (1518.75, 1518.85).$$

(e) Since

$$SSR = SST - SSE = 5.2843 - 0.1232 = 5.1611$$

it follows that

$$R^2 = \frac{5.1611}{5.2843} = 97.7\%.$$

12.12.24 C

12.12.25 D

12.12.30 B

12.12.31 B

12.12.32 A

12.12.33 B

12.12.34 C

12.12.35 A

12.12.36 A

12.12.37 D

12.12.38 C

12.12.39 D

12.12.40 C

12.12.41 C

12.12.42 C

12.12.43 B

12.12.44 D

12.12.45 A

12.12.46 C

12.12.47 D