Answers to Exercises

Chapters 1–3

A. 1. $b_1 = \frac{158}{110} = 1.44$; $b_0 = \frac{102}{11} = 9.27$; $\hat{Y} = 9.27 + 1.44X$.

2.

Analysis of Variance

Source of Variation	df	SS	MS	F
Total (corrected)	10	248.18		
Regression	1	$\frac{(158)^2}{110}$	226.95	96.17*
Residual	9	21.23	2.36	

*The hypothesis, $H_0: \beta_1 = 0$ is tested with $\alpha = 0.05$ by comparing the computed $F(1, 9)$ statistic with the critical $F(1, 9)$ for $\alpha = 0.05$. From the $\alpha = 0.05$ F-table, we find $F(1, 9, 0.95) = 5.12$. Since 96.17 is greater than 5.12, reject the hypothesis $\beta_1 = 0$.

3. The 95% confidence limits for β_1 are

$$1.11 \leq \beta_1 \leq 1.77.$$

4. The 95% confidence limits for the true average value of Y at $X_0 = 3$ are

$$12.15 \leq \text{true average } Y \text{ at } X_0 = 3 \leq 15.03.$$

5. The 95% confidence limits for the difference between the true average value of Y at $X_1 = 3$ and the true average value of Y at $X_2 = 2$. First determine the algebraic difference between $\hat{Y}_1$ and $\hat{Y}_2$:

$$\hat{Y}_1 = b_0 + b_1(3), \qquad \hat{Y}_2 = b_0 + b_1(-2).$$

Thus

$$\hat{Y}_1 - \hat{Y}_2 = b_1(3 + 2) = 5b_1 = 5(1.44) = 7.20,$$

$$s^2_{(\hat{Y}_1 - \hat{Y}_2)} = 25s^2_{b_1} = 25\left(\frac{2.36}{110}\right) = 0.53635,$$

$$s_{(\hat{Y}_1 - \hat{Y}_2)} = \sqrt{0.53635} = 0.732,$$

$$ts_{(\hat{Y}_1 - \hat{Y}_2)} = (2.262)(0.732) = 1.656.$$

Thus the 95% confidence band on the true difference is

$$7.20 - 1.66 \leq \text{true difference} \leq 7.20 + 1.66$$
$$5.54 \leq \text{true difference} \leq 8.86.$$

6. Calculate the residuals and look for patterns.

X	Y	$\hat{Y}$	$Y - \hat{Y}$
−5	1	2.07	−1.07
−4	5	3.51	1.49
−3	4	4.95	−0.95
−2	7	6.39	0.61
−1	10	7.83	2.17
0	8	9.27	−1.27
1	9	10.71	−1.71
2	13	12.15	0.85
3	14	13.59	0.41
4	13	15.03	−2.03
5	18	16.47	1.53

There is no obvious alternative to the model.

7. If the tentative assumption of a first-order model is reasonable, there is little point in using eleven different experimental levels. Of course, we need at least two levels to estimate the parameters in the model, and at least one more to detect curvature in the true model, if curvature exists. By taking repeat observations at some or all levels, we can obtain a pure error estimate of σ^2 to use in checking lack of fit. Thus for an experiment of about the same size, one possibility would be to choose three widely spaced levels—the extremes of the X-range and the center, for example—and to take four observations at each of these levels. This would lead to an analysis of variance table of the form below.

ANOVA (n = 12)

Source of Variation	df
Total (corrected)	11
Regression	1
Residual	10
Lack of fit	1
Pure error	9

Since we now have only 1 degree of freedom for lack of fit, this is not entirely satisfactory. Slightly better would be the choice of three runs at each of four levels:

ANOVA (n = 12)

Source of Variation	df
Total (corrected)	11
Regression	1
Residual	10
Lack of fit	2
Pure error	8

There are many other possibilities. See Section 1.8.

B. 1. Randomized order.

2a. $\hat{Y} = 0.5 + 0.5X$.

2b.

Source	df	SS	MS
Corrected total $\Sigma(Y_i - \bar{Y})^2$	19	83.2	
Due to regression $\dfrac{[\Sigma(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\Sigma(X_i - \bar{Y})^2}$	1	40.0	40.0*
Residual	18	43.2	2.4

2c. (1) $\hat{Y} = 3.0 \pm 0.73 = 2.27$ to 3.73.
(2) $\hat{Y} = 5.0 \pm 1.26 = 3.74$ to 6.26.

3a.

Residual	18	43.2	2.4	
Lack of fit	3	1.2	0.4	NS
Pure error	15	42.0	2.8	

No significant lack of fit.

3b. Yes.

4a. Same ANOVA and conclusion as in 3a.

4b. Confidence limits not applicable. Error variance is dependent on level of Y.

4c. First-order model suitable

5a.

Residual	18	43.2	2.4	
Lack of fit	3	20.0	6.67	*
Pure error	15	23.2	1.55	

Significant lack of fit indicates inadequacy of model.

5b. Since the model is incorrect, confidence intervals will be invalid.

5c. A second-order model is suggested.

C. Best fitting straight line is

$$\hat{Y} = b_0 + b_1 X = 129.7872 - 24.0199X.$$

ANOVA

Source	df	SS	MS	F
Total	12	3396.62		
Regression	1	3293.77	3293.77	
Residual	11	102.85	9.35	
Lack of fit	5	91.08	18.22	9.30*
Pure error	6	11.77	1.96	

The model is inadequate.

D. First note that the sum of squares function for the no-intercept model is

$$\begin{aligned}\Sigma(Y_i - \beta X_i)^2 &= \Sigma\{Y_i - bX_i + bX_i - \beta X_i\}^2 \\ &= \Sigma(Y_i - bX_i)^2 + (b - \beta)^2 \Sigma X_i^2 \\ &\quad + 2\Sigma(Y_i - bX_i)(b - \beta)X_i.\end{aligned}$$

The cross-product term can be rewritten as

$$2(b - \beta)\Sigma(X_i Y_i - bX_i^2) = 2(b - \beta)\{\Sigma X_i Y_i - b\Sigma X_i^2\}$$

and vanishes by definition of b. What is left is a function of β that is minimized when $\beta = b$. This proves the least squares result. For the rest of the question, we see first that $n + 1 = (1 + a)^2$. If we use (U, V) to denote the new data, then, for example,

$$\begin{aligned}\overline{U} &= (\text{Sum of all new observations})/(n + 1) \\ &= (n\overline{X} + n\overline{X}/a)/(n + 1) \\ &= n\overline{X}/\{a(1 + a)\}, \\ S_{UU} &= \Sigma X_i^2 + (m\overline{X})^2 - (n + 1)\overline{U}^2 \\ &= \Sigma X_i^2 + n^2\overline{X}^2/a^2 - n^2\overline{X}^2(1 + a)^2/\{a^2(1 + a)^2\} \\ &= \Sigma X_i^2.\end{aligned}$$

Similarly $S_{UV} = \Sigma X_i Y_i$ and so $b_1 = S_{UV}/S_{UU} = b$.

(The estimate $b_0 = \overline{V} - b\overline{U} = n(\overline{Y} - b\overline{X})/\{a(1 + a)\}$ is not zero in general. Ignore it! This trick gets the right slope, but not the right zero intercept in general!)

For our example, the equations obtained are $\hat{Y} = 0.22857X$ for the three data points, and $\hat{Y} = 0.1714 + 0.22857X$ for the four data points. Note that the standard error and t-statistic for $b = 0.22857$ remain the same. You might consider why this happens.

E. There are only seven observations, so one cannot hope for too much here. For response Y_2, we get:

1. $\hat{Y} = 38.067 - 358.2X$. We leave the plot to the reader.

2. Residuals 19.36, 2.11, 0.25, −15.99, −13.55, −0.11, 7.92. Their sum is −0.01, zero within rounding error.

3.

Source	df	SS	MS	F	
$b_1\|b_0$	1	891.5	891.5	5.06	($p = 0.074$)
Residual	5	881.4	176.3	—	
Total, corrected		1772.8	—	—	

4. $\text{se}(b_0) = 9.146 \quad \text{se}(b_1) = 159.3$

5. $\text{se}(\hat{Y}) = \left\{\frac{1}{7} + \frac{(X_0 - 0.048)^2}{0.006946}\right\}^{1/2} (176.3)^{1/2}.$

X_0	$\hat{Y}$	se($\hat{Y}$)	95% Limits ($t = 2.571$)
0.01	34.49	7.864	14.27, 54.71
0.03	27.32	5.780	12.46, 42.18
0.05	20.16	5.029	7.23, 33.09
0.07	12.99	6.121	−2.98, 28.96
0.09	5.83	8.364	−15.67, 27.33
0.10	2.25	9.686	−22.65, 27.15

6. For F see (3). Note that $t = -358.2/159.3 = -2.249 = -F^{1/2}$, $R^2 = 891.5/1772.8 = 0.5029$.

7. The pure error SS = 162.33 + 21.78 = 184.1 with 2 + 1 = 3 df.
The lack of fit SS = 881.4 − 184.1 = 697.3 with 5 − 3 = 2 df.
$F = (697.3/2)/(184.1/3) = 5.68$ with (2,3) df, close to the 10% upper-tail point of 5.46. Technically then, no lack of fit is shown. Of course, we are working with very few df here.

8. The straight line gives some basis ($p = 0.07$ for the F-test) for the idea that the trend is downward, that is, that higher costs for alcohol are associated with fewer deaths. Whether the deaths really depend on the costs is not known. It would be nice to think we *could* reduce the number of deaths by raising the price. The early data vary much more than the later data. This would seem to deny our assumption that the σ^2 is constant, but with so few points we cannot be sure. Obviously the practical conclusion is to raise prices in France, Italy, Germany, and Belgium and see what happens!

F. 1. $\hat{Y} = -21.33 + 5X$

2. $2.984 \le \beta_1 \le 7.016$

ANOVA

Source	df	SS	MS	F
Corrected total	11	69.67		
Due to regression	1	52.50	52.50	
Residual	10	17.17	1.72	
Lack of fit	4	5.50	1.375	0.706 (not significant
Pure error	6	11.67	1.945	at $\alpha = 0.05$)

The model seems to be adequate.

G. 1. $\hat{Y} = 323.628 + 131.717X$.

ANOVA

Source	df	SS	MS	F
Corrected total	16	2,305,042		
Due to regression	1	1,099,641.1	1,099,641.10	13.68 significant
Residual	15	1,205,400.9	80,360.06	at $\alpha = 0.05$

2.

ANOVA

Source	df	SS	MS	F
Corrected total	16	2,305,042		
Due to regression	1	1,099,641.1		
Residual	15	1,205,400.9		
Lack of fit	1	520,648.6	104,129.72	1.52 not significant
Pure error	10	684,752.3	68,475.23	at $\alpha = 0.05$

A straight line relationship seems reasonable.

H. Prediction equation: $\hat{Y} = 1.222 + 0.723\ X$

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Corrected total	13	2.777			
Due to regression	1	1.251	1.251	9.850	4.75
Residual	12	1.526	0.127		

$9.850 > F(1, 12, 0.95) = 4.75$; $\therefore$ reject $H_0: \beta_1 = 0$ if no lack of fit.

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Corrected total	13	2.777			
Due to regression	1	1.251			
Residual	12	1.526			
Lack of fit	7	0.819	0.117	0.830	4.88
Pure error	5	0.707	0.141		

$0.830 < F(7, 5, 0.95) = 4.88$, $\therefore$ lack of fit is not significant.

Conclusion: Use the prediction equation

$$\text{Cup loss(\%)} = 1.222 + (0.723)[\text{bottle loss(\%)}].$$

I. Prediction equation: $\hat{Y} = 17.146 + 11.836X$.

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Corrected total	12	22,126.308			
Due to regression	1	6,034.379	6,034.379	4.125	4.840
Residual	11	16,091.929	1,462.903		

$4.125 < F(1, 11, 0.95) = 4.840$; $\therefore$ do not reject $H_0: \beta_1 = 0$. The regression is not significant.

$$R^2 = \frac{\text{SS due to regression}}{\text{Corrected total SS}} = \frac{6034.379}{22,126.308} = 27.27\%.$$

Conclusions: (*i*) The model is not useful. (*ii*) Further investigation of alternative variables will be necessary. (*iii*) Check pure error.

J. 1. $\hat{Y} = 2.5372000 - 0.004718X$

2.

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Corrected total	14	0.209333			
Due to regression	1	0.110395	0.110395	14.50	4.67
Residual	13	0.098938	0.007611		

$14.50 > F(1, 13, 0.95) = 4.67$; $\therefore$ reject $H_0: \beta_1 = 0$. The regression is significant, if there is no lack of fit.

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Residual	13	0.098938			
Lack of fit	5	0.018938	0.003788	0.38	3.69
Pure error	8	0.080000	0.010000		

$0.38 < F(5, 8, 0.95) = 3.69$; $\therefore$ Lack of fit not significant.

3. The 95% confidence interval on the true mean value of Y, calculated at four points: $X = 0$, $X = \overline{X}$, $X = 400$, $X = 460$:

At $X = 0$ $\quad \hat{Y} \pm (2.160)(0.527) = \hat{Y} \pm 1.138$
At $X = \overline{X}$ $\quad \hat{Y} \pm (2.160)(0.022) = \hat{Y} \pm 0.048$
At $X = 400$ $\quad \hat{Y} \pm (2.160)(0.039) = \hat{Y} \pm 0.084$
At $X = 460$ $\quad \hat{Y} \pm (2.160)(0.048) = \hat{Y} \pm 0.104$

K. 1. Plot the data and draw line by eye. The line drawn by eye here may well be somewhat different from that fitted by least squares later.

2. $\Sigma X_u = 1244.5 \qquad \Sigma Y_u = 30.458$
$\Sigma X_u^2 = 73{,}920.05 \qquad \Sigma Y_u^2 = 27.573638$
$\Sigma X_u Y_u = 1032.4865$

For later numerical work it is wise to keep all digits in the sums of squares calculations.

3. $b_1 = -0.00290351$
$b_0 = \overline{Y} - b_1 \overline{X} = 1.00210$

The line contains (for example) the points (0, 1.0021) and (100, 0.7117).

4. There do not appear to be any peculiarities severe enough to warrant corrective action. (There is a slight tendency for small residuals to be associated with small X's, however, and this might bear further investigation.)

5.

ANOVA

Source	df	SS	MS
Regression (b_0)	1	27.28500	
Regression ($b_1 \mid b_0$)	1	0.23914	0.23915
Residual	32	0.04950	$s^2 = 0.001547$
Total	34	27.57364	

6. $\text{se}(b_1) = s/\{\Sigma X_u^2 - (\Sigma X_u)^2/n\}^{1/2} = 0.00023$,
$\text{se}(b_0) = s/[\Sigma X_u^2/\{n\Sigma X_u^2 - (\Sigma X_u)^2\}]^{1/2} = 0.01089$.

7. $\text{se}(\hat{Y}_0) = s\{1/n + (X_0 - \overline{X})^2/(\Sigma X_u^2 - (\Sigma X_u)^2/n)\}^{1/2}$.
The formula for any particular X_0 is obtained by substituting for the known quantities (except X_0). Then the 95% confidence band for the true mean value of Y at X_0 is given by

$$\hat{Y} \pm t(32, 0.975)\text{se}(\hat{Y}_0).$$

For $t(32, 0.975)$ we can use $t(30, 0.975) = 2.042$ or interpolate in the table. The plot of confidence limits will look something like Figure 3.1.

8. The F-test statistic for overall regression is $0.23915/0.001547 = 154.6$, as compared to $F(1, 30, 0.95) = 4.17$. We could interpolate for 32 df, but it is clearly not necessary here. We therefore reject the null hypothesis that $\beta_1 = 0$. $R^2 = 0.83$, so that 83% of the variation about the mean $\overline{Y}$ is explained by our linear regression.

L. We calculate the approximate pure error SS as 0.01678 with 10 df. So we have the table:

ANOVA

Source	df	SS	MS
Total (corrected)	33	0.2886	
Regression $(b_1\|b_0)$	1	0.23915	0.23915
Residual	32	0.04950	$s^2 = 0.001547$
Lack of fit	22	0.03272	$\text{MS}_L = 0.001487$
Pure error	10	0.01678	$s_e^2 = 0.001678$

To test lack of fit we calculate an F-statistic of $\text{MS}_L/s_e^2 = 0.8862$; clearly no lack of fit is indicated since $F(22, 10, 0.95) = 2.75$. We thus pool the lack of fit SS with the pure error SS to calculate s^2. *Conclusion:* The data appear to be adequately described by a straight line regression of Y upon X. One could use the fitted relationship for predicting the true mean value of Y for any particular X_0, and the confidence bands drawn previously give an indication of what the accuracy of such predictions would be, assuming the model to be correct.

M. The conclusions are that (1) R^2 can equal 1 if there are no repeat runs in the data, but (2) R^2 cannot achieve 1 if nonidentical repeat runs exist, because the model cannot explain the pure error sum of squares. These conclusions are true in general regression situations as the following algebra shows.

Let the observations be

$Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ at the first location in X-space
$Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ at the second location in X-space
$\vdots$
$Y_{k1}, Y_{k2}, \ldots, Y_{kn_k}$ at the kth location in X-space

$$R^2 = 1 - \frac{\sum_{r=1}^{k}\sum_{u=1}^{n_r}(Y_{ru} - \hat{Y}_{ru})^2}{\sum_{r=1}^{k}\sum_{u=1}^{n_r}(Y_{ru} - \overline{Y}_{ru})^2} = 1 - \frac{\text{Residual SS}}{\text{Total corrected SS}}.$$

Now $\hat{Y}_{ru} = \hat{Y}_r$, the same for each u. Let $\overline{Y}_r = \sum_{u=1}^{n_r} Y_{ru}/n_r$ be the mean response at the rth location. Then

$$\begin{aligned}\sum_{u=1}^{n_r}(Y_{ru} - \hat{Y}_{ru})^2 &= \sum_{u=1}^{n_r}\{(Y_{ru} - \overline{Y}_r) + (\overline{Y}_r - \hat{Y}_r)\}^2 \\ &= \sum_{u=1}^{n_r}(Y_{ru} - \overline{Y}_r)^2 + \sum_{u=1}^{n_r}(\overline{Y}_r - \hat{Y}_r)^2 \\ &\quad + 2(\overline{Y}_r - \hat{Y}_r)\sum_{u=1}^{n_r}(Y_{ru} - \overline{Y}_r) \\ &= \sum_{u=1}^{n_r}(Y_{ru} - \overline{Y}_r)^2 + n_r(\overline{Y}_r - \hat{Y}_r)^2,\end{aligned}$$

the last summation being zero.

Thus

$$R^2 = 1 - \frac{\sum_{r=1}^{k}\sum_{u=1}^{n_r}(Y_{ru} - \overline{Y}_r)^2 + \sum_{r=1}^{k} n_r(\overline{Y}_r - \hat{Y}_r)^2}{\sum_{r=1}^{k}\sum_{u=1}^{n_r}(Y_{ru} - \overline{Y})^2}.$$

It follows that R^2 can attain 1 if (i) each $Y_{ru} = \overline{Y}_r$ *and* (ii) $\overline{Y}_r = \hat{Y}_r$.

(i) is true, however, only if either there are no repeats, that is, all $n_r = 1$, or all repeats at a location are identical, for *all* locations.
(ii) implies the model fits all the means perfectly, which can happen, sometimes.

Thus, in general, when pure error exists, $R^2 < 1$.

N. Test for lack of fit before checking the regression. If there *is* lack of fit, the F-test for regression and the calculations for confidence intervals and so forth are not valid. We calculate pure error as follows:

X	Pure Error Contribution		df
10	$\frac{1}{2}[-2 - (-4)]^2$	$= 2$	1
20	$\frac{1}{2}[1 - 3]^2$	$= 2$	1
30	$\frac{1}{2}[2 - 5]^2$	$= 4.5$	1
40	$0^2 + 1^2 + 2^2 - 3^2/3$	$= 2$	2
50	$(-2)^2 + (-3)^2 + (-4)^2 - \frac{(-9)^2}{3}$	$= 2$	2
Pure error		$= 12.5$	7

Split-up of Residual SS

Source	df	SS	MS	F
Lack of fit	3	73.177	24.392	$F(3, 7) = 13.66$, significant lack of fit shown
Pure error	7	12.5	1.786	
Residual	10	85.677		

The model suffers from lack of fit; the next step is to plot residuals and check patterns, to see if the model can be improved.

O. The fitted equation is $\hat{Y} = 2.0464 + 0.1705X$.

ANOVA

Source	df	SS	MS	F
Regression $(b_1 \mid b_0)$	1	16.514	16.514	9.47, significant at 2.76% level
Residual	5	8.723	$s^2 = 1.745$	
Total, corrected	6	25.237		
$SS(b_0)$	1	272.813		
Total	7	298.050		

$$R^2 = \frac{16.514}{25.237} = 0.6544, \qquad \begin{aligned}\hat{Y}(0) &= 2.05,\\ \hat{Y}(100) &= 19.10,\end{aligned}$$

$$\text{Est } V\{\hat{Y}(X_0)\} = s^2\left\{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma (X_i - \bar{X})^2}\right\}$$

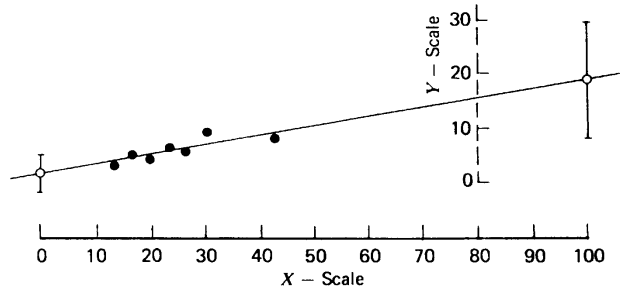
$$= 1.745\left\{\frac{1}{7} + \frac{(X_0 - 24.614286)^2}{568.168571}\right\};$$

Est $V(\hat{Y}(0)) = 1.745\{0.142857 + 1.066344\} = 2.110056 = (1.452603)^2$;
Est $V(\hat{Y}(100)) = 1.745\{0.142857 + 10.002324\} = 17.70334 = (4.207534)^2$.

The 95% confidence limits for true mean value of Y are

$$\hat{Y} = t_\nu(1 - \tfrac{1}{2}\alpha)\sqrt{\text{Est } V(\hat{Y})} = \hat{Y} \pm 2.571\sqrt{\text{Est } V(\hat{Y})}.$$

$X = 0$: $2.05 \pm 3.73 = -1.68$ to 5.78;
$X = 100$: $19.10 \pm 10.82 = 8.28$ to 29.92. (See plot below.)



Solution O

These are wide (compared, e.g., with range of Y, 6 units) even if model is correct at 0 and 100, which is uncertain for two reasons:

1. 0 and 100 are *well* outside data range
2. There is no reason to believe a linear relationship will be preserved at crucial places like 0 and 100 from practical considerations.

Conclusion: The predictions at $X = 0$ and $X = 100$ must be regarded with caution.

P. 1. $r^2_{XY} = \dfrac{\{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})\}^2}{\{\Sigma (X_i - \bar{X})^2\}\{\Sigma (Y_i - \bar{Y})^2\}} = \dfrac{\text{SS}(R|b_0)}{\Sigma (Y_i - \bar{Y})^2} = R^2.$

2. $r_{Y\hat{Y}} = \dfrac{\Sigma (\hat{Y}_i - \bar{\hat{Y}}_i)(Y_i - \bar{Y})}{\{\Sigma (Y_i - \bar{\hat{Y}}_i)^2\}^{1/2}\{\Sigma (Y_i - \bar{Y})^2\}^{1/2}}.$

Now $\hat{Y}_i = b_0 + b_1 X_i$, $\bar{\hat{Y}}_i = b_0 + b_1\bar{X} = \bar{Y}$. Thus we can substitute for $\hat{Y}_i - \bar{\hat{Y}}_i = b_1(X_1 - \bar{X})$, cancel out b_1 top and bottom, and we are left with r_{XY}.

Q. Solution depends on the data collected.

R. (Partial solution) The fitted equation is $\hat{A} = 14.410649 + 0.130768T$.

ANOVA

Source	df	SS	MS	F
Total	43	39,554.000	—	
b_0	1	38,221.488		
$b_1\|b_0$	1	764.646	764.646	55.2, significant at 1%
Residual	41	567.866	$s^2 = 13.851$	
Lack of fit	20	243.866	12.19	<1, not significant
Pure error	21	324.000	15.43	

$R^2 = 0.574$.

There appear to be no peculiarities in the residuals. There is a significant fit. Of the variation about the mean, 57.4% is explained out of a possible maximum of 100(1332.512 − 324)/1332.512 = 75.7%. (The remaining variation is due to pure error.) Thus a reasonably satisfactory fit has been obtained although there is room for improvement.

Residuals in same order, rounded to the nearest unit:

−2	−3
−2, −1	0, 1
−7, −2	2, −7
0, 3	−7, 0
−5, −1, 5	−2
−2	2, −1
3, 2, 6	6, −2
7, 4	5, −5, 2
1, −1, 1	−4
2, 7, 1	−2
6, −5, −2	0

S. 1. $\hat{X}_0 = 45.38$, $(X_L, X_U) = (39.56, 50.05)$.

2. $g = 0.074418$. For $g = 0$, $(X_L, X_U) = (39.83, 49.86)$, fairly close to the values in (1), but less conservative.

3. $\hat{X}_0 = 45.38$ [same as in (1)]. An additional 1 is inserted within the right brackets in Eq. (3.2.8) to provide end points $(X_L, X_U) = (19.33, 70.28)$, which are very widely spaced.

4. $g = 0.074418$ [same as in (2)]. For $g = 0$, $(X_L, X_U) = (21.23, 68.46)$. The large sizes of the two intervals concerned make these look reasonably close to those in (3), in spite of the large numerical differences.

T. The only way the confidence bands can look parallel is if the second term of $V(\hat{Y}_0)$ is small and, since $150 \leq X_0 \leq 170$, whatever $\overline{X}$ is, the top of the second term will be large for some X_0 values. So $S_{XX} = \Sigma (X_i - \overline{X})^2$ must be very large, and the answer is (3).

U. 1. $b_1^* = b_1(1 - q)$. In the ANOVA table, $SS(b_1^*|b_0) = SS_{XY}^*{}^2/S_{XX}^* = S_{XY}^2/S_{XX}$, so no change in the ANOVA, or F-tests. The fitted values and confidence calculations are changed, however. Replace X_i by $X_i/(1 - q)$ in all formulas.

2. There would be complicated changes throughout due to the fact that $(1 - q_i)$ cannot be pulled through the summations. Replace X_i by $X_i/(1 - q_i)$ in all formulas.

3. Replace $(1 - q_i)^2$ by $1 - 2q_i$ where it occurs. If $q_i = q$ pull all q factors through the summations.

V. We can fit the straight line $\hat{Y} = 9.930 - 0.010987X$. The analysis of variance table is as follows:

ANOVA

Source	df	SS	MS	F
b_0	1	4,230.1602		
$b_1\|b_0$	1	1.1777	1.1777	0.73
Residual	48	77.2321	1.6090	
Lack of fit	29	45.7771	1.5785	0.95
Pure error	19	31.4550	1.6555	
Total	50	4,308.5700		

$F(30, 18, 0.95) = 2.11$; $F(1, 48, 0.95) = 4.05$. There is no apparent lack of fit, and the regression slope is not significant.

Conclusion: These data do *not* confirm the idea that length of life is related to length of lifeline.

Notes: (*i*) The contributions to pure error at $X = 75$ and 82 are (comparatively) extremely large because of two extreme observations, 6.45 and 13.20, respectively. (*ii*) An improved

analysis would take account of the covariate "body length" or some similar covariate and would adjust the observations for this possible source of variation. It is anticipated that this would not alter our main conclusion.

W. Parts 1 and 2 lead you to think in terms of using *age* as a predictor, and to consider the possibility of transforming price and/or age, perhaps as in (3).

3. $\mathbf{Y} = (3.91, 3.56, \ldots, 1.61)'$.
$\mathbf{Z} = (82, 72, \ldots, 12)'$.
$\hat{Y} = 1.143181 + 0.0346564Z$.
$100\mathbf{e} = (-8, -8, 27, -8, 25, 13, -31, 8, -32, 22, 3, -23, 6, 5)'$.

ANOVA

Source	df	SS	MS	F
b_0	1	80.6880	—	
$b_1\|b_0$	1	6.4075	6.4075	157.4
Residual	12	0.4884	$s^2 = 0.0407$	
Total	14	87.5859		

The regression is highly significant and $R^2 = 0.9292$.

4. Note the difficulty in answering apparently straightforward questions when transformations have been used. The per year rate of increase of price is b_1 (price) and depends not only on b_1 but also price. Thus it is simpler to say that $Y = \ln$ (price) increases at rate b_1 per year.

5. Again, some care is needed. If we forecast Y (on the basis of the fitted equation) at $Z = 38$, we would have $\hat{Y}_{38} = 2.46$. However, in 1975 the actual $Y = \ln(20) = 3.00$, so that the prediction falls short by 0.54, a comparatively large amount. If a straight line relationship holds in 1975, it does not look like the same one, but it is difficult to tell on the basis of one new point whether the whole line has risen so that b_0 alone has increased (possible) or if b_1 alone has increased (unlikely, since younger port would then be cheaper!) or both b_0 and b_1 have increased (possible). More data are needed to investigate further.

X. See the quoted source for a full discussion of this exercise. For a way of generating such data, see "Computer generation of data sets for homework exercises in simple regression," by S. R. Searle and P. A. Firey, *The American Statistician*, **34,** February 1980, 51–54.

Y. Answer is implicit in the question. It is true generally, as we now show in matrix algebra (see Chapters 5 and 8).

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$
$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\epsilon},$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Thus $\mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}$. This is a useful result to know.

Z. **1.** Not shown.

2. $\hat{Y} = -10.9 + 18.449X$.

3. The residuals range from -1340.81 to 539.35 and add to 0.000

4.

Source	df	SS/10^6	MS/10^6	F
Regression $\|b_0$	1	22.01	22.01	97.68
Residual	15	3.38	0.225	
Total, corrected	16	25.39		

5. $R^2 = 22.01/25.39 = 0.8669$.

6. $s_e^2 = 584291.745/4 = 146072.9363$;

$F = \{2795543.255/11\}/14607.9363 = 1.74$
with df (4, 11). No lack of fit.

7. Yes, $F(1,15) = 97.68$, $p < 0.001$.

8. $\text{se}(b_0) = 141.7$, $\text{se}(b_1) = 1.867$.
Note that the nonsignificant $b_0 = -10.9$ is not practically reasonable (implying a negative consumption when the population is zero) and that refitting a straight line through the origin would make sense here. The result is $\hat{Y} = 18.365X$.

9.

	Limits of 95% Bands	
X	Lower	Upper
25	193	707
50	665	1158
100	1503	2165
150	2269	3244
200	3012	4346
250	3746	5456

10. The Swiss data point would not have much effect as it is down close to other small results. The U.S. Data point is at an extreme and changing it would have an effect. The point is *influential*; see Cook's statistic elsewhere.

11. The e versus $\hat{Y}$ plot has funnel-shaped characteristics, which imply more variation in larger Y-values. Consideration should be given to transforming the Y's; see *transformations*.

12. $\hat{X} = 6.401 + 0.046989Y$ or $Y = 136.20 + 21.281\hat{X}$;
$c_1 = (18.449/0.046989)^{1/2} = 19.815$.
The third line is $\hat{Y} - \overline{Y} = c_1(X - \overline{X})$ or $\hat{Y} - 804.93 = 19.815(X - 44.224)$, namely, $\hat{Y} = 71.357 + 19.815X$. This lies between the other two lines; all three lines intersect at $(\overline{X}, \overline{Y})$.

AA. 1. $\hat{Y} = -2.679 + 9.5X$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total	6	651.714		
Regression: $b_1 \mid b_0$	1	631.750	631.750	158.33
Residual	5	19.964	3.99	

Since $158.33 > 6.61$, the regression is significant, $\alpha = 0.05$.

3. No evidence to suggest a more complicated model needed.

BB. 1. $b_1 = 9.13$; $\hat{Y} = 9.13X$.

Obs. No.	X	Y	$\hat{Y}$	Residuals
1	3.5	24.4	31.955	−7.555
2	4.0	32.1	36.520	−4.420
3	4.5	37.1	41.085	−3.985
4	5.0	40.4	45.650	−5.250
5	5.5	43.3	50.215	−6.915
6	6.0	51.4	54.780	−3.380
7	6.5	61.9	59.345	2.555
8	7.0	66.1	63.910	2.190
9	7.5	77.2	68.475	8.725
10	8.0	79.2	73.040	6.160

3. The plot of the residuals against $\hat{Y}$ indicates the omission of the β_0 term in the model.
4. The model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is recommended, restricting its use to within the X_1 range, 3.5 to 8.0, shown by the data.

   If the true model really has $\beta_0 = 0$, then data will have to be obtained nearer to the zero response before any insight into the right model for the range of X from 0 to a large value of X can be obtained.

CC. 1. $\hat{Y} = -252.298 + 8.529X$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total	9	219270.5000		
$b_1 \mid b_0$	1	200772.3188		
Residual	8	18498.1812		
Lack of fit	4	18454.6812	4613.6703	424.2455
Pure error	4	43.5000	10.8750	

The lack of fit test indicates that the model is inadequate.

The plot of the residuals against $\hat{Y}$ indicates a definite trend from negative residuals to positive residuals as the value of $\hat{Y}$ increases.

There is also evidence of an outlier; namely, $Y = 415$ when $X = 90$. This point should be investigated further.

DD. 1. $\hat{Y} = 7.950 - 0.0179T$.

2.

ANOVA

Source of Variation	df	SS	MS	Calc. F	$F_{0.95}$
Total	8	6.260			
Regression	1	1.452	1.452	2.114	5.59
Residual	7	4.808	0.687		
Lack of fit	3	1.763	0.588	<1	
Pure error	4	3.045	0.761		

The regression is nonsignificant, $R^2 = 23.2\%$.

3. $s_{b_1} = 0.01228$. $-0.04689 \leq \beta_1 \leq 0.01119$.

4.

Batch No.	Y_i	$\hat{Y}_i$	$Y_i - \hat{Y}_i$
1	2.10	2.95	−0.85
2	3.00	3.49	−0.49
3	3.20	2.59	0.61
4	1.40	2.24	−0.84
5	2.60	2.42	0.18
6	3.90	2.95	0.95
7	1.30	2.24	−0.94
8	3.40	2.59	0.81
9	2.80	2.24	0.56

No discernible pattern in the residuals.

5. $0.193 \leq \hat{Y}_0 \leq 4.453$.
6. No. The slope of the line fitted is not significant and, in any case, 360 is well beyond the temperature range, making the use of the fitted equation even more dangerous.

EE. 1. $\hat{Y} = 44.353 + 2.6172X$.

3. −0.75, 0.73, 2.63, −1.63, 3.11, −6.41, −5.67, 6.54, 1.92, −1.86, 0.78, −7.53, 4.21, 0.95, 6.43, −7.10, 7.90, 9.38, −3.93, 0.07, −6.71, −6.71, 3.62. Their sum = −0.03.

4.

Source	df	SS	MS
b_0	1	107,170.00	—
$b_1 \mid b_0$	1	728.09	728.09
Residual	21	580.34	27.64
Lack of fit	18	459.84	25.55
Pure error	3	120.50	40.17
Total	23	108,478.43	

5. $R^2 = 728.09/1308.43 = 0.5565$.

6. $F = 25.55/40.17 = 0.64$. This value does not cause us to suspect the model, at least as far as this test is concerned.

7. As far as we can tell, yes. $F = 728.09/27.64 = 26.34$, which exceeds $F(1, 21, 0.95) = 4.32$ handily. We reject the idea that $\beta_1 = 0$.

8. Use s^2. We get $\text{se}(b_0) = \{s^2 \Sigma X_i^2/(nS_{XX})\}^{1/2} = 4.785$ and $\text{se}(b_1) = \{s^2/S_{XX}\}^{1/2} = 0.510$. The corresponding t-ratios are 9.27 and 5.13. [Note $5.13^2 = 26.32$, the F-value in (7) apart from rounding error.]

9. $\text{se}(\hat{Y}_0) = s\{1/23 + (X_0 - \overline{X})^2/S_{XX}\}^{1/2}$.

X_0	$\hat{Y}_0$	95% Limits	
6	60.1	56.0	64.1
7	62.7	59.5	65.9
8	65.3	62.7	67.9
9	67.9	65.6	70.2
10	70.5	68.1	73.0
11	73.1	70.1	76.2
12	75.8	72.0	79.6

10. The new trees lie well away from the rest of the data. The small tree is clearly of low height compared with trees of similar diameter. The large tree would lie close to the fitted line, as recorded. If damaged, it would indicate that the model might have to be rethought to curve upward if larger diameter, nondamaged trees were included in future data collection.

FF. The two individual lines are

$$\hat{Y} = 44.353 + 2.6172X,$$

$$\hat{X} = -5.378 + 0.21261Y,$$

or $$Y = 25.295 + 4.7034\hat{X};$$

$$c_1 = (2.6172/0.21261)^{1/2} = 3.5085.$$

Thus $c_0 = \overline{Y} - c_1\overline{X} = 68.261 - 3.5085(9.1348) = 36.211$.
The compromise line is thus

$$\hat{Y} = 36.211 + 3.5085X.$$

GG. Detroit has low turnover but also a low wage, so it seems likely that this is the (atypical) city with high unemployment. The fitted line is $\hat{Y} = 110.79 - 11.515X$. $R^2 = 0.825$, $s^2 = 65.50$ (2 df). When $X = 6$, $\hat{Y} = 41.7$ and $\text{se}(\hat{Y}) = 4.40$. The confidence interval is thus $41.7 \pm 4.303(4.40) = (22.8, 60.6)$. This is very wide, but not surprisingly so, since the fit is based on only four data points.

HH. $\hat{Y} = 1919.8 + 3.1376X$ is the fitted equation. If conditions remained the same, we could

use this as a predictive equation *within the current limits of the X-data.* When $X = 0$, $\hat{Y} = 1920$ and the requested limits are (1521, 2319). Does this mean that the Protestants will receive no money at all, or lose money (!), if Catholic attendance drops low enough? This seems most improbable! The model is unlikely to be valid below the data we have. For example, the true model could turn sharply downward, with X staying positive, below the present data. We are grateful to R. Peter Hypher for this interesting example.

II. The underlying idea is that there might be some relationship between how well the TA was rated and the course number (ranging from the elementary 201 to the most advanced 824). The fitted line $\hat{Y} = 3.6854 + 0.00140X$ explains only 16.3% of the variation even though the F is technically significant at about 2.2%. (See Chapter 11 for more discussion on this.) A plot of the data makes it obvious that observations 6, 7, and 11 are very influential in determining the slope of the line. This would show up in high values of Cook's statistic (see Chapter 8). These three observations are from courses that require highly trained teaching assistants. The omission of these three influential points actually improves the R^2 value to 21.0% with an F significant at about 1.2% from $\hat{Y} = 2.9895 + 0.003932X$. So overall we might claim some modest effect is apparent, probably because the higher the course number, the more experienced the teaching assistant needs to be. In the second fit, the two observations from course 424 now become influential. Omitting those gives $R^2 = 29.8\%$ and a significant F at about 0.3%, from $\hat{Y} = 2.4412 + 0.006071X$. So the effect seems fairly persistent. Note that spacing the data by an X-coordinate based on the course number may not be a good way to represent their relative difficulty or the requirements for their teaching assistants.

JJ. **1, 3.** Not provided.

2. $\hat{Y} = 76.01 + 1.27667X$.

4.

Source	df	SS	MS	F
Regression $\|b_0$	1	6,895,054	6,895,054	289.52
Residual	46	1,095,495	23,815	
Total, corrected	47	7,990,548		

5. $R^2 = 0.8629$.

6.

Lack of fit SS	1,012,189.6	(29 df)	MS = 34,903.1
Pseudo pure error SS	83,305.4	(17 df)	MS = 4,900.3

$F = 7.12$; $F(29, 17, 0.95) = 2.16$.

Significant lack of fit.

If only the exact repeats are used, the pure error MS = 6893/2 = 3446.5 (2 df).

$F = \{1{,}092{,}048.5/44\}/3{,}446.5 = 7.20$, while $F(44, 2, 0.95) = 99.47$, so this test is very insensitive with only 2 df for pure error. One would be inclined to believe that there is lack of fit, relying on the pseudo test above.

Examination of the data plot shows that the first observation is very influential, and that the impression of a curved plot comes mostly from it.

10. Apart from two observations that stand out, a reasonable broad band is seen.

11. The analysis is little altered by these changes. Lack of fit persists in alternative analyses. The fitted straight line gives a reasonable impression of the data ($R^2 = 0.8629$) but there is a lot of variation unexplained at some of the lower data points by it.

KK. The original fit has a slope of 0.979 and an $R^2 = 0.985$. Removal of observation 9, which has the smallest X-value, the largest residual, and is the most influential (via Cook's D, see Section 8.3) leads to a regression with slope 1.013 and $R^2 = 0.995$. Now the second largest observation becomes most influential. Although eight of the papers suffered circulation losses, papers 2 and 9, which did not suffer losses, slightly distort the effects of these losses in the regression. The original fit is probably good enough for prediction, although

some analysts would argue the point. A fit through the origin is also not unreasonable in this problem; this fit has a slope of 0.990.

LL. The fitted equation is $\hat{Y} = 43.84 + 37.23X$.

ANOVA

Source	df	SS	MS	F
Total	13	857,500		
b_0	1	687,700		
$b_1 \mid b_0$	1	155,258	155,258	
Residual	11	14,542	1,322	
Lack of fit	6	13,075	2,179	7.44*
Pure error	5	1,467	293	

* Significant lack of fit at the 5% level.

Residuals in order are: −31, −8, −28, −6, −16, 24, −3, 43, 63, 36, 18, −36, −56.

Conclusions: A plot of the residuals in order, or a plot of the data, both show a clear "quadratic curve" pattern and, as we have already noted, there is significant lack of fit. We need to improve the model. One way would be to fit a quadratic curve in X to the data. If we do this we obtain the fitted equation

$$\hat{Y} = -49.05 + 83.18X - 4.07X_1^2.$$

Further analysis shows that no lack of fit is indicated and the overall regression is highly significant with $F(2, 10) = 237.4$. This model explains $100R^2 = 97.94\%$ of the variation about the mean. For an alternative analysis in which additional information is used, see Exercise H in "Exercises for Chapters 5 and 6."

MM.

$$b_0 = -1.752, \quad b_1 = 0.908$$

$$a_0 = -1.58, \quad a_1 = 1.097$$

$$Y = 89.19 + 0.910(X - 96.25)$$

All four sets of residuals $(Y_i - \hat{Y}_i, X_i - \hat{X}_i)$, and both set of residuals from a gmfr line sum to zero. All four plots of residuals versus fitted values suggest a nonconstant variance structure.

Chapter 4

A. **1.** False. Matrices are of different sizes.
2. False. **A** is 3 × 2, **C** is 3 × 3.
3. True.
4. True.
5. False. Only square matrices have inverses.
6. False. The right-hand side should be 2 × 2 not 3 × 3.

B. **1.** Impossible. **B** is 3 × 3 and **C** is 2 × 2.

2.
$$\mathbf{BB}' = \begin{bmatrix} -1 & 1 \\ 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 13 & 12 \\ -1 & 12 & 13 \end{bmatrix}.$$

3. Impossible. **A** is 3 × 3. **B'B** is 2 × 2.

4.
$$\mathbf{BC} = \begin{bmatrix} -1 & 1 \\ 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 17 & 12 \\ 18 & 13 \end{bmatrix}.$$

5. $\mathbf{AA}^{-1}\mathbf{BC} = \mathbf{IBC} = \mathbf{BC}$, given above.

6. $\mathbf{CB}' = (\mathbf{BC}')' = (\mathbf{BC})' = \begin{bmatrix} -1 & 17 & 18 \\ -1 & 12 & 13 \end{bmatrix}$
(or multiply it out).

7. Impossible. $\mathbf{C}$ is 2×2. $\mathbf{A}$ is 3×3.

8. $\begin{bmatrix} -1 & 1 \\ 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} 5 & -7 \\ 5 & -6 \\ 0 & 1 \end{bmatrix}$.

9. Partitioned matrix.
Invert middle to get $\frac{1}{4}$. Corners are $\mathbf{C}$, invert to give $\mathbf{C}^{-1}$. Put these together to give

$$\begin{bmatrix} -2 & 0 & 3 \\ 0 & \frac{1}{4} & 0 \\ 3 & 0 & -4 \end{bmatrix}.$$

10. $\mathbf{A}$ is symmetric so $\mathbf{A}' = \mathbf{A}$. Thus $\mathbf{A}'\mathbf{A}(\mathbf{A}')^{-1}\mathbf{A}^{-1} = \mathbf{AAA}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{I}_{3\times 3}$.

C. $\mathbf{A}$ is 2×3, $\mathbf{b}$ is 2×1, $\mathbf{C}$ is 2×2, $\mathbf{D}$ is 3×3.
So (1), (5), and (6) are all false. The matrices are not the right size for the designated operations. (3) is also false:

$$\mathbf{AD} = \begin{bmatrix} 1 & 4 & 1 \\ 0 & 8 & 1 \end{bmatrix}.$$

The (2, 1) element is not 6, as given. The remaining equations, (2), (4), (7), (8), and (9), are all true. However, in (7) one must check that $\mathbf{C}^{-1}$ does, in fact, exist.

D.
$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 14 & 790 \\ 790 & 49300 \end{bmatrix}^{-1} = \frac{1}{66100} \begin{bmatrix} 49300 & -790 \\ -790 & 14 \end{bmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 295 \\ 18030 \end{bmatrix}.$$

$$\mathbf{b} = \begin{bmatrix} 299800/66100 \\ 19370/66100 \end{bmatrix} = \begin{bmatrix} 4.535552 \\ 0.293041 \end{bmatrix}.$$

(Plot data, line, residuals.)

ANOVA

Source	df	SS	MS	F
Total	14	6641.00		
b_0	1	6216.07		
$b_1 \mid b_0$	1	405.45	405.45	250 significant
Residual	12	19.48	$s^2 = 1.62$	
Lack of fit	4	0.81	0.20	Not significant
Pure error	8	18.67	2.33	

$$\mathbf{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}s^2 = \begin{bmatrix} 1.208 & -0.019 \\ -0.019 & 0.000343 \end{bmatrix}$$

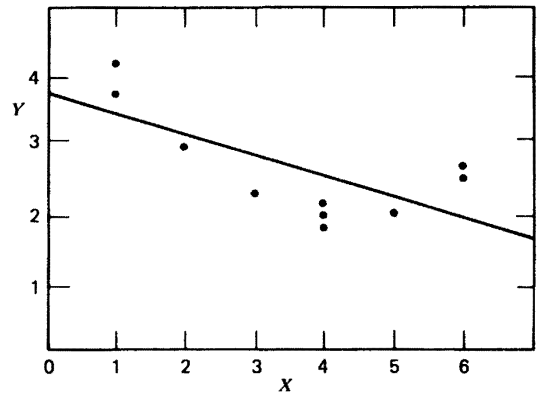
When $X = 65$,

$$V(\hat{Y}) = (1, 65) \begin{bmatrix} 1.208 & -0.019 \\ -0.019 & 0.000343 \end{bmatrix} \begin{bmatrix} 1 \\ 65 \end{bmatrix} = 0.1872;$$

$$\hat{Y}(65) = 23.583; \quad t(12, 0.975) = 2.179.$$

The 95% confidence interval for $E(Y \mid X = 65)$ is $23.583 \pm 2.179(0.1872)^{1/2} = 22.640$ to 24.526.

E. $\mathbf{b} = \begin{bmatrix} 10 & 36 \\ 36 & 160 \end{bmatrix}^{-1} \begin{bmatrix} 26.5 \\ 86.1 \end{bmatrix} = \frac{1}{304} \begin{bmatrix} 160 & -36 \\ -36 & 10 \end{bmatrix} \begin{bmatrix} 26.5 \\ 86.1 \end{bmatrix} = \begin{bmatrix} 3.75132 \\ -0.305921 \end{bmatrix}$.



Solution E

Y_i	4.2	3.8	3.0	2.3	1.8	2.0	2.2	2.0	2.5	2.7
$\hat{Y}_i$	3.45	3.45	3.14	2.83	2.53	2.53	2.53	2.22	1.92	1.92
$e_i = Y_i - \hat{Y}_i$	0.75	0.35	−0.14	−0.53	−0.73	−0.53	−0.33	−0.22	0.58	0.78

$\Sigma e_i = -0.02$

ANOVA

Source	df	SS	MS	F
b_0	1	70.225		
$b_1 \mid b_0$	1	2.845		
Lack of fit	4	2.740	$MS_L = 0.685$	15.222*
Pure error	4	0.180	$s_e^2 = 0.045$	
Total	10	75.990		

* We must test lack of fit first. If lack of fit exists, most other calculations (e.g., F-test for regression, confidence intervals, confidence bands) are not valid and should not be performed at all. Now, from the F-table, $F(4, 4, 0.95) = 6.39$. Thus there is significant lack of fit because $15.222 > 6.39$.

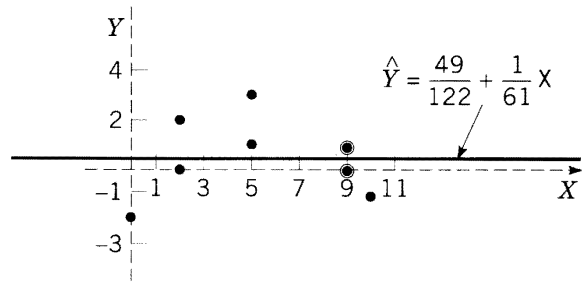
$R^2 = 2.845/5.765 = 0.4935$. The straight line explains only 49.35% of the variation about the mean, that is, not much.

We see, from the residuals, which have a clear positive–negative–positive pattern, that the data have a quadratic tendency, which a straight line model cannot follow. This is obvious from the plot, of course.

We conclude that the straight line model is inadequate and not usable and that a model $Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \epsilon$ should be tried next. (If we do this, we get $\hat{Y} = 5.462 - 1.6380X + 0.192840X^2$ with a good fit.)

F. 1. $\mathbf{b} = \begin{bmatrix} 10 & 60 \\ 60 & 482 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 32 \end{bmatrix} = \begin{bmatrix} 49/122 \\ 1/61 \end{bmatrix}$.

$$\hat{Y} = \frac{49}{122} + \frac{1}{61} X.$$



Solution F

3. Basic ANOVA:

$$\mathbf{b'X'Y} = \left[\frac{49}{122}, \frac{1}{61}\right]\begin{bmatrix}5\\32\end{bmatrix} = \frac{245+64}{122} = 2\tfrac{65}{122}.$$

ANOVA

Source	df	SS
b_0, b_1	2	$2\frac{65}{122}$
Residual	8	$18\frac{57}{122}$
Total	10	21

4.

$$SS(b_0) = (\Sigma Y)^2/n = 2\tfrac{1}{2}.$$

X	Pure Error (p.e.) Contribution at X	df
2	2	1
5	2	1
9	1	3
—	5 = Pure error SS	5 = p.e. df

ANOVA

Source	df	SS	MS
b_0	1	$2\frac{1}{2}$	
$b_1 \mid b_0$	1	$\frac{2}{61} = .033$	
Lack of fit	3	$13\frac{57}{122} = 13.467$	$4\frac{179}{366}$
Pure error	5	5	1
Total	10	21	

5. $F = \dfrac{MS_L}{s_e^2} = 4\frac{179}{366} < F(3, 5, 0.95) = 5.41.$

Thus lack of fit is not significant by this test. We recombine lack of fit SS and residual SS to give

$$s^2 = 18\tfrac{57}{122}/8 = 2\tfrac{301}{976} = 2.308525.$$

6. $F = \dfrac{2/61}{2\frac{301}{976}} < 1$. Obviously not significant. Yes, test is valid because lack of fit is not significant.

7. $V(\hat{Y}_0) = \left\{\dfrac{1}{n} + \dfrac{(X_0 - \bar{X})^2}{\Sigma (X_i - \bar{X})^2}\right\}\sigma^2 = \left\{\dfrac{1}{10} + \dfrac{(X_0 - 6)^2}{122}\right\}\sigma^2.$

When $X_0 = \sqrt{122} + 6$, this reduces to $1.1\sigma^2$, which is estimated by $1.1s^2 = (1.1)(2\frac{301}{906}) = (1.1)(2.308525) = 2.5393775$;

$$t(8, 0.975) = 2.306.$$

So the interval required is $\hat{Y}_0 \pm 2.306\sqrt{2.5393775}$

or $\hat{Y}_0 \pm 3.6747$.

Yes, it is valid because lack of fit is not significant. However, it is a long way outside the data and so is somewhat dangerous as we cannot guarantee the model out there.

8.

X_i	Y_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$
0	−2	0.40	−2.40
2	0	0.43	−0.43

X_i	Y_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$
2	2	0.43	1.57
5	1	0.48	0.52
5	3	0.48	2.52
9	1	0.54	0.46
9	0	0.54	−0.54
9	0	0.54	−0.54
9	1	0.54	0.46
10	−1	0.56	−1.56
			$0.06 = \Sigma e_i$

(Two decimal places is quite enough; even one decimal place would do.) The plot of e_i versus Y_i looks like the plot in (2) with different axes, of course. (Ask yourself why.)

9. $\mathbf{V(b)} = (\mathbf{X'X})^{-1}\sigma^2 = \begin{bmatrix} 10 & 60 \\ 60 & 482 \end{bmatrix}^{-1}\sigma^2 = \frac{\sigma^2}{1220}\begin{bmatrix} 482 & -60 \\ -60 & 10 \end{bmatrix}$

$$= \frac{\sigma^2}{610}\begin{bmatrix} 241 & -30 \\ -30 & 5 \end{bmatrix}.$$

10. $R^2 = \frac{2/61}{18\frac{1}{2}} = \frac{2}{61}\cdot\frac{2}{37} = \frac{4}{2257} \approx 0.18\%.$

11. Identical to R^2 for a straight line model.

12. No significant regression, no lack of fit. Nevertheless, the residuals pattern indicates a systematic departure from randomness that needs exploring further in spite of the nonsignificance of the lack of fit test. The model fitted is not of much value.

Chapters 5 and 6

A. 1. $b_0 = 14$, $b_1 = -2$, $b_2 = -\frac{1}{2}$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	10	190		
Due to regression	2	122	61.0	7.17*
Residual	8	68	8.5	

3. Test of significance:

$$\text{Compare } F = \frac{\text{MS regression}}{\text{MS residual}} \text{ with } F(2, 8, 0.95) = 4.46.$$

Since 7.17 is greater than the critical F, we reject the hypothesis of no planar fit and use the fitted equation

$$\hat{Y} = 14 - 2X_1 - \tfrac{1}{2}X_2.$$

4. $R^2 = \frac{122}{190} = 64.21\%.$

5a. Estimated variance of $b_1 = 1.4365$.

b. Estimated variance of $b_2 = 0.3587$.

c. Estimated variance of $\hat{Y} = 1.95075$.

6.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	10	190.00		
Regression	2	122.00	61.00	7.18
Due to b_1	1	116.08	116.08	13.64*
b_2 given b_1	1	5.92	5.92	<1
Residual	8	68.00	8.50	

7.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	10	190.00		
Regression	2	122.00	61.00	7.18
Due to b_2	1	98.33	98.33	11.57*
b_1 given b_2	1	23.67	23.67	2.78
Residual	8	68.00	8.50	

8. *Conclusions:*

(i) While the regression equation, $\hat{Y} = 14 - 2X_1 - \frac{1}{2}X_2$, is statistically significant, the mean square error is larger than that obtained when $\hat{Y} = 9.162 - 1.027X_1$ is used.

(ii) Independent estimates of β_1 and β_2 are not obtainable from these data. If independent estimates of β_1 and β_2 are desired, a balanced experiment in X_1 and X_2 should be done.

(iii) When problems arise as to a choice of a model, more experimental work of a balanced nature is usually necessary.

B.

$$\hat{X} = 1.0607 + 0.0056Y - 0.0013Z.$$

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	11	0.294867		
Regression	2	0.236409	0.118204	18.20
Due to Y	1	0.236275	0.236275	36.38*
Due to $Z \mid Y$	1	0.000134	0.000134	<1
Due to Z	1	0.236006	0.236006	36.34*
Due to $Y \mid Z$	1	0.000403	0.000403	<1
Residual	9	0.058458	0.006495	

Conclusion: The inclusion of both Y and Z in the model is not useful. This is further demonstrated by the correlation coefficient, $r_{yz} = -0.9978$.

C.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	14	85386.0000		
Regression	2	49791.1751	24895.58755	8.39*
Due to X_1 (alone)	1	48186.1482	48186.1482	16.24*
Due to $X_2 \mid X_1$	1	1605.0249	1605.0269	<1
Residual	12	35594.8249	2966.2354	

Conclusions

(i) The predictive model

$$\hat{Y} = 124.063977 + 3.512038X_1 + 0.834632X_2$$

explains only 58.31% of the total corrected variability in G.C.E. examination scores. While it proves to be statistically significant for an α-risk of 0.011, the standard deviation of the residuals is 54.46 and expressed as a percentage of the mean exam score, 9.725%. This shows that there is a great deal of unexplained variation, and thus the equation will not be very useful for prediction.

(ii) The addition of X_2, the previous performance in S.C. English Language, adds little to the predictability of a candidate's total mark in the G.C.E. examination. A simple model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ would do almost as well.

D. 1. $\hat{Y} = -94.552026 + 2.801551X_1 + 1.072683X_2$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	7	2662.14		
Regression	2	2618.98	1309.49	151.74*
Residual	5	43.16	8.63	

Since $F(2, 5, 0.95) = 5.79$, the overall regression is statistically significant, that is, $151.74 > 5.79$.

3. $R^2 = \dfrac{\text{SS Regression}}{\text{Corr. Tot.}} = \dfrac{2618.98}{2662.14} = 98.38\%$.

E. 1. $\hat{Y} = 67.234527 + 0.906089(X_1 - 164) - 0.064122(X_2 - 213)$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	15	8429.14444		
Regression	2	6796.77105	3398.385525	26.90*
Residual	13	1632.37339	126.336415	

Multiple $R^2 = 80.5\%$.
Standard deviation of residuals = 11.239947.
The fitted model is statistically significant. However, 20% of the variability remains unexplained; more work needs to be done on this problem.

3.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	15	8429.14444		
Regression	2	6796.77105	3398.385525	26.90*
X_1	1	6777.72877	6777.72877	53.65*
$X_2 \mid X_1$	1	19.04228	19.04228	NS
X_2	1	25.10057	25.10057	NS
$X_1 \mid X_2$	1	6771.67048	6771.67048	53.60*
Residual	13	1632.37339	126.33645	

X_1 is the more important variable.

4. *Conclusions*

(i) In the region of this machine's operation as indicated by the levels of Plate Clearance and Plate Temperature, the Plate Clearance has a pronounced effect on the percent properly sealed.

(ii) There is little evidence present that the Plate Temperature has an additive effect on the percent properly sealed.

There are some indications in the data that a different model should be fitted. It is helpful to examine the observations after they have been rearranged into the format below.

		Sealer Plate Temperature		
		176–208	210–220	225–240
Sealer Plate Clearance	130–148	35 42.5 43.5	34.5	56.7 51.7
	156–178	81.7 94.3	44.3 91.4	52.7
	186–194	82.0	98.3 83.3	95.4 84.4

One can see a definite interaction occurring between the clearance and the temperature. Thus a second-order model would be more appropriate.

F. 1. $\hat{Y} = 72.25 + 0.0286X_1 + 0.0487X_2$.

2.

Source	df	SS	MS	F
$b_1, b_2 \mid b_0$	2	35.0	17.5	<1
$b_1 \mid b_0$	1	1.7	1.7	<1
$b_2 \mid b_0, b_1$	1	33.3	33.3	<1
$b_2 \mid b_0$	1	34.8	34.8	<1
$b_1 \mid b_0, b_2$	1	0.6	0.6	<1
Residual	14	1509.1	$s^2 = 107.8$	
Lack of fit	5	898.6	179.7	2.65
Pure error	9	610.5	67.8	
Total, corrected	16			

There is no apparent lack of fit.

3. Residuals are 15, −4, 1, −11, −11, 6, 6, 3, −13, 5, −11, 0, 7, 21, 0, −9, −2. (Examination is left to the reader.)

4. Neither X_1 nor X_2 has any value in explaining Y on the basis of this data set. Together they explain only 100 $R^2 = 2.3\%$ of the variation about the mean.

G. For the model with both X_1 and X_2 in, we obtain: $\hat{Y} = 65.13 + 0.286X_1 + 0.487X_2$ with analysis of variance table as follows:

ANOVA

Source	df	SS	MS	F
b_0	1	79.973.88		
$b_1, b_2 \mid b_0$	2	35.01	17.51	0.16
Lack of fit	5	898.61	179.72	2.65
Pure error	9	610.50	67.83	
Total	17	81,518.00		

Because $F(5, 9, 0.95) = 3.48$, there is no reason to suspect lack of fit. Thus $s^2 = (898.61 + 610.50)/(5 + 9) = 107.79$. Test for overall regression: $F = 17.51/107.79 = 0.16$, not significant. The residuals show nothing worth remarking.

$$\text{SS}(b_1 \mid b_0, b_2) = 0.21; \quad \text{SS}(b_2 \mid b_0, b_1) = 33.32.$$

Both partial F-tests are nonsignificant. We conclude that the model $\hat{Y} = \bar{Y} = 68.588$ is as good as any.

Note: It is possible for the overall F-test for $H_0: \beta_1 = \beta_2 = 0$ to be nonsignificant, but the

partial F-test for one of the hypotheses $H_0 : \beta_j = 0, j = 1, 2$, to be significant. This would denote a weak relationship with the corresponding X_j.

H. The fitted equation is

$$\hat{Y} = -5.95 + 54.35X - 27.40Z.$$

Now $\mathbf{b}'\mathbf{X}'\mathbf{Y} = 21{,}205{,}018{,}600/24{,}780 = 855{,}731$.

This last figure is the $SS(b_0, b_1, b_2)$. Because of the pattern of Z's it will be found that the pure error sum of squares is exactly the same as before, 1467 with five degrees of freedom. (This is not true in general; usually the addition of a new variable, like Z here, will lead to fewer degrees of freedom for pure error, because responses Y_i with exactly the same X-values generally have different Z-values. In our data, such responses always have the same Z-values, which is uncommon.) The overall analysis of variance table is as follows:

ANOVA

Source	df	SS	MS	F
b_0	1	687,700		
$b_1, b_2 \mid b_0$	2	168,031	84,015	
Lack of fit	5	302	60	<1 NS*
Pure error	5	1,467	293	
Total	13	857,500		

* So $s^2 = (302 + 1{,}467)/(5 + 5) = 177$.

Test for overall regression: $F(2, 10) = 84{,}015/177 = 474.7$, which is very highly significant.

The extra sum of squares test for $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ is as follows:

$$\begin{aligned} SS(b_2 \mid b_1, b_0) &= SS(b_2, b_1 \mid b_0) - SS(b_1 \mid b_0) \\ &= 168{,}031 - 155{,}258 \\ &= 12{,}773. \end{aligned}$$

$F(1, 10) = MS(b_2 \mid b_1, b_0)/s^2 = 12{,}773/177 = 72.2$, which is very highly significant. Thus addition of Z as a predictor is extremely worthwhile.

Conclusions: The "number of men working" is an important variable. Because $b_2 = -27.40$, it has a negative effect. Note that this does *not* mean that the men do negative work but it does mean that their presence tends to cause less work to be done than might otherwise be anticipated. It would probably be a little more informative to fit the equation in the form

$$Y = \beta_0 + \beta_1(X - Z) + \beta_2 Z + \epsilon \tag{1}$$

so that the two predictors $(X - Z)$ and Z are the numbers of women and men working, respectively. Our original fitted equations in X and Z given above can be rewritten as

$$\hat{Y} = -5.95 + 54.35(X - Z) + 26.95Z,$$

which is what we would have obtained by fitting Eq. (1) by least squares directly. We see that, for the type of data we have here, the women do about twice as much work as the men! The indicated practical conclusion is to use women employees instead of men (or perhaps just different men) in the future, and see how that turns out.

I. 1. Fit model with both X_1 and X_2:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 68 & -67 \\ 0 & -67 & 68 \end{bmatrix}^{-1} \begin{bmatrix} 46 \\ -66 \\ 69 \end{bmatrix} = \begin{bmatrix} \frac{46}{7} \\ 1 \\ 2 \end{bmatrix}.$$

ANOVA

Source	df	SS	MS	F
b_0	1	302.29		
$b_1, b_2 \mid b_0$	2	72.00	36.00	83.72
Residual	4	1.71	$s^2 = 0.43$	
Total	7	376.00		

F is significant at 1% level so we reject $H_0 : \beta_1 = \beta_2 = 0$.

2. Fit model with X_1 alone:

$$\hat{Y} = 46/7 - (66/68)X_1,$$

$$\text{SS}(b_1 \mid b_0) = 64.06.$$

Thus $\text{SS}(b_2 \mid b_1, b_0) = 72 - 64.06 = 7.94$.
Test for $\beta_2 = 0$ (β_1 in model): $F = 7.94/0.43 = 18.53$; significant at 5% level but not at 1% because $F(1, 4, 0.99) = 21.20$.

3. Fit model with X_2 alone:

$$\hat{Y} = 46/7 + (69/68)X_2,$$

$$\text{SS}(b_2 \mid b_0) = 70.01.$$

Thus $\text{SS}(b_1 \mid b_2, b_0) = 72 - 70.01 = 1.99$.
Test for $\beta_1 = 0$ (β_2 in model): $F = 1.99/0.43 = 4.64$ not significant at 5% level, because $F(1, 4, 0.95) = 7.71$.

4. Implications: If X_2 is in, we don't need X_1.
If X_1 is in, X_2 helps out significantly.
Thus X_2 is clearly the more useful variable and alone explains $R^2 = 70.01/73.71 = 0.9498$ of the variation about the mean, whereas X_1 alone explains 0.8691, and X_1 and X_2 together explain 0.9768. Note that X_1 and X_2 are highly correlated in this data set.

J.

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 8 & -4 \\ 0 & -4 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 105 \\ -7 \\ 17 \end{bmatrix} = \begin{bmatrix} 10.50 \\ 0.25 \\ 2.25 \end{bmatrix}.$$

Source	df	SS	MS	F
b_0	1	1102.50		
$b_1, b_2 \mid b_0$	2	36.50		
Lack of fit	2	6	3.0	$F = 2.50$ (<5.79)
Pure error	5	6	1.2	
Total	10	1151		

No lack of fit; $s^2 = 1.714$

$$\text{Extra SS } F = \{(36.50 - 6.125)/1\}/1.74 = 17.72 \text{ (df 1,7)},$$

$$> 5.59, \text{ reject } H_0.$$

This F is the square of the corresponding t_7-statistic since the df are 1, 7.

Note that the extra SS F for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ is $F = \{(36.50 - 36.125)/1\}/1.714 = 0.219$, not significant. So X_1 could be dropped to give the equation $\hat{Y} = 10.5 + 2.125X_2$.

K. Add a column of 1's to the X_1, X_2 columns to get **X**.

1.
$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 50.3 \\ 9.9 \\ 5.7 \end{bmatrix}.$$

2.
$$\mathbf{b} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} 50.3 \\ 9.9 \\ 5.7 \end{bmatrix} = \begin{bmatrix} 10.06 \\ 2.10 \\ 0.75 \end{bmatrix}.$$

3. $\mathbf{b}'\mathbf{X}'\mathbf{Y} = 531.083$.

4. $s^2 = (531.19 - 531.083)/(5 - 3) = 0.107/2 = 0.0535$.

5. $\text{se}(b_0) = (0.2 \times 0.0535)^{1/2} = 0.103$,
$\text{se}(b_1) = (0.5 \times 0.0535)^{1/2} = 0.164$,
$\text{se}(b_2) = (1 \times 0.0535)^{1/2} = 0.231$.

6. $\hat{Y}_0 = (1, 0.5, 0)\mathbf{b} = 11.11$.

7. $\text{se}(\hat{Y}_0) = \{\mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 s^2\}^{1/2} = 0.132$.

8. Set $\beta_2 = 0$, refit, get $\text{SS}(b_0, b_1) = 530.5205$; $\text{SS}(b_2|b_1, b_0) = 0.5625$.

9. The **V** matrix is diag(1, 1, 0.25, 1, 1) with $\mathbf{V}^{-1}$ = weight matrix = diag(1, 1, 4, 1, 1). The new estimates are $\mathbf{b}_W = (9.9625, 2.1, 0.75)$.

[*Note*: Here are some (X_1, X_2, Y) data for a similar exercise: (−2, −4, 22), (−1, −1, 19), (0, 0, 19), (1, 1, 24), (2, 4, 26).]

L. 1. The negative residuals occur at low levels of concentration and the positive residuals occur at high levels of concentration.

2. $\hat{Y} = 2.693374 - 0.277361X_1 + 0.365028X_2$.

3.

ANOVA

Source of Variation	df	SS	MS	Calc. F	$F_{0.95}$
Total (corrected)	8	6.2600			
Regression	2	4.7381	2.3690	9.34	5.15
$b_1\|b_0$	1	1.4521	1.4521	5.73	5.99
$b_2\|b_0, b_1$	1	3.2860	3.2860	12.96	5.99
Residual	6	1.5219	0.2536		

a. Since there are no replicates, the lack of fit test cannot be done.

b. A model $\hat{Y} = \overline{Y}$ explains 62.41/68.67 = 90.88% of the crude variation in the data measured from $Y = 0$. Of the remaining variation, the model $\hat{Y} = b_0 + b_1X_1 + b_2X_2$ explains 75.69% of it, or a total of 97.78% of the crude variation.

c. The addition of β_2 to the model improves the fit as shown by R^2 going from 23.30% to 75.69%.

4. $R^2 = 75.69\%$.

5. $\text{se}(\tilde{b}_1) = 0.00795$, $\text{se}(\tilde{b}_2) = 0.10141$.

6.

Batch	Y	$\hat{Y}$	$Y - \hat{Y}$
1	2.100	2.518	−0.418
2	3.000	3.350	−0.350
3	3.200	2.693	0.507
4	1.400	1.774	−0.374
5	2.600	2.781	−0.181
6	3.900	3.248	0.652
7	1.300	1.044	0.256
8	3.400	3.058	0.342
9	2.800	3.234	−0.434

7. Var $\hat{Y}$(coded) = 0.044871.

M. 1. The estimated point (b_0, b_1) is

$$(b_0, b_1) = (13.623005, -0.079829).$$

2. The 90% confidence contour for (β_0, β_1) is given by $(\boldsymbol{\beta} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \mathbf{b}) \leq ps^2F(p, \nu, 1 - \alpha)$. Let

$$\begin{pmatrix}\gamma_0\\ \gamma_1\end{pmatrix} = \boldsymbol{\gamma} = \boldsymbol{\beta} - \mathbf{b} = \begin{pmatrix}\beta_0 - b_0\\ \beta_1 - b_1\end{pmatrix}.$$

This is equivalent to making a transformation to a new origin at **b**. So now

$$\boldsymbol{\gamma}'\mathbf{X}'\mathbf{X}\boldsymbol{\gamma} \leq ps^2F(p, \nu, 1 - \alpha). \tag{1}$$

$$\begin{aligned}\text{LHS} = \boldsymbol{\gamma}'\mathbf{X}'\mathbf{X}\boldsymbol{\gamma} &= (\gamma_0, \gamma_1)\begin{bmatrix} n & \Sigma_{i=1}^n X_i \\ \Sigma_{i=1}^n X_i & \Sigma_{i=1}^n X_i^2\end{bmatrix}\begin{pmatrix}\gamma_0\\ \gamma_1\end{pmatrix}\\ &= [n\gamma_0 + \gamma_1 \Sigma X_i, \gamma_0 \Sigma X_i + \gamma_1 \Sigma X_i^2]\begin{bmatrix}\gamma_0\\ \gamma_1\end{bmatrix}\\ &= (n\gamma_0 + \gamma_1 \Sigma X_i)\gamma_0 + (\gamma_0 \Sigma X_i + \gamma_1 \Sigma X_i^2)\gamma_1\\ &= n\gamma_0^2 + 2\gamma_0\gamma_1 \Sigma X_i + \gamma_1^2 \Sigma X_i^2.\end{aligned}$$

Then Eq. (1) is

$$n\gamma_0^2 + 2\gamma_0\gamma_1 \Sigma X_i + \gamma_1 \Sigma X_i^2 \leq ps^2F(p, \nu, 1 - \alpha) = c, \text{ say.}$$

The confidence region thus consists of the interior points of the ellipse whose boundary is

$$n\gamma_0^2 + 2\gamma_0\gamma_1 \Sigma X_i + \gamma_1^2 \Sigma X_i^2 = c. \tag{2}$$

in the (γ_0, γ_1) space.

To find points on the boundary, we select some value of β_0 and thus of γ_0, and solve the quadratic Eq. (2) for γ_1. There will be two roots (providing an upper and a lower point on the ellipse):

$$\gamma_{11} = \{-\gamma_0 \Sigma X_i - \sqrt{(\gamma_0 \Sigma X_i)^2 - (\Sigma X_i^2)(n\gamma_0^2 - c)}\}/\Sigma X_i^2,$$
$$\gamma_{12} = \{-\gamma_0 \Sigma X_i + \sqrt{(\gamma_0 \Sigma X_i)^2 - (\Sigma X_i^2)(n\gamma_0^2 - c)}\}/\Sigma X_i^2.$$

(Imaginary roots mean we have chosen a γ_0 value that lies outside the ellipse.) We then put $\beta_0 = \gamma_0 + b_0$, $\beta_{1j} = \gamma_{1j} + b_1$ to give rise to two points (β_0, β_{11}), (β_0, β_{12}) in the original (β_0, β_1) space. These are plotted, a new β_0 is chosen, and the whole cycle is repeated until the shape of the ellipse is clear and can be sketched in. We know $\Sigma X_i = 1315$, $\Sigma X_i^2 = 76{,}323.42$, $c = ps^2F(p, \nu, 1 - \alpha) = 2(0.7926)(2.55) = 4.04225$. Via a short Fortran program, we obtain the following output.

β_0	β_{11}	β_{12}
12.4	−0.0614	−0.0561
12.5	−0.0643	−0.0567
12.6	−0.0668	−0.0576
12.7	−0.0691	−0.0587
12.8	−0.0713	−0.0600
12.9	−0.0734	−0.0613
13.0	−0.0755	−0.0627
13.1	−0.0775	−0.0641
13.2	−0.0794	−0.0657

β_0	β_{11}	β_{12}
13.3	−0.0813	−0.0672
13.4	−0.0832	−0.0688
13.5	−0.0850	−0.0705
13.6	−0.0867	−0.0722
13.7	−0.0884	−0.0739
13.8	−0.0901	−0.0757
13.9	−0.0917	−0.0775
14.0	−0.0933	−0.0794
14.1	−0.0948	−0.0813
14.2	−0.0963	−0.0832
14.3	−0.0977	−0.0853
14.4	−0.0991	−0.0873
14.5	−0.1004	−0.0895
14.6	−0.1015	−0.0918
14.7	−0.1025	−0.0942
14.8	−0.1033	−0.0969
14.9	−0.1035	−0.1001

3. The 95% confidence limits for β_1 are

$$b_1 \pm t(23, 0.975)s/\{\Sigma (X_i - \overline{X})^2\}^{1/2}$$

or $-0.0798 \pm (2.069)(0.0105)$, providing the interval $-0.1015 \leq \beta_1 \leq -0.0581$.
The 95% confidence limits for β_0 are

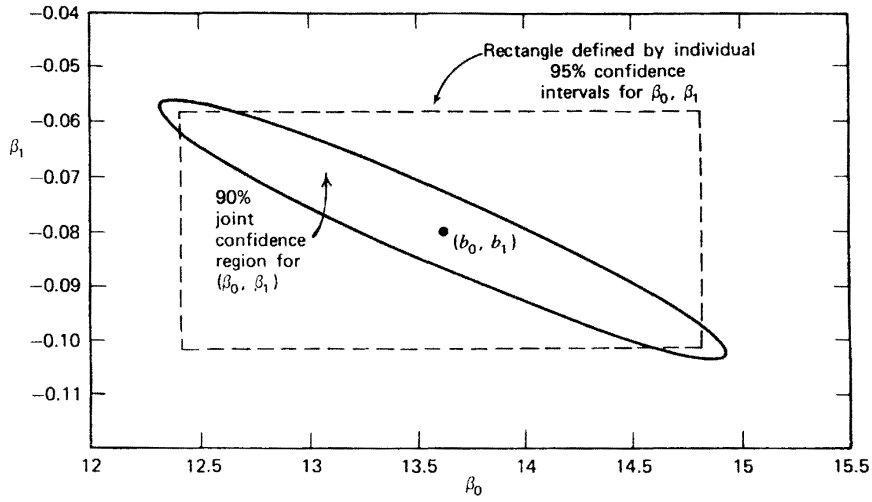
$$b_0 \pm t(23, 0.975)s/\{\Sigma X_i^2/n \Sigma (X_i - \overline{X})^2\}^{1/2},$$

or $13.623 \pm (2.069)(0.5814)$, providing the interval $12.420 \leq \beta_0 \leq 14.826$.
The rectangle produced by these two intervals is shown in the figure (Solution M).

Comments

(i) The joint 90% confidence region for the parameters β_0 and β_1, is shown as the long thin ellipse and encloses values (β_0, β_1), which the data regard as jointly reasonable for the parameters.

(ii) If we interpret the 95% confidence intervals of β_0 and β_1 simultaneously (and wrongly) as a "joint 90.25% confidence region" (rectangle), it is clear that we shall be led astray to some extent in this example. See Section 5.5.



Solution M

N. We have that $\hat{Y}_i - \overline{\hat{Y}}_i = \hat{Y}_i - \overline{Y}$

$$Y_i - \overline{Y} = (\hat{Y}_i - \overline{Y}) + e_i.$$

The numerator of $r_{Y\hat{Y}}$ is the sum of cross-products of these two and it reduces to $\Sigma(\hat{Y}_i - \overline{Y})^2$, the other terms vanishing due to the facts that (a) the residuals are orthogonal to the $\hat{Y}_i$'s and (b) $\Sigma e_i = 0$. [To prove (a): $\hat{\mathbf{Y}}'\mathbf{e} = (\mathbf{HY})'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{Y}'\mathbf{H}'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{0}$ because $\mathbf{H}$ is symmetric ($\mathbf{H}' = \mathbf{H}$) and idempotent ($\mathbf{H} = \mathbf{H}^2$).] The square root of the numerator now cancels out part of the denominator, and what is left is $\mathbf{H}$.

O. The vector on the right-hand side of the appropriate normal equations consists of the three elements $\Sigma e_i = 0$, $\Sigma e_i \hat{Y}_i = 0$, and $T_{12} = \Sigma e_i \hat{Y}_i^2$. Thus all three estimated coefficients depend on T_{12}, which is thus a measure of the amount of quadratic trend in the e_i versus $\hat{Y}_i$ plot.

P. $$\text{SS(regression)} = \mathbf{b}'\mathbf{X}'\mathbf{Y} = ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{H}\mathbf{Y}.$$

Remember $\mathbf{H} = \mathbf{H}^2 = \mathbf{H}^3 = \cdots = \mathbf{H}^m \cdots$. Also,

$$\hat{\mathbf{Y}} = \mathbf{Xb} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{HY}.$$

So $\hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \hat{\mathbf{Y}}\mathbf{H}'\mathbf{HY} = \hat{\mathbf{Y}}\mathbf{HHY} = \hat{\mathbf{Y}}\mathbf{HY}$ = SS(regression).
$\hat{\mathbf{Y}}'\mathbf{H}^3\mathbf{Y} = \hat{\mathbf{Y}}'\mathbf{HY} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$.

Q. $$\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} = (\mathbf{X}' - \mathbf{X}')\mathbf{Y} = 0$$

R. Write $\mathbf{X}_i'$ for the ith row of $\mathbf{X}$. Then

$$\begin{aligned}\sum_{i=1}^{n} V(\hat{Y}_i) &= \sum_{i=1}^{n} \mathbf{X}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i\sigma^2 \\ &= \text{trace}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\sigma^2 \qquad (1) \\ &= \text{trace}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\}\sigma^2 \\ &= p\sigma^2\end{aligned}$$

and we divide both sides by n. For step (1) see Appendix 6A.

S. This can be proved algebraically or demonstrated on the computer using one set of errors and different β's.

T. Follows directly from the hint.

U. Follows directly as indicated.

V. We get $\hat{Y} = 6.059 - 30.22U + 34.39U^2$, where $U = (X - 0.048)/0.048$.
Note that, in fitting this model, we have coded the X's to U's. Some programs will not carry out the calculation properly if some sort of coding is not done, because the correlation between X and X^2 for these data is high, 0.990, something that typically happens when polynomial terms are combined in a model and when orthogonal polynomials are not used. The minimum $\hat{Y} = -0.58$ occurs when $U = -(-30.22)/\{2(34.39)\} = 0.4394$, that is, when $X = 0.048 + 0.048(0.4394) = 0.069$. Of course, the value of the minimum is senseless, being negative, but we might hope that, in fact, a positive or zero minimum might really lie around $X = 0.069$. The data set is too small. Nevertheless, it provides a nice example of what might be achievable if one had a lot of good quality data.

W. Solutions in Appendix 7B.

X. **1.** $\Sigma (e_i - \bar{e})(Y_i - \bar{Y}) = \Sigma e_i (Y_i - \bar{Y})$ (because $\bar{e} = 0$ if a β_0 term is in the model)

$= \Sigma e_i Y_i$ ($\bar{e} = 0$)

$= \mathbf{e}'\mathbf{Y}$

$= \mathbf{e}'\mathbf{e}$ [because

$\mathbf{e}'\mathbf{e} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{Y}$

$= \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$

$= \mathbf{Y}'\mathbf{e}$

$=$ Residual SS $= \mathbf{e}'\mathbf{Y}$, where

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$\Sigma (e_i - \bar{e})^2 = \Sigma e_i^2 = \mathbf{e}'\mathbf{e}$

$\Sigma (Y_i - \bar{Y})^2 =$ Total corrected SS

$$r_{eY} = \frac{\mathbf{e}'\mathbf{e}}{\{(\mathbf{e}'\mathbf{e})\Sigma (Y_i - \bar{Y})^2\}^{1/2}} = \left\{\frac{\text{Residual SS}}{\text{Total corrected SS}}\right\}^{1/2} = \{1 - R^2\}^{1/2}.$$

2. Slope $= \mathbf{e}'\mathbf{e}/S_{YY} = (1 - R^2)$.

3. $\Sigma (e_i - \bar{e})(\hat{Y}_i - \bar{\hat{Y}}) = \Sigma e_i \hat{Y}_i$ (by similar reduction to that above)

$= \mathbf{e}'\hat{\mathbf{Y}}$

$= \mathbf{Y}'(\mathbf{I} - \mathbf{H})'\mathbf{H}\mathbf{Y}$ [because $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$]

$= \mathbf{Y}'(\mathbf{H} - \mathbf{H}^2)\mathbf{Y} = 0$ so that $r_{e\hat{Y}} = 0$ also.

Reference: Jackson and Lawton (1967).

Y. The basic reason for the computer's failure to provide estimates is that the experimental design and model used provide a singular $(\mathbf{X}'\mathbf{X})$ matrix, which cannot be inverted. Here, $X_1^4 = 10X_1^2 - 9$ and $X_2^4 = 10X_2^2 - 9$, for each row of $\mathbf{X}$.

Both models have the same singularity problems.

Z. **1.** $\hat{Y} = 22.561235 + 1.668017X - 0.067958X^2$.

2.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	18	204.481053		
Regression	2	201.994394	100.997197	649.8*
Residual	16	2.486659	0.155416	

The regression is statistically significant.

3. Test for lack of fit (breakdown of residual term in the above ANOVA table)

Source	df	SS	MS	F
Residual	16	2.486659		
Lack of fit	8	1.733325	0.216666	2.30
Pure error	8	0.753334	0.094168	

The lack of fit is nonsignificant since $2.30 < F(8, 8, 0.95) = 3.44$. Thus the quadratic model is sufficient for predictive purposes.

4. $\hat{Y} = 23.346374 + 1.045463X$.

ANOVA

Source of Variation	df	SS	MS	F
Total	18	204.481053		
Regression	1	195.242967	195.242967	359.29*
Residual	17	9.238086	0.543417	
Lack of fit	9	8.484752	0.942750	10.01*
Pure error	8	0.753334	0.094168	

The lack of fit is statistically significant since $10.01 > F(9, 8, 0.95) = 3.39$.

Residuals from the fitted equation also demonstrate the inadequacy of the model.

By plotting the residuals against the values of X, the predictor variable, one sees a definite curvature in the residuals which indicates the need for a second-order term in X.

5. *Conclusions:* The cloud point can be predicted by a second-order function of the percent $I - 8$ in the base stock,

$$\hat{Y} = 22.561235 + 1.668017X - 0.067958X^2.$$

There is no indication that a more complicated model is needed.

AA. First note that the total sum of squares is $4^2 + 2^2 + 1^2 + 5^2 = 46$ and this is the maximum sum of squares for all sets of linear functions. $SS(L_1, L_2) = \mathbf{z}'\mathbf{CY}$, where

$$\begin{bmatrix} 10 & -2 \\ -1 & 4 \end{bmatrix}\mathbf{z} = \begin{bmatrix} 15 \\ 6 \end{bmatrix} \quad \text{or} \quad \mathbf{z} = \begin{bmatrix} 2 \\ 2.5 \end{bmatrix} \quad \text{while } \mathbf{CY} = \begin{bmatrix} 15 \\ 6 \end{bmatrix}.$$

So $SS(L_1, L_2) = 30 + 15 = 45$.

$SS(L_1, L_2, L_3)$ can be done similarly, but we can also note that the vector for L_3, $(-3, 0, 0, 3)$, is orthogonal to the vectors for L_1 and L_2, so we can work this out separately as $(\mathbf{c}_3'\mathbf{Y})^2/\mathbf{c}_3'\mathbf{c}_3 = 3^2/18 = 0.5$ and add it on to get $SS(L_1, L_2, L_3) = 45.5$. Any fourth linearly independent linear function L_4 will add exactly 0.5 to this subtotal. Pick any L_4 and try it. An easy (orthogonal) choice is $L_4 = Y_2 - Y_3$.

Chapter 7

A. 1. The plots are not given here.

2. The histogram of residuals is somewhat skewed toward negative values, and a similar characteristic appears in the curved normal plot. The "residuals versus order" plot shows one comparatively low residual value (the 11th). There is also a drop (representing overprediction) in the last four residual values; this may have resulted from better control over the steam plant in the later part of the observed period. The lowest five negative residuals give rise to the skewness mentioned above. The "residuals versus $\hat{Y}$" plot is also affected by these five negative residuals. For that reason, the impression of a widening scatter is probably illusory.

3. $d = 1.39$, $k = 3$, $n = 25$. The d-value is inconclusive at the (two-tailed) 10%, 5%, and 2% levels and is close to being nonsignificant at the 2% level. Thus there is weak evidence of positive serial correlation, perhaps.

4. $n_1 = 10$, $n_2 = 15$, $r = 8$.

$\mu = 13$, $\sigma^2 = 5.5$.

$z = (8 - 13 + \frac{1}{2})/(5.5)^{1/2} = -1.919$. The two-tailed p-value is 0.055. As in (3), weak evidence (for a 5% person) of positive serial correlation is indicated.

B. For plots, see the source reference, especially p. 372–373 in which Figures 3, 4, and 5 of Watts and Bacon correspond, respectively, to the plots for our data sets 1, 2, and 3. The runs test results look like this:

Data Set:	1	2	3
r	9	22	13
n_1	19	29	16
n_2	28	17	30
n	47	46	46
z(lower)	−4.33	—	−2.76
z(upper)	—	—	—

Only the two z-values shown are relevant, the others being nowhere near their corresponding tail areas. There are pronounced indications of positive serial correlation in data sets 1 and 3. When this test is applied to unequally spaced data, inaccurate results can arise. Here, however, the long stretches of equally spaced data are reassuring.

C. See the source reference.

D. $\mu = 28.857$, $\sigma^2 = (3.688)^2$, $r = 38$.
$z = (38 - 28.857 - 0.5)/3.688 = 2.3435$. Yes.

E. $z = (5 - 13.48 + 0.5)/2.443 = -3.266$. Yes.

F. $d = 2.33$ is in the upper tail, so use $4 - d = 1.67$. The 2.5% lower tail bounds are $(d_L, d_U) = (1.51, 1.65)$. So $4 - d$ is not significant.

G. The Durbin–Watson statistic cannot exceed 1, a significant value for $n = 51$, $k = 5$. Advice: Conclude evidence of positive serial correlation on the basis of this test.

H. The Durbin–Watson statistic for these residuals is $d = 2225/834 = 2.67$, so $4 - d = 1.33$. This is at the upper end of the range $(d_L, d_U) = (1.16, 1.33)$ for $n = 24$, $k = 1$, in the 2.5% table. So it is not significant, and the answer is no.

Chapter 8

A. The matrix **H** should behave as described in Section 8.1. When outliers are in the data, the residuals plots may be slightly different.

B. The last point (6, 4.5) is very influential. With it, the line has positive slope; without it, negative. Some observations between $X = 3$ and $X = 6$ would be useful here.

C. A majority of Cook's statistics in a regression are typically small. You seek ones that are large *compared with a majority*. Some computer printouts indicate influential observations and you can see by experience how big the Cook's statistics needs to be to trigger that indicator. There is, however, no firm rule on how big the Cook's statistics should be for an observation to be influential, and the choice of a critical level is arbitrary. When you find an influential observation, try to see where it is in the X-space. Is it an isolated point? Does it have a large residual too? (An influential point *could* have a small or zero residual if it pulls the model to it, remember.) If the influential point is isolated can we "fill in" new data to bridge the gap to the main pattern of the data?

D. See source reference.

Chapter 9

A. H_0 is $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$, so that a suitable (nonunique)

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}.$$

B. The reduced model is $E(Y) = \beta_0 + 2\beta_2 X_1 + \beta_2 X_2 = \beta_0 + (2X_1 + X_2)\beta_2$. The fitted equation is $\hat{Y} = 10.06 + 0.980769(2X_1 + X_2)$ with residual sum of squares 0.162 (3 df). Thus $F = \{(0.162 - 0.107)/1\}/\{0.107/2\} = 1.03$. So the hypothesis is reasonable (not rejected).

C. For the full fit, the residual sum of squares is 47.8635 (8 df). The reduced model is $Y = \beta_0 +$

$\beta_1(X_1 + X_3) + \beta_2(X_2 + X_4) + \epsilon$ with fit $\hat{Y} = 274.1 - 1.106(X_1 + X_3) - 2.104(X_2 + X_4)$ and residual sum of squares 2510.1 (10 df). It follows that $F = \{(2510.1 - 47.8635)/(10 - 8)\}/\{47.8635/8\} = 205.77$ (2, 8 df). H_0 is rejected. The hypothesis pairs the two aluminates $(X_1 + X_3)$ and the two silicates $(X_2 + X_4)$. Both pairs are highly negatively correlated so H_0 looks reasonable at first sight, at least to a layperson.

D. $\hat{Y} = 1.7213 + 0.22434X$. Note that the slope has been much reduced compared with the ordinary least squares fit $\hat{Y} = 1.4364 + 0.3379X$ because of the smaller weight attached to the (possibly faulty) last observation.

E. $\hat{Y} = 1.8080 + 0.19750$. With even less weight on the largest observation, the slope is reduced even further. Down-weighting observations does not always change the fit. For example, an observation with a zero residual can be omitted (weighted zero) without affecting the fitted line.

F. It is obvious that this must be true, but it can be formally proved by applying Eq. (9.5.2) with $\mathbf{C} = (0, 1, 0)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, and $d = 1$. It then emerges, after reduction, that

$$\hat{\beta}_0 = b_0 + C_{01}(1 - b_1)/C_{11},$$
$$\hat{\beta}_1 = 1,$$
$$\hat{\beta}_2 = b_2 + C_{21}(1 - b_1)/C_{11},$$

where b_0, b_1, b_2 are the least squares estimates when there is no restriction on β_1 and $(\mathbf{X}'\mathbf{X})^{-1} = \{C_{ij}\}$ for $i, j = 0, 1, 2$. It can then be shown that the same estimates arise from the alternative fit. This requires some tedious algebra. It is easier to work a numerical example both ways on the computer.

G. See F solution first. Again obviously true, and again provable in a similar manner. Forcing a model through specific points can be unwise. (The comments that follow arise from a conversation with J. K. Little.)

Suppose we decide to fit a straight line $Y = \beta X + \epsilon$ through the origin when there are data points at $X = 0$. Because

$$b = \Sigma X_i Y_i / \Sigma X_i^2,$$

the points at $X = 0$ will not contribute to the regression slope estimate at all. Thus the decision to fit a line through the origin means we effectively *ignore the data at the origin*, an unwise move.

Another way of seeing this is that the sum of squares function

$$S(\beta) = \Sigma(Y_i - \beta X_i)^2$$

contains a constant portion $Y_1^2 + \cdots + Y_q^2$, say, for the q data values at $X = 0$, and so these data do not contribute to the fitting process.

A similar point arises when we fit a nonlinear model function that takes zero value at $\boldsymbol{\xi} = \mathbf{0}$ for any value of the parameter vector $\boldsymbol{\theta}$.

The overall moral is that, if we have data at the predictor variables origin, we should be *especially* cautious about fitting models forced *through* the origin, because the fitting process ignores important information. (In general, we advise that, for linear models at least, β_0 should *never* be omitted. One can always test whether β_0 is zero after the fit has been made.)

By extension of the idea above, any criterion such as "minimize

$$\Sigma W_i\{Y_i - f(\mathbf{X}_i, \boldsymbol{\theta})\}^\gamma,$$

where W_i are given weights, independent of $\boldsymbol{\theta}$, and where γ is a chosen constant" applied to data such that $f(\mathbf{X}_i, \boldsymbol{\theta})$ = constant for some i, independent of $\boldsymbol{\theta}$, will exhibit similar behavior. Forcing a regression equation through any point can thus create potential problems.

Chapter 10

A. We can go directly to part 2 and evaluate $E(\mathbf{b}) = \boldsymbol{\beta} + \mathbf{A}\boldsymbol{\beta}_2$ for both designs. We get

$$A = \frac{8}{8 + n_0}\begin{bmatrix} 1 & 1 & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where $c = -0.5$ for design A and $c = 0$ for design B. So

$$E(b_0) = \beta_0 + \frac{8}{8 + n_0}(\beta_{11} + \beta_{22} + c\beta_{12})$$

while b_1 and b_2 are unbiased. The symmetric design B is slightly better from this viewpoint. Adding center points reduces the bias directly, through the factor $8/(8 + n_0)$. Design B is also better in that all estimates are uncorrelated and have the same or smaller variances compared with design A. Evaluate $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ for both designs to see this. For design B, $V(b_0) = (8 + n_0)^{-1}\sigma^2$, $V(b_1) = \sigma^2/8 = V(b_2)$, whereas for design A, $V(b_0) = (8 + n_0)^{-1}\sigma^2$, $V(b_1) = \sigma^2/6 = V(b_2)$ and $\text{cov}(b_1, b_2) = \sigma^2/12$.

B. For design A,

$$\mathbf{A} = \begin{bmatrix} 8 + n_0 & 0 \\ 0 & 8 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ -4/(8 + n_0) \end{bmatrix}.$$

$$\text{So } E(b_0) = \beta_0, \text{ (unbiased)}, E(b_2) = \beta_2 - \frac{4\beta_1}{8 + n_0}.$$

For design B, $\mathbf{X}'\mathbf{X}_2 = \mathbf{0}$ so both b_0 and b_2 are unbiased. Again, the orthogonal design B is better from this viewpoint.

C.

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2 = \begin{bmatrix} 6 & 0 \\ 0 & 70 \end{bmatrix}^{-1}\begin{bmatrix} 70 \\ 0 \end{bmatrix} = \begin{bmatrix} 35/3 \\ 0 \end{bmatrix}.$$

$$E(b_0) = \beta_0 + 35\beta_2/3,$$

$$E(b_1) = \beta_1 \qquad \text{(unbiased)}.$$

D. We need

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -3 & -1 & 1 & 3 & 5 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

$$\mathbf{X}_2' = [25, 9, 1, 1, 9, 25], \qquad \boldsymbol{\beta}_2 = \beta_2,$$

$$n = 6, p = 2,$$

$$\mathbf{A} = \begin{bmatrix} 35/3 \\ 0 \end{bmatrix}.$$

We then find that

$$E\,(\text{MS due to } b_0) = \sigma^2 + (6\beta_0 + 70\beta_2)^2/6,$$

$$E\,(\text{MS due to } b_1|b_0) = \sigma^2 + 70(\beta_1^2 + 7\beta_2^2),$$

$$E\,(\text{MS due to lack of fit}) = \sigma^2 + 1792\beta_2^2/3.$$

Note that, if $\beta_2 = 0$, a test for $H_0: \beta_1 = 0$ becomes feasible as described in the text. Note also that only the first expectation contains β_0.

E. The key difference between this quadratic term and the previous one in Exercise C is that the current vector $\mathbf{X}_2$, which is $(5, -1, -4, -4, -1, 5)'$, is orthogonal to both columns of $\mathbf{X}$ so that $\mathbf{X}'\mathbf{X}_2 = 0$. This means that $\mathbf{A} = \mathbf{0}$ and so no biases exist. Note that if we rewrite the quadratic model function of Exercise C as

$$\{\beta_0 + 35\beta_2/3\} + \beta_1 X + (8\beta_2/3)\{0.375(X^2 - 35/3)\}$$

with all estimates now orthogonal since $\mathbf{X}'\mathbf{X}$ is diagonal, it becomes apparent that the estimate of the first coefficient is an estimate of $\beta_0 + 35\beta_2/3$, as given in Exercise C.

F. 1. Either express the lack of fit mean square as a quadratic form and apply the result in Section 10.3 directly, or evaluate

$$\begin{aligned} E(\mathbf{Y}'\mathbf{Y}) &= E(\mathbf{Y}'\mathbf{I}\mathbf{Y}) \\ &= E(\mathbf{Y}')E(\mathbf{Y}) + \text{trace}\,\mathbf{I}\sigma^2 \\ &= (\mathbf{X}\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\beta}_2)'(\mathbf{X}\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\beta}_2) + n\sigma^2 \end{aligned}$$

and then obtain E(lack of fit sum of squares) by difference, assuming the other results in the table, and remembering to multiply by the degrees of freedom where necessary. Remember that $(\mathbf{I} - \mathbf{H}) = (\mathbf{I} - \mathbf{H})^2$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and that $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$ and

$$(\mathbf{I} - \mathbf{H})\mathbf{X}_2\boldsymbol{\beta}_2 = (\mathbf{X}_2 - \mathbf{X}\mathbf{A})\boldsymbol{\beta}_2.$$

2. $E(\text{residual mean square}|\beta_2 = 0) = (f\sigma^2 + e\sigma^2)/(f + e) = \sigma^2$.

G.

$$\mathbf{A} = \begin{bmatrix} 5 & 83 \\ 83 & 2189 \end{bmatrix}^{-1} \begin{bmatrix} 0.5 \\ 5.0 \end{bmatrix} = \begin{bmatrix} 0.167530 \\ 0.004068 \end{bmatrix}.$$

b_0 is biased by 0.168β; b_1 is biased by 0.004β.

Chapter 11

A. Results depend on α and the degrees of freedom used. For example, if a plane $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ is fitted to 20 observations, $\nu_1 = 2$, $\nu_2 = 17$ and $F(\nu_1, \nu_2, 1 - 0.05) = 3.59$. Using Eq. (11E) gives $R^2 = 2(3.59)/(2(3.59) + 17) = 0.2969$. So a "just significant" regression explains only about 30% of the variation about the mean $\overline{Y}$.

B. Here $\nu_1 = 2$, $\nu_2 = 30$, $\alpha = 0.05$. $F(2, 30, 0.95) = 3.32$. So $R^2 = 2(3.32)/(6.64 + 30) = 0.1812$, quite low.

C. Because $n = 46$ and the number of parameters fitted is 6, then $\nu_1 = 5$, $\nu_2 = 40$ and we need an F such that, at minimum, $0.90 = 5F/(5F + 40)$, that is, $F = 72$. Note that, since there are 25 degrees of freedom for pure error, there will be a maximum possible value for R^2 in such a set of data. It may well lie below the desired 90%!

D. 1. There are $n = 50$ observation and $\nu_1 = 5$, $\nu_2 = 44$. The observed $F = 10F(5, 44, 0.95) = 24.3$ and $R^2 = 5(24.3)/\{5(24.3) + 44\} = 0.7341$.

2. There are $5(7) + 9 = 44$ degrees of freedom for pure error and these comprise the entire residual degrees of freedom. Thus no further increase in R^2 is possible.

Chapter 12

A. 1. The model contains ten parameters. Examination of the data reveals only eight different data sites. Thus it is impossible to fit the model as stated.

2. $s^2 = 4.325$ with 10 degrees of freedom, calculated from pure error.

B.

$$a_0 = 12.565663, \quad b_0 = 0.038437, \quad c_1 = -0.032454,$$
$$a_1 = -0.006327, \quad b_1 = -0.013571, \quad c_2 = 0.001248,$$
$$a_2 = -0.090698, \quad b_2 = 0.001376, \quad c_3 = 0.000198.$$

ANOVA

Source	df	SS	MS	F
Regression$\|b_0$	8	44.40797	5.55100	
Residual	11	0.03989	0.00363	
Lack of fit	6	0.03461	0.00577	5.46, significant at $\alpha = 0.05$ level
Pure error	5	0.00528	0.001056	
Total (corrected)	19	44.44785		

$R^2 = 44.40797/44.44785 = 0.9991$. This is an interesting set of data. Practically all the variation is explained by the model, but the lack of fit test is significant. This would lead one to first question whether the repeat runs are really *that* precise. If they are, the analysis says that there is still variation that is in excess of natural variation, which could be explained. (From a practical point of view, however, most of the variation *is* explained, so why not use the equation? The author did.) A plot of the residuals in the X-space shows up a possible X_1X_2 interaction and suggests the possible addition of terms $(\delta_0 + \delta_1 Z + \delta_{11}Z^2)X_1X_2$ to the model. If this is done, δ_{11} cannot be estimated separately because of column dependence in the **X** matrix and the least squares estimates are

$$a_0 = 11.918982, \quad b_0 = 0.049215,$$
$$a_1 = 0.057033, \quad b_1 = -0.014627,$$
$$a_2 = -0.090698, \quad b_{11} = 0.001376,$$
$$c_0 = -0.001660, \quad d_0 = 0.000513,$$
$$c_1 = -0.001769, \quad d_1 = -0.000050.$$
$$c_{11} = 0.000198,$$

The extra SS for d_0 and d_1 given the other estimates is 0.02183, which leads to an F of $5.43 > F(1, 9, 0.95) = 5.12$, just significant. The new lack of fit F is $3.03 < F(4, 5, 0.95) = 5.19$, not significant, and R^2 goes up very slightly to 0.9996. The regression F-value (given a_0) is 2214, highly significant.

C. $\widehat{\log Y} = -4.927 + 16.85\{1000/(T + 460)\}$.

ANOVA

Source	df	SS	MS	
b_0	1	238.731		
$b_1\|b_0$	1	1.785	1.785	
Lack of fit	2	0.064	0.032	1.24, not significant
Pure error	20	0.515	0.026	
Total	24	241.095		

Test for regression $F = 1.785/(0.579/22) = 67.87 > F(1, 22, 0.95) = 4.30$, very significant. $R^2 = 0.7552$.

The residuals show a declining scatter as T increases, an indication that the log Y transformation has not adequately removed the inhomogeneity of variance in the original set of data.

Further (or different) transformation and use of weighted least squares are two possibilities to consider.

The increasing slope of residuals when plotted against fitted values within each temperature is a spurious "effect" that is caused by the grouping. This essentially (for equal groups) forces the straight line to attempt to fit the group means, so inevitably the smaller observations *within* the group will here have lower residuals, and so on. The overall (ungrouped) plot shows only the funnel effect noted above.

D. The design is *not* rotatable. If it were, the axial points would be at distance 1.682 from the origin, not 1.2154.

Response Y_1: A second-order model is not needed. The plane $\hat{Y}_1 = 76.220 - 7.013X_1 - 3.324X_2 - 4.305X_3$ explains 76.8% of the variation about the mean. $\hat{Y}$ increases when all three X's are reduced.

Response Y_2: Again, a plane $\hat{Y}_2 = 63.573 - 10.133X_1 - 5.381X_2 - 6.009X_3$ explains a lot of the variation about the mean, 82.1%, and the quadratic curvature is nonsignificant. Reducing X's again increases the predicted response.

Response Y_3: There is not much variation in these data. A plane explains only 46.6% of the variation around the mean. Adding second-order terms increases this to 83.7% even though the second-order terms as a group are not significant, and nor are any of them marginally. The best second-order term is the one in X_1^2. Adding this we find that the equation $\hat{Y} = 96.626 + 0.1598X_1 - 0.3091X_2 - 0.3803X_3 + 0.5036X_1^2$ explains 64.3% of the variation about the mean. This may be a reasonable compromise using the original response variable. There is very little variation in the response data so a transformation would do little to help.

E. Use X_1 = (fat −12)/4, X_2 = (flour −20)/10, X_3 = (water −50)/4, and X_4 = (rpm − 130)/40 to translate the first point into (−1, −1, −1, −1), and so on. The axial points are at a distance 1 and the design is not rotatable because a distance 2 is needed to achieve that. A plane explains 64.5% ($R^2 = 0.645$) and the full quadratic 82.9%. Only b_2 and b_4 show large marginal t-values in the planar fit. Adding the quadratic terms weakens the first-order coefficients but draws attention to the $b_{24}X_2X_4$ term. A possible compromise is to fit $\hat{Y} = 552.44 - 37.78X_2 - 84.39X_4 + 2.30X_2^2 + 32.80X_4^2 + 34.63X_2X_4$, which explains $R^2 = 0.718$.

F. 1. The fitted surface is

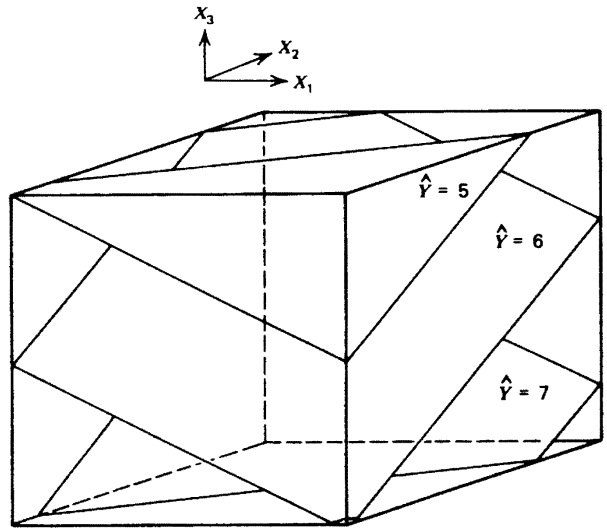
$$\begin{aligned}\hat{Y} = {} & 6.087 - 0.240X_1 + 0.340X_2 - 0.495X_3 \\ & -0.036X_1^2 + 0.021X_2^2 + 0.118X_3^2 \\ & + 0.040X_1X_2 - 0.070X_1X_3 + 0.045X_2X_3.\end{aligned}$$

2. The analysis of variance table takes the form:

ANOVA

Source	df	SS	MS	F
Mean (b_0)	1	758.173		
First order	3	5.671	1.890	
Second order|b_0	6	0.299	0.050	
Lack of fit	5	0.839	0.168	2.37
Pure error	5	0.354	0.071	
Total	20	765.336		

3. No lack of fit is shown and the second-order terms do not make a significant contribution to explaining the variation in the data when the mean square due to second order is compared to the residual mean square $s^2 = (0.839 + 0.354)/(5 + 5) = 0.119$, which has ten degrees of freedom.



Solution F.1

4. When a (reduced) first-order model is fitted to the data, the fitted equation takes the form

$$\hat{Y} = 6.157 - 0.240X_1 + 0.340X_2 - 0.495X_3,$$

there is no significant lack of fit, and the new estimate of σ^2 is $s^2 = 0.093$ based on 16 degrees of freedom.

Note: The fitted (planar) contours take the form illustrated in Solution F1. The cube shown has vertices $(\pm 2, \pm 2, \pm 2)$.

An alternative representation is shown in Solution F.2. The three parts of this figure show straight line contours for X_1 and X_2 in the three planes $X_3 = -1, 0$, and 1. The contours must be imagined as continuous between these three planes. Design points that fall on the planes are shown as dots. Two axial points that do not lie on any of the planes are not shown. Note that these contours cover a smaller region than that shown by the other figure.

A series of cross-sectional diagrams of this type is especially useful when second-order contours must be examined. For larger numbers of predictors than three, the methods of canonical reduction are especially useful.

We leave the examination of the residuals to the reader.

G. Yes.

H. The model chosen is

$$\hat{Y} = 120.627 + 490.412X_2 - 5.716X_3 - 1107.847X_2^2.$$

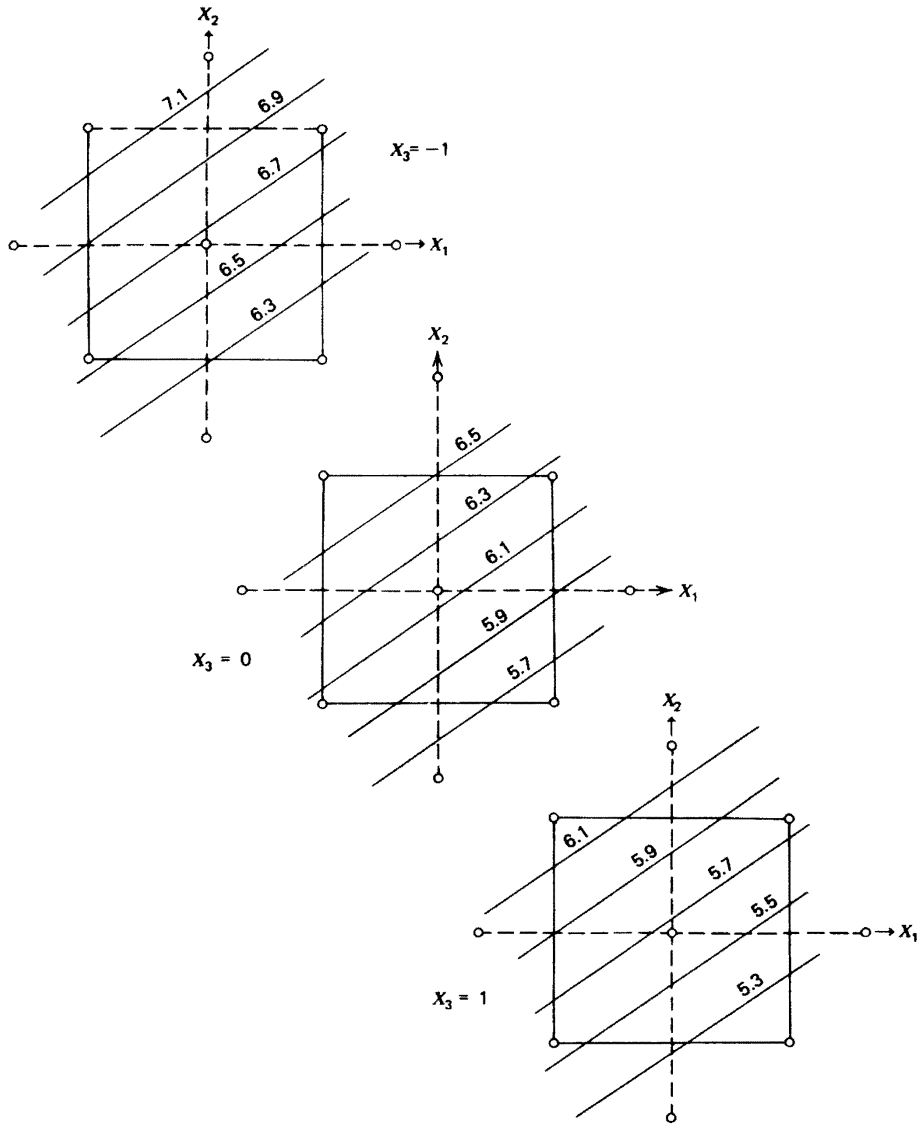
A plot of the residuals reveals runs of + and − signs indicating the presence of unconsidered X-variables. Adding second-order terms in X_2 and X_3 is of only marginal help. This equation has $R^2 = 90.27\%$, with a standard deviation of 6.2233.

The model in Anderson and Bancroft is

$$\begin{aligned}\hat{Y} - 84.204 = {} & 2.463(X_1 - 1.86) - 75.369(X_2 - 0.188) \\ & + 1.584(X_3 - 7.64) \\ & - 1.380(X_1X_2 - 0.3507).\end{aligned}$$

This model is not as good a fit. The residuals have a definite pattern and $R^2 = 75.49\%$.

Warning: The example in Anderson and Bancroft was used to illustrate regression calculations; there was no intention of building a best model.



Solution F.2

Chapter 13

A.

X	Y	$Z = \log_{10} Y$	$V = \log_{10} \log_{10} Y$
1830	30	1.4771	0.1694
1905	130	2.1139	0.3251
1930	400	2.6021	0.4153
1947	760	2.8808	0.4595
1952	1500	3.1761	0.5019
1969	25000	4.3979	0.6432

1. The plot of Y versus X shows all the "action" as having been in the most recent years and the plot is not particularly informative. The need for a transformation is apparent. In such a situation, proportional increases, determined by Z, are often more informative.

2. The plot of Z versus X is much easier to assess and is preferable to the plot of Y versus X.
3. $U \equiv V$ is not bad, although you may find a better one.
4. $\hat{V} = -5.4874 + 0.00307284X$.
   The residuals plots overall and against time are somewhat unsatisfactory but with only six residuals and four degrees of freedom between them, we cannot hope to say much more.
5. **ANOVA**

Source	df	SS	MS	F
$b_1 \mid b_0$	1	0.1185	0.11850	41.58
Residual	4	0.0114	0.00285	
Total (corrected)	5	0.1299		

   $41.58 > F(1, 4, 0.99) = 21.20$, significant at 1%. $R^2 = 0.9122$. Most of the variation in the data is explained by the model.

   Note that covariance $(b_0, b_1) = -0.999717$, very high. This is caused by the remoteness of the origin, and a reparameterization by taking a new origin closer to $\overline{X} = 1922.167$ would help. (See the closely related comments in Section 24.4.)
6. Use inverse interpolation by setting

$$\log\{\log(186{,}000 \times 3600)\} = -5.4874 + 0.00307284\hat{X}$$

$$\hat{X} = 2094.$$

7. This is a risky extrapolation of a small, difficult set of data under a tentative transformation that may well not hold up for new data. It depends on the trend being maintained for another 100 years or so. Don't rely on it!

B. In both solutions, values of $\lambda = -1(0.1)1$ were used. A finer grid would give greater accuracy. We leave detailed examination of the residuals to the reader.

First response (N990), $n = 23$. The best λ is around -0.6, where $S(\hat{\lambda}, \mathbf{V}) = 23.87$ is the residual sum of squares. A conservative (large) 95% confidence band based on $S(\lambda, \mathbf{V}) = 23.87 \exp(3.84/20) = 28.92$ stretches from -1.1 to -0.2, very roughly. The inverse square root $\lambda = -0.5$ might be a good compromise here, explaining 0.975 of the variation about the mean.

Second response (Silica B), $n = 24$. The best λ is around -0.3, where $S(\hat{\lambda}, \mathbf{V}) = 46.1$. A 95% conservative confidence band based on $S(\lambda, \mathbf{V}) = 46.1 \exp(3.84/21) = 67.06$ stretches from roughly -0.7 to 0.1 so that both -0.5 and 0 are possible choices. The corresponding R^2 values range from about 0.974 to 0.971 with better values in between. For $\hat{\lambda} = 0.3$, a very significant regression explains $R^2 = 0.981$ of the variation about the mean. The more appealing inverse square root value $\lambda = -0.5$ comes close with $R^2 = 0.979$.

If the same transformation were to be used for both responses, the inverse square root would be a sensible choice.

C. (Y_1 data)

λ	$L_{\max}(\lambda)$
−1.0	−22.9
−0.5	−11.0
−0.4	−9.5
−0.3	−8.0
−0.2	−6.5
−0.1	−5.3
0.0	−4.8
0.1	−5.3
0.2	−6.5
0.3	−8.1
0.4	−9.8
0.5	−11.5
1.0	−24.8

An appropriate 95% confidence interval is $-0.35 \leq \lambda \leq 0.32$. We select $\lambda = 0$, that is, use the transformation $W = \ln Y_1$. The fitted equation is $\hat{W} = 4.234 + 0.204X_1 + 0.098X_2 - 0.139X_3 - 0.070X_4$.

$R^2 = 0.9963$. All partial F-values for individual coefficients are highly significant. With only six degrees of freedom in eleven residuals, residual examination cannot be expected to be very revealing and it isn't.

(Y_2 data: solution not provided.)

D. The best value of λ is at about 2.1. The widest confidence interval based on $S(\lambda, \mathbf{V}) = 86.50 \exp(3.84/24) = 114$ is roughly $1.4 \leq \lambda \leq 3.0$. (Substituting the residual degrees of freedom 14 for $n = 24$ changes this to about 1.5 to 2.8, still quite wide.) Using $\lambda = 1$ permits a second-order fit that explains $R^2 = 0.906$ and using $\lambda = 2.1$ raises this to 0.940. This fact, combined with the wide confidence band for λ makes it clear that using a transformation is probably not worthwhile.

E. $\hat{\lambda} = 0.11$. The 95% confidence band is about $-0.16 \leq \lambda \leq 0.39$, so we can take $\lambda = 0$, that is, use $W = \ln M$. Then,

$\hat{W} = -5.728 + 2.031 \ln T$,

$R^2 = 0.8056$,

$SS(b_1|b_0) = 16.240$ (1 df),

Residual SS = 3.919 (41 df),

Total, corrected = 20.159 (42 df),

$F = 169.9 > F(1, 41, 0.99) = 4.08$, very significant regression.

Plots of the residuals versus T and $\hat{Y}$ show nothing unusual. The Durbin–Watson $d = 1.94$ obtained from the listed order is not significant.

F. $\widehat{\log Y} = 1.9929 + 0.5428 \log X_1 + 0.2740 \log X_2$.

Only 37.52% of the variation is explained by this model. The regression is not significant, the overall F-value being $(6.49/2)/(12.03/13) = 3.51$, which is $< F(2, 13, 0.95) = 3.81$.

The sixth residual is extremely large and negative (-2.626), indicating that the corresponding Y seems far too low.

Plots of residual versus $\widehat{\log Y}$ exhibit curvature.

It seems doubtful that the errors are additive and it would be sensible to try the nonlinear model $Y = \alpha X_1^{\beta} X_2^{\gamma} + \epsilon$ for which the methods of Chapter 24 are needed. Initial estimates for the parameters (needed for nonlinear estimation) would be taken from the model fitted above, namely, $\alpha_0 = 10^{1.9929} \div 100$, $\beta_0 = 0.5428$, and $\gamma_0 = 0.2740$.

G. Model $Y = \alpha X_1^{\beta} X_2^{\gamma} X_3^{\delta} \cdot \epsilon$.

By taking logarithms to the base e we can convert the model into the linear form.

$$\ln Y = \ln \alpha + \beta \ln X_1 + \gamma \ln X_2 + \delta \ln X_3 + \ln \epsilon$$

or

$$\widehat{\ln Y} = 8.5495297 + 0.1684244 \ln X_1 - 0.537137 \ln X_2 - 0.0144135 \ln X_3.$$

X_2 or peripheral wheel velocity.

All except X_3 = feed viscosity, which provides an F-ratio of

$$2.15 < F(1, 31, 0.95) = 4.16.$$

With 95.52% variation explained and a small standard deviation of 1.563% of the response mean, this looks like a good prediction equation. Plots of residuals reveal no peculiarities.

H. $\hat{Y} = 0.4875 + 0.13X_1 - 0.08X_2 + 0.125X_3 - 0.0975X_4$ explains $R^2 = 0.907$ of the variation about the mean; $s = 0.115$.

Chapter 14

A. Set up a dummy variable Z with values 1, 3, 4, 6, 10, 13, 14, 15, 16, 17, 19 corresponding

to November 17, 19, . . . , December 5. The model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \alpha_1 Z + \alpha_2 Z^2 + \epsilon$ is then fitted. (Note that a Z term as well as Z^2 term is needed to properly represent the quadratic trend. We also would need an α_0 term if we did not already have a β_0 in the model.) Alternatively, we could use dummies $(Z - 1)$ or $(Z - \bar{Z})$; these would change the interpretation of the constant term only.

B. To account for the tillers, we employ two dummy variables Z_1 and Z_2, assigning $Z_1 = 1$ for tiller 1 and zero elsewhere, and $Z_2 = 1$ for tiller 2 and zero elsewhere. A third dummy, Z_3, is assigned the value one for the Waldron variety and zero for the Ciano variety. The N rate variable is called N. We can then fit by least squares the equation

$$\hat{Y} = -333.5 + 185.01X - 7.801X^2 - 0.1294N$$
$$+ 106.0Z_1 + 33.7Z_2 + 19.6Z_3.$$

(We quote sufficient significant figures to produce predictive accuracy of the same order as that of the original data, that is, to the nearest unit.) The analysis of variance table is as follows:

ANOVA

Source	df	SS	MS	F
$b_1 \mid b_0$	1	1,112,031	1,112,031	176.71
$b_{11} \mid b_0, b_1$	1	33,968	33,968	5.40
$b_2 \mid b_0, b_1, b_{11}$	1	20,392	20,392	3.24
$a_1, a_2, a_3 \mid$ others	3	14,994	4,998	0.79
Residual	11	69,225	6,293	
Total (corrected)	17	1,250,610		

$F(1, 11, 0.95) = 4.84$; $F(1, 11, 0.99) = 9.65$.

We see that the total contribution of the dummy variables is not significant. Omission of the dummies leads to an $R^2 = 0.933$ from the fitted equation $\hat{Y} = -338.3 + 200.3X - 7.594X^2 - 0.3696N$, whereas retention of the dummies provides $R^2 = 0.945$, an unimportant increase. Note, however, that if *only* the three dummy variables are used, they account for $R^2 = 0.571$ of the variation about the mean. This is due to correlation between the dummies and X; note, for example, the low tiller 3 responses.

The N rate is, not surprisingly, highly correlated ($r = 0.672$) with X and ($r = 0.629$) with X^2. Its nonsignificant contribution could be omitted from the equation, as could those of the dummies. If this is done, we have $\hat{Y} = -337 + 196.0X - 8.13X^2$ with $R^2 = 0.916$. A plot of the data confirms what the analysis of variance table shows: that the plot has only a slight quadratic bend in it and that the first-order term takes up most of the variation. The values of Y, $\hat{Y}$, se($\hat{Y}$), e, and the standardized residuals are tabulated below for further examination by the reader.

Y	$\hat{Y}$	se($\hat{Y}$)	e	e/se(e)
370	343	25	27	0.34
659	652	32	7	0.09
935	803	50	132	1.97
390	263	24	127	1.59
753	683	32	70	0.91
733	819	62	−86	−1.53
182	192	26	−10	−0.13
417	462	31	−45	−0.58
686	697	31	−11	−0.14
188	148	30	40	0.51
632	546	33	86	1.12
538	690	31	−152	−1.96

Y	$\hat{Y}$	se($\hat{Y}$)	e	e/se(e)
27	22	46	5	0.07
141	148	30	−7	−0.09
262	263	24	−1	−0.01
34	133	31	−99	−1.28
222	192	26	30	0.37
242	355	26	−113	−1.43

C. 1.

$$\mathbf{X} = \begin{array}{c} \begin{matrix} X_0 & X_1 & X_2 & X_3 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix} \end{array}$$

$$\hat{Y} = 248 + 2X_1 - 10X_2 - 7.33X_3.$$

2.

ANOVA

Source of Variation	df	SS	MS	F
Total (corrected)	8	1466.0		
Regression	3	826.7	257.57	NS
$b_1 \mid b_0$	1	54.0 } 504.0	252.00	NS
$b_2 \mid b_0, b_1$	1	450.0 }		
$b_3 \mid b_0, b_1, b_2$	1	322.7	322.70	NS
Residual	5	639.3	127.9	

Operator differences are not statistically significant; that is, $\dfrac{252.0}{127.9} = 1.97$ is less than $F(2, 5, 0.95) = 5.79$.

Operator No. 1: $\hat{Y} = 248 + 2(1) = 250$;

Operator No. 2: $\hat{Y} = 248 - 10(1) = 238$;

Operator No. 3: $\hat{Y} = 248 + 2(-1) - 10(-1) = 256$.

3. There is not sufficient evidence to say line speed affects bar appearance with an α risk of 0.05.

4. Residual plots indicate a second-order model with line speed a better choice.

D. Such problems can often be done several ways, for example:

1. Add the $\mathbf{Z}_0$ vector to all the others. The resulting six vectors are clearly linearly independent, so the system will work.

2. More tediously, solve

$$a\mathbf{Z}_0 + b\mathbf{Z}_1 + c\mathbf{Z}_2 + d\mathbf{Z}_3 + e\mathbf{Z}_4 + f\mathbf{Z}_5 = \mathbf{0}. \quad (1)$$

If $a = b = c = d = e = f = 0$, then columns are linearly independent and the system works. If any of $a, b, \ldots, f$ are *nonzero*, it does not work.

Here we get

$$a + b - c - d - e - f = 0,$$
$$a - b + 2c - d - e - f = 0,$$
$$a - b - c + 3d - e - f = 0,$$
$$a - b - c - d + 4e - f = 0,$$
$$a - b - c - d - e + 5f = 0,$$
$$a - b - c - d - e - f = 0.$$

Subtraction of the last equation from the others in succession shows that $b = c = d = e = f = 0$, whereupon $a = 0$ from the last equation. It works.

3. Find the determinant of the 6×6 matrix. If it is zero, the system fails; if nonzero, it works. Here the determinant has value -720 and so the system works. To get the determinant in a computer, request the eigenvalues. The determinant is the product of these.

E. Remember to put in the Z_0 column of 1's. Refer to the D solution for methods. It works.

F. 1. Leave in u and v and write out the six equations that would hold ($a\mathbf{Z}_0 + \cdots + f\mathbf{Z}_5 = 0$) for linear dependence. Solve them to obtain $a = b = c = d = e = f = 0$, for *any* values of u and v. So the system works if $u = v = 0$ in particular.

2. Since it works for any u and v, the answer is no.

3. Not relevant.

G. System B is given in the text as a suitable one.

System A also is ok. Putting subscripts on the columns we can see that $X_{0A} = X_{0B}$, $X_{1A} = X_{1B} - 5X_{3B}$, $X_{2A} = X_{2B} - X_{3B}$ and $X_{3A} = X_{3B} - X_{0B}$. Thus the columns of A are linearly independent columns derivable from linear combinations of the columns of B. This can also be done by showing that $\hat{Y}_A$ and $\hat{Y}_B$ are identical.

H.

$$\mathbf{X} = \begin{array}{c} \begin{matrix} X_0 & X_1 & X_1^2 & X_2 & X_3 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 \\ 1 & 3 & 9 & 0 & 0 \\ 1 & 4 & 16 & 0 & 0 \\ 1 & 5 & 25 & 1 & 1 \\ 1 & 5 & 25 & 2 & 1 \\ 1 & 5 & 25 & 3 & 1 \end{bmatrix} \end{array}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_{11} X_1^2 + \beta_2 X_2 + \beta_3 X_3 + \epsilon.$$

As in all dummy variable situations, many other valid answers are possible.

I. Using the dummy variable suggested in the similar example in the text, we get $\hat{Y} = -0.5 + 2X_1 + X_2 + 0.2X_3$ with separate equations $\hat{Y}_1 = -0.5 + 2X_1$ and $\hat{Y}_2 = 9.7 + X_2$ for the two lines that intersect at $X_1 = 5.2$ (or $X_2 = 0.2$).

J. We can use dummies with values

$$X_1 = 1, 2, 3, 4, 5, 6, 7, 8, 8, 8, 8, \ldots, 8,$$
$$X_2 = 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 3, \ldots, 64,$$
$$\mathbf{X}_3 = 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, \ldots, 1.$$

to get

$$\hat{Y} = 0.421429 + 0.039643X_1 + 0.004062X_2 + 0.016366X_3.$$

$SS(b_1, b_2, b_3 \mid b_0) = 1.00910$ (3 df).
Residual SS = 0.03829 (68 df).
$F = 597.3 > F(3, 68, 0.99) = 4.10$ (approximately). Significant regression. $R^2 = 0.9634$. Overall an excellent fit appears to have been achieved by these criteria (but see below).

The lines intersect at an age of 7.96 ($X_1 = 8.46$).

The Durbin–Watson statistic is $d = 2.514$; $4 - d = 1.486$ is significant at the 5% level indicating negative serial correlation in the residuals, which needs further investigation. The corresponding upper-tail runs test is also significant ($r = 46$, $n_1 = 32$, $n_2 = 40$, $z = 2.27$, upper-tail probability, 0.0116). The presence of negative serial correlation affects the validity of the regression tests above to some extent. A possible next step would be to estimate the value of ρ and use generalized least squares.

Note: In the original paper, the data were fitted by a quadratic curve and a line rather than two straight lines as here, and no assumption was made as to which points were on which portions of the model. For this more general fit, see the quoted reference.

K. We can use the same dummy layout as in the previous exercise except that the values run only to 32 observations here. Now

$$\hat{Y} = 0.421429 + 0.039643X_1 + 0.004277X_2 + 0.012905X_3.$$

$SS(b_1, b_2, b_3 \mid b_0) = 0.33925$ (3 df).

Residual SS = 0.02152 (28 df).

$F = 147.13 > F(3, 28, 0.99) = 4.57$; significant regression.

$R^2 = 0.9403$. Again an excellent fit by these criteria, but see below.

The lines intersect at an age of 7.86 ($X_1 = 8.36$).

The Durbin–Watson statistic is $d = 2.722$; $4 - d = 1.278$ is almost but not quite significant at the 5% level (where $d_L = 1.24$). The corresponding upper-tail runs test is significant at the 2.5% level (one tail), however. ($r = 23$, $n_1 = 18$, $n_2 = 14$, $z = 2.10$, upper-tail probability, 0.0179). The possibility of negative serial correlation affects the validity of the regression tests above to some extent. A possible next step would be to estimate the values of ρ_s and use generalized least squares.

(Read the note in the previous solution.)

L. If we attach a dummy variable Z to distinguish the two groups, we can look at all four possibilities at once. For example, with $Z = 0$ for set A and $Z = 1$ for set B, we can fit the model $Y = \beta_0 + \beta_1 X + \alpha_0 Z + \alpha_1 XZ + \epsilon$. The fitted equation is $\hat{Y} = 1.142 + 0.506X - 0.0418Z - 0.0360XZ$. The last two coefficients have large standard errors. We can test if a single line is sufficient by ignoring the Z and ZX terms to fit $\hat{Y} = 1.075 + 0.492X$. The extra sum of squares $F = (0.1818/2)/(0.3272/4) = 1.11$. So a single straight line seems appropriate.

M. For the model, see Exercise N. For testing for two "parallel" quadratics, H_0: $\alpha_1 = \alpha_{11} = 0$ is appropriate.

N. **1.** $\alpha_0 = \alpha_1 = \alpha_{11} = 0$.
2. $\beta_{11} = \alpha_0 = \alpha_1 = \alpha_{11} = 0$.
3. By fitting the model as given and setting first $Z = -1$ and then $Z = 1$.
4. No, their Z values would be different.

O. See Section 14.2.

P. Solution is implicit in the question.

Q. Models: $Y_{iu} - \overline{Y}_i = \beta_i(X_{iu} - \overline{X}_i)$, $\quad i = 1, 2, \ldots, m$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad u = 1, 2, \ldots, n$

1.

X_1	Y_1	$(X_{iu} - \overline{X}_i)$	$(Y_{iu} - \overline{Y}_i)$
3.5	24	−1.529	−17.286
4.1	32	−0.929	−9.286
4.4	37	−0.629	−4.286
5.0	40	−0.029	−1.286
5.5	43	0.471	1.714
6.1	51	1.071	9.714
6.6	62	1.571	20.714

$$\overline{X}_1 = 5.029, \qquad \overline{Y}_1 = 41.286, \qquad n_1 = 7.$$

$$b_1 = \left\{\sum_{u=1}^{7} (X_{iu} - \overline{X}_i)(Y_{iu} - \overline{Y}_i)\right\} \Big/ \left\{\sum_{u=1}^{7} (X_{iu} - \overline{X}_i)^2\right\}$$

$$= 81.542858/7.434 = 10.969.$$

$$\mathrm{SS}(b_1) = b_1^2 \left\{\sum_{u=1}^{7} (\overline{X}_{iu} - \overline{X}_i)^2\right\}$$

$$= (120.318961)(7.434) = 894.451.$$

2.

X_2	Y_2		$(X_{iu} - \overline{X}_i)$	$(Y_{iu} - \overline{Y}_i)$
3.2	22	$\overline{X}_2 = 5.533$	−2.333	−19.333
3.9	33	$\overline{Y}_2 = 41.333$	−1.633	−8.333
4.9	39	$n_2 = 6$	−0.633	−2.333
6.1	44		0.567	2.667
7.0	53		1.467	11.667
8.1	57		2.567	15.667

$$b_2 = \left\{\sum_{u=1}^{6} (X_{iu} - \overline{X}_i)(Y_{iu} - \overline{Y}_i)\right\} \Big/ \left\{\sum_{u=1}^{6} (X_{iu} - \overline{X}_i)^2\right\}$$

$$= 119.033334/17.573334 = 6.773520.$$

$$\mathrm{SS}(b_2) = b_2^2 \left\{\sum_{u=1}^{6} (X_{iu} - \overline{X}_i)^2\right\}$$

$$= (45.880573)(17.573334) = 806.274633.$$

3.

X_3	Y_3	$(X_{iu} - \overline{X}_i)$	$(Y_{iu} - \overline{Y}_i)$
3.0	32	−2.775	−18.750
4.0	36	−1.775	−14.750
5.0	47	−0.775	−3.750
6.0	49	0.225	−1.750
6.5	55	0.725	4.250
7.0	59	1.225	8.250
7.3	64	1.525	13.250
7.4	64	1.625	13.250

$$\overline{X}_3 = 5.775, \qquad \overline{Y}_3 = 50.750, \qquad n_3 = 8.$$

$$b_3 = \left\{\sum_{u=1}^{8} (X_{iu} - \overline{X}_i)(Y_{iu} - \overline{Y}_i)\right\} \Big/ \left\{\sum_{u=1}^{8} (X_{iu} - \overline{X}_i)^2\right\}$$

$$= (135.650000)/(18.495) = 7.334415.$$

$$\mathrm{SS}(b_3) = (53.793643)(18.495) = 994.913427.$$

$$b = \left\{\sum_{i=1}^{m=3}\sum_{u=1}^{n_i}(X_{iu} - \overline{X}_i)(Y_{iu} - \overline{Y}_i)\right\} \Big/ \left\{\sum_{i=1}^{m=3}\sum_{u=1}^{n_i}(X_{iu} - \overline{X}_i)^2\right\}$$

$$= (81.542858 + 119.033334 + 135.650)$$

$$\div (7.434 + 17.573334 + 18.495)$$

$$= 336.226192/43.502334 = 7.728923.$$

$$\text{SS}(b) = b^2\left\{\sum_{i=1}^{3}\sum_{u=1}^{n_i}(X_{iu} - \overline{X}_i)^2\right\}$$

$$= (59.736251)(43.502334) = 2598.666299.$$

$$\text{SS due to all } b_i|b = \sum_{i=1}^{3}\text{SS}(b_i) - \text{SS}(b)$$

$$= 894.451000 + 806.274633 + 994.913427$$

$$- 2598.666343 = 96.972717.$$

$$\text{Residual} = \text{Total SS} - \text{SS}(b) - \text{SS due to all } b_i|b$$

$$= 2792.261906 - 2598.666343 - 96.972717 = 96.622846$$

ANOVA

Source	df	SS	MS	F
b	1	2598.666343	2598.666343	403.42
All $b_i\|b$	2	96.972717	48.486359	7.53
Residual	15	96.622846	6.441523	
Total	18	2792.261906		

$H_0: \beta_i = \beta$ $\quad F_2 = 7.53 > F(2, 15, 0.95) = 3.68, \therefore H_0$ is rejected.

R. If we code boot A as $X_1 = -1$ and boot B as $X_1 = 1$ and enter the temperature (as listed) as X_2, the model $\hat{Y} = 10.6 + 0.9208X_1 - 0.15186X_2$ explains 45.1% of the variation. While this is not very good, it seems difficult to do better with the variables available. Lower responses are better and the coefficient of X_1 is positive. So boot A is better, and significantly so ($p = 0.027$). Fitting five dummy variables orthogonally to separate the subjects leaves the fitted coefficients of X_1 and X_2 unchanged and raises R^2 to 57.6%. Boot A is still better, of course.

S. We fit $E(Y) = \beta_0 + \beta_1 X + Z(\alpha_0 + \alpha_1 X)$, where $Z = -1, 1$, because this choice makes the (Z and XZ) columns orthogonal to the (1 and X) columns. Note that the columns *within* the parentheses are *not* orthogonal to each other, however.

Line 1 is then $\hat{Y} = 7.461 + 1.016X$
Line 2 is then $\hat{Y} = 9.670 + 0.320X$.

We can express H_0 as $H_0: \alpha_1 = 0$, and fit the reduced model to get the extra sum of squares for a_1 given b_0, b_1, a_0. The results are:

Source	df	SS	MS	F
$a_0, b_1\|b_0$	2	4.771	2.386	170.4
$a_1\|a_0, b_1, b_0$	1	1.147	1.147	81.9
Residual	4	0.056	0.014	
Total, corrected	7	5.974		

Clearly, H_0 is rejected and the two separate lines are needed.

T. **(a)** We can fit $\hat{Y} = 0.878 + 0.0670X + 3.641Z - 0.0147XZ \quad (R^2 = 0.945)$, where Z defines the dummy variable column $\mathbf{Z} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)'$. The XZ term can be dropped, leading to the pair of parallel lines $\hat{Y} = 1.329 + 0.0589X + 2.972Z$ (for $Z = 1$ and $Z = 0$). This reduced equation explains $R^2 = 0.938$ of the variation about the mean $\overline{Y}$.

(b) Now, $\hat{Y} = -10.216 + 3.718U + 9.820Z - 1.780UZ$, where $U = \ln X$ (with $R^2 = 0.972$), but no terms can be dropped, so that two different lines are needed. Thus, with these data, little advantage is gained by using $\ln X$ rather than X. We remind the reader that our example data were interpolated from a published graph and that results with the actual data, not given in the paper, may well differ.

U. Using $Z_1 = 1$ for cherry and -1 for yellow oval, and using $(Z_2, Z_3) = (1, 0)$, $(0, 1)$, and $(0, 0)$ for the early, middle, and late planting times, respectively, leads to the fitted equation

$$\hat{Y} = \underset{(5.63)}{9.214} - \underset{(4.5)}{1.5}\, Z_1 + \underset{(7.2)}{33.786}\, Z_2 + \underset{(10.6)}{21.29}\, Z_3.$$

The standard errors (shown in parentheses) and the analysis of variance table (not shown) reveal "no differences" between the two tomato types as far as production numbers are concerned. The dummies Z_2 and Z_3 should be regarded as a unit and not as separate variables. By choosing Z_2 and Z_3 in various ways we can make one look more important than the other at our choice. Thus dropping one of these variables is usually pointless. As a unit they show significant differences caused by different planting times.

The experimental layout is not well-planned. It would have been better to increase the direct comparisons between cherry and yellow oval tomatoes by planting some yellow ovals in the middle period and some cherries in the late period.

Chapter 15

A. 1.

$$\hat{Y} = 6360.3385 + 13.868864X_1 + 0.211703X_2 - 126.690360X_3 - 21.817974X_4.$$

2. The above least squares equation shows an R^2 of 76.702710, which is the highest one in the given regression information. The overall $F = 9.8770256$ is statistically significant. The partial F-values are also all statistically significant. None of the 95% confidence limits on the β coefficients includes zero. The standard deviation as percent of response mean = 7.536%, which is lower than any other in the given information.

3. The random vector X_5 does not contribute significantly to the explanation of variation. Actually, it contributes less than 1% and it increases the standard deviation as percent of the response from 7.536 to 7.759. The partial F-test also shows that this variable is not statistically significant.

4. This part is left to the reader.

B. 1. The stepwise procedure enters X_1 ($F = 12.60$), enters X_2 ($F = 2.04$), enters X_3 ($F = 3.62$). Now, however, X_1 has weakened with partial $F = 0.06$. X_1 is rejected and both X_2 and X_3 remain. The final equation is

$$\hat{Y} = 63.021 + 11.517X_2 - 0.816X_3.$$

2.

Source of Variation	df	SS	MS	F	Partial F
Total (corrected)	8	1279.20010			
Regression	2	1079.12600	539.56300	16.181	
$b_2 \mid b_0$	1	754.40445	754.40445	22.624	17.197
$b_3 \mid b_0, b_2$	1	324.72155	324.72155	9.738	9.738
Residual	6	200.07410	33.34568		

Residuals plots reveal no problems.

C. A study of the correlation matrix shows right away that X_1 and X_5 are perfectly correlated, caused by the fact that workers 1 and 5 are either *both* on duty or *both* absent in every run. Thus their effects cannot be separately assessed and, for regression purposes, we can drop X_5 immediately (or X_1, it makes no difference which). In these data, there are only eight distinct runs, there being twelve degrees of freedom for pure error. The sum of squares for pure error is 131.929 (12 df), so that $s_e^2 = 10.994$. The accompanying table shows the residual sums of squares for the various models in variables X_1, X_2, X_3, and X_4.

Variables in Model[a]	Residual df	Residual SS[b]	$100R^2$
—	19	42,644.00	—
1	18	8,352.28	80.14
2	18	36,253.69	14.99
3	18	36,606.19	14.16
4	18	27,254.91	36.09
12	17	7,713.10	81.91
13	17	762.55	98.21
14	17	6,071.56	85.76
23	17	32,700.17	23.32
24	17	24,102.10	43.48
34	17	16,276.60	61.83
123	16	761.41	98.21
124	16	5,614.59	86.83
134	16	163.93	99.62
234	16	15,619.01	63.37
1234	15	163.10	99.62

[a] In addition to β_0, which is in all models.
[b] Before removal of pure error.

If we adopt a stepwise procedure, we are led through this sequence:

(a) Enter X_1. $F_1 = (42{,}644 - 8{,}352.28)/(8352.28/18) = 73.90$. Retain X_1.
(b) Add X_3. $F_3 = (8352.28 - 762.55)/(762.55/17) = 169.20$. Retain X_3.
$F_1 = (36{,}606.19 - 762.55)/(762.55/17) = 799.08$. Retain X_1.
(c) Add X_4. $F_4 = (762.55 - 163.93)/(163.93/16) = 58.43$. Retain X_4.
$F_3 = (6071.56 - 163.93)/(163.93)/16) = 576.60$ Retain X_3.
$F_1 = (16{,}276.60 - 163.93)/(163.93/16) = 1572.64$ Retain X_1.

D. We give here the results of applying backward elimination, forward selection, and stepwise regression. For an analysis that uses the C_p statistic, see either of the source references.

1. *Backward Elimination*

For the full equation (123456), SS(residual) = 0.307 with 31 − 7 = 24 df. The residual sum of squares for each five-variable equation is:

Variables	SS(residual)	
12345	0.412	
12346	0.311	← This is the best, so test variable 5.
12356	0.364	Partial F-test for 5, given 12346:
12456	0.313	
13456	0.545	$F_{1,24} = \dfrac{0.311 - 0.307}{0.307/24} = 0.31.$
23456	0.939	

The partial F-test indicates that variable 5 is not necessary for a good fit. Drop variable 5.

Our model now includes just the variables (12346), so SS(residual) = 0.311 with 31 − 6 = 25 df. This model has the following four-variable submodels:

Variables	SS(residual)	
1234	0.441	
1236	0.365	
1246	0.323	← Test variable 3.
1346	0.555	Partial F-test for 3, given 1246:
2346	0.995	
		$F_{1,25} = \dfrac{0.323 - 0.311}{0.311/25} = 0.96.$

Variable 3 appears unnecessary. Drop variable 3.

Our model now is (1246). SS(Residual) = 0.323 with 31 − 5 = 26 df. This model has the following three-variable submodels:

Variables	SS(residual)	
124	0.450	
126	0.367	← Test variable 4.
146	0.558	Partial F-test for 4, given 126:
246	1.192	
		$F_{1,26} = \dfrac{0.367 - 0.323}{0.323/26} = 3.54.$

This F-value is significant at the 0.10 level, but not at the 0.05 level. This is a borderline case. If the equation is to be used just to summarize the data, it would probably be wise to keep variable 4 in the equation. The backward elimination procedure then stops and the final model is (1246). If the equation is to be used for prediction, however, the fact that variable 4 (a dummy variable that expresses an overall difference in response between the two sets of runs) is not reproducible would probably cause it to be dropped.

If variable 4 is dropped, the resulting model is (126), and SS(residual) = 0.367, with 31 − 4 = 27 df.

This model has the following two-variable submodels:

Variables	SS(residual)	
12	0.499	← Test variable 6.
16	0.576	Partial F-test for 6, given 12:
26	9.192	
		$F_{1,27} = \dfrac{0.499 - 0.367}{0.367/27} = 9.71$

This F-value is significant at the 0.01 level. Conclude variable 6 is necessary and cannot be dropped. Procedure ends and final model is (126).

2. *Forward Selection.*

To find the variable most correlated with the response, we consider all six one-variable equations:

Variable	SS(residual)	Variable	SS(residual)
1	0.607	4	1.522
2	10.795	5	9.922
3	10.663	6	9.196

The model with variable 1 is the best. The F-statistic for variable 1 is:

$$F_{1,29} = \frac{11.058 - 0.607}{0.607/29}$$

$$= 499.31$$

which is highly significant. We now try adding each of the 5 remaining variables:

Variables	SS(residual)	
12	0.499	← Test variable 2.
13	0.600	Partial F-test for 2, given 1:
14	0.582	
15	0.597	$F_{1,28} = \frac{0.607 - 0.499}{0.499/28} = 6.06.$
16	0.576	

This F-value is significant at the 0.05 level. Add variable 2 to the model. Our model now includes the variables (12). Try adding each of the four remaining variables:

Variables	SS(residual)	
123	0.498	
124	0.450	
125	0.447	
126	0.367	← Test variable 6.
		Partial F-test for 6, given 12:
		$F_{1,27} = \frac{0.499 - 0.367}{0.367/27} = 9.71.$

This F-value is significant at the 0.01 level. Add variable 6 to the model. Our model is now (126). Try adding each of the three remaining variables:

Variables	SS(residual)	
1236	0.365	
1246	0.323	← Test variable 4.
1256	0.364	Partial F-test for 4, given 126:
		$F_{1,26} = \frac{0.367 - 0.323}{0.323/26} = 3.54.$

This is a borderline case. (See discussion above where the same decision had to be made in the backward elimination procedure.) If we decide not to include variable 4, the procedure stops and the final model is (126). If we include variable 4, the procedure continues, and we try adding each of the two remaining variables to the model (1246):

Variables	SS(residual)	
12346	0.311	← Test variable 3.
12456	0.313	Partial F-test for 3, given 1246:
		$F_{1,25} = \frac{0.323 - 0.311}{0.311/25} = 0.96.$

This F-value is not significant. Do not add variable 3. Stop procedure. Final model is (1246).

(*Note*: At several stages, the best-fitting equation turned out to be the same under the forward selection procedure as under the backward elimination procedure. This will not necessarily be the case with other sets of data.)

3. Stepwise Regression.

This procedure starts out like the forward selection procedure described above, with the inclusion of variable 1, then variable 2. At this point we do a partial F-test for 1, given 2. Since SS(residual) for 2 is 10.795, and SS(residual) for (12) is 0.499,

$$F_{1,28} = \frac{10.795 - 0.499}{0.499/28} = 577.73.$$

This F-value is highly significant, so we cannot remove variable 1. We shall assume that the critical F-value for deletion does not exceed the critical F-value for entry (as recommended in Section 15.4) so we need not test variable 2, the last variable entered, for possible deletion. Resuming the forward selection procedure, we enter variable 6 as above, then test to see whether either of the previously entered variables (1 or 2) can be removed.

$$\text{Partial } F\text{-test for 1 given 26: } F_{1,27} = \frac{9.192 - 0.367}{0.367/27} = 649.25.$$

$$\text{Partial } F\text{-test for 2 given 16: } F_{1,27} = \frac{0.576 - 0.367}{0.367/27} = 15.38$$

Both of these F-values are significant at the 0.01 level, so we cannot remove either of them. Resume the forward selection procedure, choosing variable 4 as the next candidate for entry. If we decide that the partial F-value (3.51) does not warrant the addition of variable 4, the stepwise regression procedure stops and the final model is (126).

If we decide to keep variable 4, continue the stepwise regression procedure by testing to see if any of the previously entered variables (1, 2, or 6) can be dropped.

$$\text{Partial } F\text{-test for 1 given 246: } F_{1,26} = \frac{1.192 - 0.323}{0.323/26} = 69.95.$$

$$\text{Partial } F\text{-test for 2 given 146: } F_{1,26} = \frac{0.558 - 0.323}{0.323/26} = 18.92.$$

$$\text{Partial } F\text{-test for 6 given 124: } F_{1,26} = \frac{0.450 - 0.323}{0.323/26} = 10.22.$$

These F-values are all significant at the 0.01 level. None of them can be dropped. Continuing, we try to add variable 3, but find that it does not significantly improve the fit. The final model is thus (1246).

E. Both stepwise and backward elimination procedures produce the equation

$$\hat{Y} = 3.068360 + 0.0007259X_1 + 0.0446022X_4 \qquad (1)$$

with an $R^2 = 0.7874$. However, there is lack of fit indicated, the appropriate F-statistic for the lack of fit test being $F = \{0.27351/73\}/\{0.02154/14\} = 2.435$, compared to $F(73, 14, 0.95) = 2.21$, approximately. The residuals plots show a quadratic tendency in X_1, and the Durbin–Watson statistic $d = 0.8480$ indicates evidence of positive serial correlation.

When the term $\beta_{11}X_1^2$ is added to the model, other terms now enter, also. The stepwise procedure, for example, produces the following sequence of entries:

Predictor	R^2	Change in R^2	Significance Level (α)
X_1 in	0.5999	0.5999	0.000
X_4 in	0.7874	0.1875	0.000
X_1^2 in	0.8917	0.1043	0.000
X_6 in	0.9000	0.0083	0.009
X_7 in	0.9056	0.0056	0.028

The final fitted model is

$$\hat{Y} = 2.997526 + 0.0019418X_1 - 0.00000289X_1^2 + 0.0222738X_4 + 0.0607877X_6 - 0.0224359X_7. \quad (2)$$

There is now no lack of fit shown by the F-test, the value of the statistic being $F = \{0.10944/70\}/\{0.02154/14\} = 1.016 < F(70, 14, 0.95) = 2.21$, approximately. The Durbin–Watson statistic is $d = 1.60$, an inconclusive result by the usual test. The table above shows that X_6 and X_7 contribute little to R^2 even though their entry into the model is "statistically significant," where we use the standard (but wrong, see p. 343) test. If we omit X_6 and X_7 and refit, we obtain the fitted equation

$$\hat{Y} = 3.0016115 + 0.0018344X_1 - 0.000002641X_1^2 + 0.0390777X_4. \quad (3)$$

The lack of fit test now provides the statistic $F = \{0.12877/72\}/\{0.02154/14\} = 1.162$, which is not significant compared to $F(72, 14, 0.95) = 2.21$ approximately. The Durbin–Watson statistic is $d = 1.64$, an inconclusive result by the usual test.

Thus if we wish to use the model indicated by the formal stepwise procedure, we should use Eq. (2), but Eq. (3) seems a sensible practical compromise. Examination of the corresponding printouts shows that the predictions from Eq. (3) are close to those from Eq. (2).

Additional Comments:

(*i*) An X_1^3 term can also be considered. If it is, the stepwise entry sequence is X_1 ($R^2 = 0.5999$), X_4 (incremental change in $R^2 = 0.1875$), X_1^2 (0.1043), X_6 (0.0083), X_1^3 (0.0079), X_7 (0.0052).

(*ii*) If X_4^2 and X_4^3 are considered in addition (as well as X_1^2 and X_1^3), the stepwise entry sequence (with incremental R^2 values) is exactly the same as in (*i*) but with two extra steps, namely X_4^3 (0.0052) comes in, but now X_4 (−0.0010) goes out.

In both cases (*i*) and (*ii*), by arguing as above, we would probably use just X_1, X_4, and X_1^2, since all the other significant variables add less than 0.01 each to R^2.

F. $$\hat{Y} = 0.4368012 + 0.0001139X_1 - 0.0051897X_3 - 0.0018887X_4 + 0.0044263X_5.$$

The plot of the residuals versus $\hat{Y}$ indicates that the variance is not homogeneous. One should try weighted least squares, or perhaps a transformation on the Y_i.

This model explains only 76.9% of the total variation, and the confidence limits on β_1 and β_4 include zero. The standard deviation of the residuals is 3.3% of the mean response. Thus the model predicts well but is not as good as one would like. If one can get rid of the large variance for large Y's, the model will be much better.

G. Model: $\hat{Y} = b_0 + b_2X_2 + b_8X_8$

or $\hat{Y} = 9.4742224 + 0.7616482X_2 - 0.0797608X_8$

$R^2 = 86.0\%$

Standard deviation as percent of response mean = 6.761%.

H. 1. $\hat{Y} = 87.158859 + 0.8519104X_1 + 0.5988662X_2 + 2.3613018X_6 - 0.9755309X_9$.

where X_1 = year,
X_2 = preseason precipitation in inches,
X_6 = rainfall in July in inches,
X_9 = August temperature.

2. The most important variable is X_1, which accounts for the upward trend in corn yield. Of all the other variables, only preseason precipitation, July rainfall, and August temperature contribute significantly to the regression.
3. With an R^2 of 72.06% and standard deviation as percent of response mean of 14.903%, this prediction equation needs to be improved. New variables should be found to bring R^2 up, and to decrease the standard deviation of residuals. Investigation of the residuals may yield some insight into this problem. (See "Fast very robust methods for the detection of multiple outliers," by A. C. Atkinson, *Journal of the American Statistical Association*, **89,** 1994, 1329–1339. Observations 7, 8, and 11 "may be outliers" on p. 1336.)

I. 1. There is a lot of replication in the data. Thus an independent estimate of pure error can be obtained. The analysis of variance can be written down as:

ANOVA

Source of Variation	df
Total	47
Regression	5
Residuals	42
Lack of fit	2
Pure error	40

2. $\hat{Y} = 134.258 + 0.050X_1 - 0.012X_2 + 0.834X_3 - 0.154X_4 - 3.804X_5$.
3. The model is not adequate since the lack of fit is statistically significant at $\alpha = 0.05$.

Source of Variation	df	SS	MS	F
Total	47	2850.3107		
Regression	5	1817.1055		
Residual	42	1033.2052		
Lack of fit	2	383.7052	191.8526	11.82*
Pure error	40	649.5000	16.2375	

4. This model explains only 63.75% of the variation, and it is not a good one. The residuals show definite nonrandom patterns.
5. This experiment is poorly designed; there are too many replicates and not enough different design points.

J. The *prediction* equation obtained by the stepwise procedure using a critical F of 2.00 for acceptance and rejection is

$$\hat{Y} = 250.1875 - 2.3124998\left(\frac{X_1 - 146}{3}\right) - 14.687499\left(\frac{X_2 - 69.5}{3.5}\right) - 2.8124997\left(\frac{X_6 - 289.5}{93.5}\right).$$

The optimum rate using this prediction equation will be at the point $\hat{Y} = 270$; $X_1 = 143$, $X_2 = 66$, $X_6 = 196$; and the other variables held at their mean levels, namely, $X_3 = -10$, $X_4 = 132.5$, $X_5 = 91.5$.

K. $\hat{Y}_1 = -2.80512 + 0.15176X_1 + 3.60191X_3$,
$\hat{Y}_2 = -2.84492 + 0.11344X_1 + 3.67343X_3$.

L. 1. Both the stepwise and the backward elimination procedures (with F-to-enter and F-to-remove values set to 4) produce the equation

$$\hat{Y} = -1.2471 + 0.510X_2 + 0.768X_4,$$

with an $R^2 = 0.9770$, and $s^2 = 0.01$. So an excellent fit is obtained.

2. High correlations (in excess of 0.8) are observed between the following variables (indicated by the subscripts):

$$r_{23} = 0.901, \quad r_{34} = 0.850, \quad r_{37} = 0.835, \quad r_{47} = 0.879.$$

The main hint from these is that question 3 may be redundant and could be dropped.

3. If the relationship *were* causal, the regression would be saying this: Teach a well organized course and answer questions in a helpful manner and you will be assured of a good overall grade as an instructor. The conclusion appears to be a very reasonable one!

M. 1. Look at the equations 145 (RSS = 0.569), 245 (1.030), 345 (1.383), 456 (1.352). The combination 145 is best (lowest RSS) and the consequent

$$F_{1|45} = \{(1.397 - 0.569)/1\}/\{0.569/27\} = 39.26,$$

which is significant, exceeding 4.21.

2. The RSS for 12345 is 0.412. The least increase for any four of the five variables is attained by 1245 (0.413). So $F_{3|1245} = \{(0.413 - 0.412)/1\}/\{0.412/25\} = 0.06$. Variable 3 would be removed.

3. $S_{YY} = 11.058$. $R^2 = (11.058 - 0.441)/11.058 = 0.9601$.

4. Using 3456 (1.342) and 123456 (0.307) we get

$$C_p = \frac{1.342}{0.307/24} - (31 - 10) = 83.91.$$

This is not a useful regression.

5. For 1246, RSS = 0.323. Dividing the latter by $31 - 5 = 26$ gives $s^2 = 0.0124$.

N. 1. $\hat{Y} = -50.359 + 0.6711545X_1 + 1.295351X_2$.
No lack of fit, $F = 1.42$ with 12, 6 df. $R^2 = 0.8506$.

2. The residuals show that the fitted equation is least satisfactory at $(X_1, X_2, X_3) = (70, 20, 91)$, that is, at run 21; thus one would be reluctant to use the equation in that neighborhood. Future runs should be chosen to provide more balanced coverage of the X-space.

Additional Notes: For a more detailed and extensive analysis of these data, see Chapters 5 and 7 (p. 138) of Daniel and Wood (1980). This particular set of data is one of the most analyzed regression problems in the literature! It has provided wonderful ammunition for those critical of least squares, because many authors maintained that there were four questionable observations, mentioned below, not all detected immediately via a least squares fit. In a thoroughly researched and amusing article, Dodge (1996) has investigated the history of these data. Among other things, he points out that various methods of analyzing them have provided at least 26 *distinct sets of detected outliers*, the most cited set being observations 1, 3, 4, and 21. Many references are provided by Dodge. See also Atkinson (1994, pp. 1330–1331).

O. For the types of detailed calculations needed, see the fully worked solution to Exercise 15D. We summarize the steps needed for this example, only.

1. *Backward Elimination.* The RSS for 123456 is 12.508. If we look at the RSS values for all sets of five, we see the smallest increase is for 12356 with RSS = 12.542. The $F(1,19)$ value is $\{(12.542 - 12.508)/1\}/\{12.508/19\} = 0.05$, not significant; so drop 4.

Proceeding similarly from 12356 we are led to drop 1. Then at the $\alpha = 0.10$ level, we quit at 2356. At the $\alpha = 0.05$ level we proceed to 25, where we quit.

2. *Stepwise Regression.* We first select 2, then 5. Variable 3 is next up. At the $\alpha = 0.05$ level we quit at 25. At the $\alpha = 0.10$ level we proceed from 25 to 235 to 2356.
3. C_p *Statistic.* The candidate sets of predictors with low C_p values are the following:

$p = 3$	$p = 4$	$p = 5$	$p = 6$
25 ($C_p = 5.44$)	125 (6.28)	1235 (5.39)	12345 (6.85)
	235 (4.31)	2345 (5.09)	12356 (5.05)
	245 (6.43)	2356 (3.28)	23456 (5.12)

Clearly, the equations 25, 235, and 2356 attract attention, with a final choice depending on how may predictors it is decided to include to achieve the indicated C_p reductions as p increases.

P. The selection procedure identifies x_1, x_2, and x_1x_2 as the most important terms. Residual No. 12 seems far too remote from the others. The four observations (Nos. 4, 8, 12, and 16) at $(x_1, x_2) = (1, 1)$ are 139.7, 141.4, 48.6, and 172.6, leading to the suspicion that 48.6 perhaps should have been 148.6. With this replacement, the fit improves to an $R^2 = 0.978$ from the previous 0.753. The largest residual is now the 16th, but the overall fit is excellent, and the two factors x_1 and x_2 and their interaction provide an excellent explanation of the data. (In the original paper, the author replaced 48.6 by the rounded average, 151.2, of the other three numbers mentioned above. This is also clearly a sensible way to deal with the matter.)

Chapter 16

A. No, because the columns are related via $-2X_1 + X_2 + X_3 = 0$. Dropping any one of the three related columns allows a unique fit. If we form the **X** matrix by adding a column $\mathbf{X}_0 = \mathbf{1}$, and apply the Gram–Schmidt procedure, we get the matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & -\frac{17}{5} & \frac{215}{186} & 0 \\ 1 & -\frac{12}{5} & -\frac{822}{186} & 0 \\ 1 & -\frac{2}{5} & \frac{824}{186} & 0 \\ 1 & \frac{13}{5} & \frac{131}{186} & 0 \\ 1 & \frac{18}{5} & -\frac{348}{186} & 0 \end{bmatrix}$$

B. No. We first add an $\mathbf{X}_0 = \mathbf{1}$ column to get **X**. The second column is already orthogonal to the first. We find

$$Z_{3T} = \tfrac{1}{2}(-5, -3, -1, 1, 3, 5)'$$
$$Z_{4T} = (0, 0, 0, 0, 0, 0)', \quad \text{so column dependence exists.}$$

In fact, the sum of the second, third, and fourth columns of **X** is zero.

C. No. We add an $\mathbf{X}_0 = \mathbf{1}$ column and columns generated by X_1^2, X_2^2 and X_1X_2 to form **X**. Note that $X_1^2 = X_2^2$ always, so the **X** matrix is singular. We can fit the model $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta(X_1^2 + X_2^2) + \beta_{12}X_1X_2 + \epsilon$ but cannot estimate β_{11} and β_{22} individually.

D. We first note that X_1 and X_2 are centered already and we center Y about $\overline{Y} = 12$. We next see that $S_{11} = 10$, $S_{22} = 34$, and $S_{YY} = 38$ and the square roots of these will be the scale

factors. After centering and scaling, the data become:

Z_1	Z_2	Y
−0.632	−0.343	0
−0.316	−0.171	−0.487
0	0	−0.487
0.316	0.171	0.324
0.632	0.343	0.649

The normal equations are

$$\begin{aligned} a_1 + 0.976\ a_2 &= 0.667, \\ 0.976a_1 + \quad a_2 &= 0.584, \end{aligned}$$

with solution $a_1 = 2.055$, $a_2 = 1.421$. The determinant of the correlation matrix is $1 - (0.976)^2 = 0.047$. The square root of this is 0.218. (See Section 5.5.)

Chapter 17

A. No, although the ridge estimators can be expressed in terms of the least squares estimators. See Appendix 17A.

B. Note that $\overline{X} = 0$ and $S_{XX}^{1/2} = 4$ is the scaling factor for X. We thus get a new column

$$\mathbf{Z} = (-0.5, -0.25, -0.25, -0.25, 0, 0.25, 0.5, 0.5)'$$

and the ridge equation is simply $b_1(\theta) = (1 + \theta)^{-1}\Sigma Z_i Y_i = 8.75/(1 + \theta)$. Then $b_0(\theta) = \overline{Y} - b_1(\theta)\overline{X} = \overline{Y} = 40$. The fitted ridge equation is thus $\hat{Y} = 40 + 8.75Z/(1 + \theta) = 40 + 2.1875X/(1 + \theta)$. The least squares fit appears when $\theta = 0$. When $\theta = 0.4$ we get $\hat{Y} = 40 + 6.25Z = 40 + 1.5625X$.

Chapter 18

A. In all cases we might consider fitting a linear model function $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$ but there are several ways this could be done.

1. Fit $f(Y_i) = \eta_i + \epsilon_i$ using a generalized linear model fit with, for example, $f(Y_i) = \ln\{Y_i/(1 - Y_i)\}$ (Chapter 18).
2. Fit the responses $f(Y_i) = \ln\{Y_i/(1 - Y_i)\}$ by generalized (weighted) least squares with weights inversely proportional to the variances $V\{f(Y_i)\} = \{Y_i(1 - Y_i)m\}^{-1}$ (Chapter 9, also see Chapter 13).
3. Transform $U_i = \sin^{-1}(Y_i^{1/2})$ and use least squares (Chapter 13).

Chapter 19

A. $\hat{Y} = 6.333 + 0.770z_1 + 0.666z_2 - z_1^2 - 0.333z_2^2 + 0.385z_1z_2$.

B.

$$\mathbf{T} = \begin{bmatrix} -a & -a & 0 & 0 \\ -b & -b & 2b & 0 \\ -c & -c & -c & 3c \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

New coordinates are, respectively, $(a, -b, -c)$, $(0, -b, -c)$, $(0, 0, -c)$, $(0, 0, 0)$, where $a = (2/3)^{1/2} = 0.816$, $b = 2^{1/2}/3 = 0.471$, $c = 0.333$.

C. Look at Figure 19.6*b* and think of it (ignoring the corner markings) as a slice of a fourth variable. Imagine two such slices, one above the other, one smaller than the other (see also Figure 19.4c). It is clear that there will be, at most, 12 extreme vertices. In fact there are fewer here mainly because the 0.80 (80%) cutoff plane for x_1 is so high that no vertices appear on it. A mechanical way to generate the extreme vertices for q dimensions is to choose any $(q - 1)$ pairs of limits temporarily ignoring the remaining pair and writing points with all possible combinations of the $(q - 1)$ pairs of levels. For example, if we ignore x_4 first, we write, converting 0.80 to 80% and so on, to get rid of decimals:

x_1	x_2	x_3	x_4	Feasible x_4?	Vertex
10	25	20	45	Yes	1
80	25	20	—		
10	45	20	25	Yes	2
80	45	20	—		
10	25	40	25	Yes	3
80	25	40	—		
10	45	40	5	No	
80	45	40	—		

The x_1, x_2, x_3 levels are all possible combinations of the restriction limits; x_4 values are added, which bring $x_1 + x_2 = x_3 + x_4 = 100$. For the point to be a vertex, we need $0.15 \leq x_4 \leq 55$. Repeating this by next ignoring x_3, then x_2, then x_1 gives further vertices as follows:

4. (10, 25, 50, 15); 5. (10, 45, 30, 15); 6. (10, 35, 40, 15); 7. (40, 25, 20, 15);
8. (20, 45, 20, 15); 9. (20, 25, 40, 15).

To get face centroids we must identify faces on which an x_i is constant at a boundary level. For example:

$x_1 = 10$: Vertices 1, 2, 3, 4, 5, 6, essentially similar to Figure 19.6*b*.
$x_2 = 25$: Vertices 1, 3, 4, 7, 9.
$x_2 = 45$: Vertices 2, 5, 8.
$x_3 = 20$: Vertices 1, 2, 7, 8.
$x_3 = 40$; Vertices 3, 6, 9.
$x_4 = 15$; Vertices 4, 5, 6, 7, 8, 9.

The centroid of the $x_3 = 20$ face, for example, is at

$$\{(10 + 10 + 40 + 20)/4, (25 + 45 + 25 + 45)/4, (20 + 20 + 20 + 20)/4, (45 + 25 + 15 + 15)/4\},$$

namely, at (20, 35, 20, 25). The overall centroid, the average of all vertex points, is at (15.6, 32.8, 31.1, 20.6) whose ingredients add to 100.1 due to rounding error.

D. $\hat{Y} = 40.91x_1 + 25.432x_2 + 28.61x_3 - 8.214x_1x_2 - 39.339x_1x_3 + 8.002x_2x_3$. At the centroid, $\hat{Y} = 27.26$. The contours are curves with highest response value at (1, 0, 0), declining as x_1 decreases. (They have not been plotted here.)

E. The points are on the vertices of the triangular mixture space, on the midpoints of the sides, and the final point is right in the middle.

F. For $m = 2$, $q = 3$, only points with coordinates 0 or $\frac{1}{2}$ or 1 whose coordinates add to 1 are allowed. Of the 27 possible combinations, six are valid, namely, the first six points mentioned

in Exercise E. The number of valid points in the general case is

$$\binom{q+m-1}{m}.$$

For $m = 2$, $q = 3$ this is

$$\binom{4}{2} = 6.$$

G. The vertices are (1, 0, 0), (0, 1, 0), and (0, 0, 1). Averaging of pairs gives $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{2}, 0, \frac{1}{2})$, and $(0, \frac{1}{2}, \frac{1}{2})$. The average of all three is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

H. Six, namely, β_1, β_2, β_3, β_{12}, β_{13}, β_{23}. Thus data from (a) will provide an exact fit while (b) will provide only one degree of freedom for error.

Chapter 20

A. **1.** $S(\beta_0, \beta_1) = (1.4 - \beta_0 - \beta_1)^2 + (2.2 - \beta_0 - 2\beta_1)^2 + (2.3 - \beta_0 - 3\beta_1)^2 + (3.1 - \beta_0 - 4\beta_1)^2$.

2. $b_0 = 0.95$, $b_1 = 0.52$.

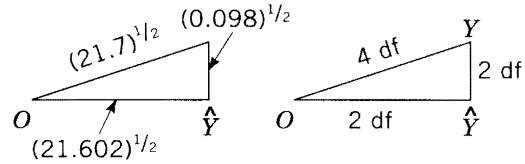
3.

$$\hat{\mathbf{Y}} = \begin{bmatrix} 1.47 \\ 1.99 \\ 2.51 \\ 3.03 \end{bmatrix}, \qquad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \begin{bmatrix} -0.07 \\ 0.21 \\ -0.21 \\ 0.07 \end{bmatrix}.$$

4.

$$\mathbf{P} = \frac{1}{10}\begin{bmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{bmatrix}.$$

5.



Solution A5

6.

Source	SS	df	MS	$F(2, 2)$
Regression	$O\hat{Y}^2 = 21.602$	2	10.801	220.43
Residual	$Y\hat{Y}^2 = 0.098$	2	0.049	
Total	$OY^2 = 21.7$	4		

$F = (O\hat{Y}^2/2)/(Y\hat{Y}^2/2)$ in diagram.

7. $\boldsymbol{\beta}_0 = (1, 0.5)'$, $\mathbf{X}\boldsymbol{\beta}_0 = (1.5, 2.0, 2.5, 3.0)'$ and this point replaces O in figures similar to those in 5, but with different lengths except for $Y\hat{Y}$. The new ANOVA is based on

$$\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0 = (\hat{\mathbf{Y}} - \mathbf{X}\boldsymbol{\beta}_0) + (\mathbf{Y} - \hat{\mathbf{Y}})$$

or, for the squared lengths (sum of squares of the elements),

$$0.1 = 0.002 + 0.098.$$

$F = 0.001/0.049 = 0.02$, not significant.

8.

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}\boldsymbol{\alpha}, \quad \text{say.}$$

So we need

$$\boldsymbol{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 1 & 0.5 & 0 & -0.5 \\ -0.3 & -0.1 & 0.1 & 0.3 \end{bmatrix}.$$

9. $\mathbf{PX} = \mathbf{X}$, so the two columns of $\mathbf{X}$ define the required two combinations.

10. $\mathbf{X}_{1.0} = (-1.5, -0.5, 0.5, 1.5)$,
$\hat{\mathbf{Y}} = 2.25\mathbf{X}_0 + 0.52\mathbf{X}_{1.0}$.

11. From 10, we get an orthogonal SS split of the regression SS as

$$21.602 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = (2.25\mathbf{X}_0)'(2.25\mathbf{X}_0) + (0.52\mathbf{X}_{1.0})'(0.52\mathbf{X}_{1.0})$$

$$21.602 = 20.25 + 1.352.$$

Thus $F = (1.352/1)/0.049 = 27.59 > F(1, 2, 0.95) = 18.51$.

B. 1.

$$\begin{bmatrix} 4 & 0 & 40 \\ 0 & 20 & 60 \\ 40 & 60 & 580 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 16 \\ -8 \\ 136 \end{bmatrix}.$$

These are dependent equations: $E_3 = 10E_1 + 3E_2$.

2. Let $b_2 = b$ (anything). Then $b_0 = 4 - 10b$, $b_1 = -0.4 - 3b$.

3.

$$\hat{\mathbf{Y}} = \mathbf{Xb} = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -1 & 7 \\ 1 & 1 & 13 \\ 1 & 3 & 19 \end{bmatrix} \begin{bmatrix} 4 - 10b \\ -0.4 - 3b \\ 6 \end{bmatrix} = \begin{bmatrix} 5.2 \\ 4.4 \\ 3.6 \\ 2.8 \end{bmatrix}.$$

Note that $\hat{\mathbf{Y}}$ is unique; the b has disappeared!

4. Because the normal equations are dependent, $\mathbf{b}$ is not unique. Thus $\hat{\mathbf{Y}} = \mathbf{Xb}$ can be *described* in many ways. There is exactly one $\hat{\mathbf{Y}}$, however, the foot of the perpendicular from $\mathbf{Y}$ onto $R(\mathbf{X})$. This simple example illustrates regression when $\mathbf{X}$ is not of full rank.

C. 1. $\mathbf{X}'\mathbf{X}$ is singular, so any generalized inverse is needed. The one that puts zeros in the last row and column and the inverse of the first 3 rows and columns of $\mathbf{X}'\mathbf{X}$ in the corresponding positions is

$$(\mathbf{X}'\mathbf{X})^- = \frac{1}{8}\begin{bmatrix} 2 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This leads to

$$\mathbf{P} = \frac{1}{4}\begin{bmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J} \end{bmatrix},$$

where each $\mathbf{J}$ is a 4 by 4 block of 1's and each $\mathbf{0}$ is a 4 by 4 block of 0's.

2. $\mathbf{P}$ is unique, although $(\mathbf{X}'\mathbf{X})^-$ is not unique. $\mathbf{PY}$, the vector from O to the foot of the perpendicular from $\mathbf{Y}$ to $R(\mathbf{P}) = R(\mathbf{X})$, is also unique.
3. For $\mathbf{c}'\boldsymbol{\beta}$ to be estimable, the $\mathbf{c}'$ vector must be expressible as a linear combination of the rows of $\mathbf{X}$. There are 12 rows but only three are distinct, namely,

$$\mathbf{d}_1' = (1, -1, 1, -1),$$
$$\mathbf{d}_2' = (1, 0, 0, 0),$$
$$\mathbf{d}_3' = (1, 1, 1, 1).$$

Checking the five $\mathbf{c}'$ vectors, we see that

$$\mathbf{c}_1' = (1, 0, 1, 0) = (\mathbf{d}_1' + \mathbf{d}_3')/2; \quad \text{estimable.}$$

$$\mathbf{c}_2' = (0, 1, 0, 1); \quad \text{not estimable.}$$

$$\mathbf{c}_3' = (1, -1, 0, 0); \quad \text{not estimable.}$$

$$\mathbf{c}_4' = (0, 1, 0, 1) = (\mathbf{d}_3' - \mathbf{d}_1')/2; \quad \text{estimable.}$$

$$\mathbf{c}_5' = (1, 1, 1, 1) = \mathbf{d}_3; \quad \text{estimable.}$$

D. 1. The basic triangle has sides of lengths $OY = (786)^{1/2}$, $O\hat{Y} = (530)^{1/2}$, $Y\hat{Y} = 16$ in spaces of dimensions 3, 2, 1, respectively.

2. $786 = 530 + 256$, for SS.
   $3 = 2 + 1$, for df.
3. The point Y lies high above the plane of the estimation space. In fact, the F for regression is $(530/2)/256 = 1.04$, because the error MS is large compared with the regression MS.
4. Yes, it is $U_3 = 0$. That is why $Y(19, 13, 16)$ projects to $\hat{Y}(19, 13, 0)$.
5. $\mathbf{X}_{2.1}' = (1.2, -0.6, 0)$. The axes are now perpendicular and $530 = 405 + 125$ provides an orthogonal breakup of the regression SS in the rectangle formed by O, $\hat{Y}$, and the projections of $\mathbf{Y}$ on to the new axes.
6. If Y is moved to (19, 13, 2) from (19, 13, 16), the perpendicular to $\hat{Y}$ is the same but Y is much closer, that is, the error vector is shorter.
7. The new $F = (530/2)/4 = 66.25$
   Moral: A good regression is one where Y lies close to the estimation space.

E. 1.

$$\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{20}\begin{bmatrix} 19 & 3 & -3 & 1 \\ 3 & 11 & 9 & -3 \\ -3 & 9 & 11 & 3 \\ 1 & -3 & 3 & 19 \end{bmatrix}.$$

2. In four-dimensional space there is only one basis vector $\mathbf{v} = (a, b, c, d)'$, say, orthogonal

to the columns of **A** and it must satisfy the orthogonality conditions

$$a + b + c + d = 0,$$
$$-3a - b + c + 3d = 0,$$
$$a - b - c + d = 0.$$

So $\mathbf{v} = (1, -3, 3, -1)'$ or any multiple of this.

3. Obviously $\mathbf{Pv} = \mathbf{0}$ so $\mathbf{v}$ projects to the origin and **P** and **A** span the same space, that is, $R(\mathbf{P}) = R(\mathbf{A})$.

F. All are! We need $\mathbf{MM^-M} = \mathbf{M}$ to hold.
Suppose we call

$$\mathbf{M^-} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then

$$\mathbf{MM^-M} = (a + b + c + d)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

It follows that any such matrix with $a + b + c + d = 1$ will work, and those given all satisfy this condition.

G. The 10 df = 2 (model) + 3 (lack of fit) + 5 (pure error). The (**1**, **X**) vectors span the estimation space, five vectors $(-1, 1, 0, 0, \ldots, 0, 0), \ldots, (0, 0, \ldots, 0, 0, -1, 1)$ span the pure error space, and we need three more vectors to span the lack of fit space. Because of the form of the **X** vector (see below), we can use the orthogonal polynomials for $n = 5$ from the table in Section 22.2 and double them up. For example, from $\psi_2' = (2, -1, -2, -1, 2)$ we make the 10-dimensional vector $\phi_2' = (2, 2, -1, -1, -2, -2, -1, -1, 2, 2)$ and similarly for ψ_3' and ψ_4'. The **X** vector is a doubled up form of ψ_1', of course.

H. The row sums of **P** are given by **P1**. The vector **1** is already in the estimation space because β_0 is in the model, so **P1** = **1**, because when we project a vector already in the estimation space it remains as is. The column sums are $\mathbf{1'P} = (\mathbf{P1})'$, since $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$ is symmetric, and so $\mathbf{1'P} = \mathbf{1'}$.

Chapter 21

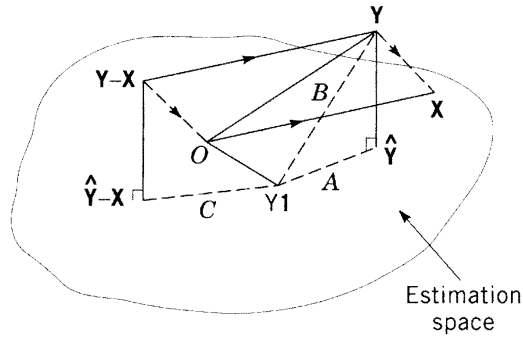
A. 1.

$$\frac{1}{20}\begin{bmatrix} 19 & 3 & -3 & 1 \\ 3 & 11 & 9 & -3 \\ -3 & 9 & 11 & 3 \\ 1 & -3 & 3 & 19 \end{bmatrix} - \frac{1}{20}\begin{bmatrix} 14 & 8 & 2 & -4 \\ 8 & 6 & 4 & 2 \\ 2 & 4 & 6 & 8 \\ -4 & 2 & 8 & 14 \end{bmatrix} = 0.25(\mathbf{X}_2, -\mathbf{X}_2, -\mathbf{X}_2, \mathbf{X}_2).$$

The dimensions are 3 − 2 = 1, so that, respectively, only 3, 2, and 1 column(s) are linearly independent.

2. Obviously $\mathbf{X}_2$ is $\omega^\perp \cap \Omega$ by definition, since it is the only vector in Ω that is orthogonal to the vectors in ω.
3. $\mathbf{X}_2$.
4. $\mathbf{P}_\Omega - \mathbf{P}_\omega$ has dimension (rank) $d = 1$ and is idempotent. A theorem says that idempotent matrices of rank d have eigenvalues 1 (d times over) and all other eigenvalues are zero. So the answer is 1, 0, 0, 0.

   More directly, we solve $\det(\mathbf{P}_\Omega - \mathbf{P}_\omega - \lambda\mathbf{I}) = 0$ to get $\lambda^3(\lambda - 1) = 0$.



Solution C1. The geometrical effect of changing the response from **Y** to **Y** − **X** in a regression on **1** and **X**. Note that the residual sum of squares is unchanged.

B. The least squares fit is $\mathbf{b} = (20.\dot{3}, 6.25, 2.41\dot{6})'$. $\mathbf{A} = (1, -2, 0)$, $\mathbf{c} = 4$.
$\mathbf{Z} = \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}' = \frac{4}{3}$, so $\mathbf{Z}^{-1} = 0.75$.
$\mathbf{c} - \mathbf{A}\mathbf{b} = -3.8\dot{3}$.
$\mathbf{b}_H = (19.375, 7.6875, 3.375)$.
Note that $b_{0H} - 2b_{1H} = 4$ as required.
(This exercise can also be done by substituting for the restriction into the model.)

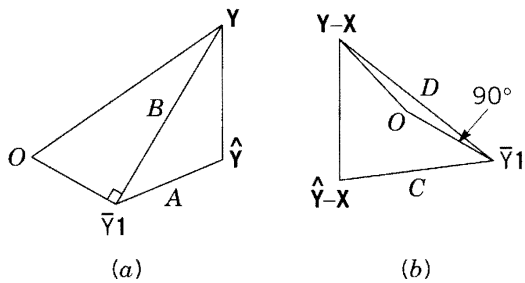
C. Look first at the vector parallelogram in Solution C1. We see that the **Y** − **X** vector is the same (in length and orientation) as the vector joining the tips of the **Y** and **X** vectors. It follows that the perpendiculars to the estimation space must also be equivalent vectors with the same lengths. The squares of these lengths are the residual sums of squares for regressions of (a) **Y** on **1** and **X** and (b) **Y** − **X** on **1** and **X**.

We now consider the R^2 statistics for the two regressions, shown in Solution C2. Obviously they can be different and R_a^2 could be larger or smaller than R_b^2 depending on the original geometry. Similar sorts of comments apply when there are k regression vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ $\mathbf{X}_k$ and $\mathbf{Y} - \Sigma\theta_i\mathbf{X}_i$ is used in the (b) regression.

As an extreme case of this moving of the response vector, consider **Y** − **Xb**, where we reduce **Y** by the particular combination of columns of **X** designated by the elements of the least squares vector **b**. Of course, the resulting response vector is the residual vector **e**; this is orthogonal to the estimation space and consequently $R^2 = 0$.

The author of the reference quoted suggests that use of R^2 is unsatisfactory because of this feature. We disagree. R^2, which measures the correlation between the selected response values and the predicted response values, accurately reflects the geometry of each regression situation shown. Looking at the geometry clarifies what the data manipulation has done.

G. The hypothesis $H_0: \beta_1 = 7, \beta_2 = 4$ cannot be tested because β_1 and β_2 are not individually estimable; **X** has dependent columns. There are 6 df available, 3 of which carry pure error. The remaining 3 df of the estimation space can be associated with the three averages obtained



Solution C2. The R^2 statistics for the two regressions: (a) $R_a^2 = A^2/B^2$, (b) $R_b^2 = C^2/D^2$.

from (a) the three estimates of $\beta_0 + \beta_1$ (namely, 101, 105, 94), (b) the two estimates of $\beta_0 + \beta_2$ (84, 88), and (c) the one estimate of $\beta_0 + \beta_3$ (32). Obviously, the comparison $\hat{\beta}_1 - \hat{\beta}_2 = \frac{1}{3}(101 + 105 + 94) - \frac{1}{2}(84 + 88) = 14$ provides an estimate of $\beta_1 - \beta_2$ and this has a standard error of $\{\hat{\sigma}^2(\frac{1}{3} + \frac{1}{2})\}^{1/2} = \{5\hat{\sigma}^2/6\}^{1/2}$. The pure error estimate of σ^2 is $\hat{\sigma}^2 = \{(1^2 + 5^2 + 6^2) + (2^2 + 2^2)\}/(2 + 1) = 70/3 = 23.333$, so that the standard error is 4.410. We can thus test whether $\beta_1 - \beta_2 = 7 - 4 = 3$. We get a t-statistic of $t = \{14 - (7 - 4)\}/4.410 = 2.494$. The corresponding $F(1, 3)$ statistic is $(2.494)^2 = 6.22$. The result is that we cannot reject $\beta_1 - \beta_2 = 3$ at the $\alpha = 0.05$ level.

This simple example (in which the n-dimensional space is divided up only into the estimation and pure error spaces) makes it clear that we can estimate only quantities that are linear combinations of the rows of $\mathbf{X}\boldsymbol{\beta}$. For some intricate algebra in the general case of nontestable hypotheses and some possible alternatives, see the papers mentioned in the source references.

H. 1.

$$70\mathbf{P} = \begin{bmatrix} 62 & 18 & -6 & -10 & 6 \\ & 26 & 24 & 12 & -10 \\ & & 34 & 24 & -6 \\ & & & 26 & 18 \\ \text{symmetric} & & & & 62 \end{bmatrix}.$$

2.

$$70\mathbf{P}_\omega = \begin{bmatrix} 42 & 28 & 14 & 0 & -14 \\ & 21 & 14 & 7 & 0 \\ & & 14 & 14 & 14 \\ & & & 21 & 28 \\ \text{symmetric} & & & & 42 \end{bmatrix}.$$

3. $\mathbf{P}_1 = \mathbf{P} - \mathbf{P}_\omega$, so

$$14\mathbf{P}_1 = \begin{bmatrix} 4 & -2 & -4 & -2 & 4 \\ & 1 & 2 & 1 & -2 \\ & & 4 & 2 & -4 \\ & & & 1 & -2 \\ \text{symmetric} & & & & 4 \end{bmatrix}.$$

4. All are true.

5. $\mathbf{P}_1\mathbf{x} = \mathbf{0}$. This is obvious because $\mathbf{x}$ is a vector in $\boldsymbol{\omega}$ and so is orthogonal to the space $R(\mathbf{P}_1)$, which is $\boldsymbol{\Omega} - \boldsymbol{\omega}$.

Chapter 22

A. Note that Company A data can be fitted in original Y or in $\ln Y$ form by orthogonal polynomials, whereas Company B data cannot be logged, because of the negative values, which are losses. Both sets of data given show severe cyclical trends, so that more than a straight line fit is needed. Data like these are notoriously difficult to predict into the future, and the polynomial, at best, explains what has happened. Prediction in other more stable cases is likely to be more useful.

B. $\hat{Y} = 11.2 + 0.9143(2X) + 1.5833\{1.5(X^2 - 35/12)\} + 0.006\{(5/3)(X^3 - 101X/20)\}$, with successive regression sums of squares 627.2, 58.51, 210.58, 0.006; and residual sum of squares 207.70 (2 df). The cubic term can be dropped, and the remaining terms rearranged in the form

$$\hat{Y} = 4.273 + 1.8286X + 2.3750X^2.$$

The sums of squares mentioned in the question are identical to the corresponding sums of squares of the same order given above.

C. Let Z = week number. The fitted equation is

$$\hat{Y} = 136.227 + 2.687Z + 0.167Z^2.$$

Source of Variation	df	SS	MS	F
Total	20	74,628.00		
a_0	1	48,609.80		
a_1	1	25,438.75	25,438.75	4,558.92*
a_2	1	489.00	489.00	87.63*
a_3	1	1.15	1.15	0.21
Residual	16	89.30	5.58	

D. 1. $\hat{Y} = -0.0037 - 2.8008t + 0.2314t^2$.

2. Residual analysis: There is no evidence for increasing the degree of the polynomial in t.

E. The analysis of variance is shown below.

Value of b_j	Source	df	SS
15.111	b_0	1	2055.11
1.866667	b_1	1	209.07
0.165945	b_2	1	76.33
0.072727	b_3	1	5.24
—	Residual	5	4.36
	Total	9	2350.00

The cubic term is not significant. A suitable model is

$$\hat{Y} = b_0 + b_1X + b_2(3X^2 - 20)$$
$$= 11.792 + 1.8667X + 0.4978X^2,$$

where the dyestuff levels are coded $X = -4, -3, -2, -1, 0, 1, 2, 3, 4$. This model accounts for $R^2 = 285.4/294.89 = 0.9678$ of the variation about the mean; $s^2 = 9.6/6 = 1.6$.

F. Partial solution.

	Estimated Coefficient	Before Compression		After Compression	
i	a_i	df	SS(a_i)	df	SS(a_i)
0	66.25	1	52,668.75	1	52,669
1	−0.02273	1	0.30		
2	−0.00774	1	0.72		
3	1.96018	1	19,780.16	1	19,780
4	0.01349	1	1.46		
5	−0.02137	1	7.26		
6	−0.06551	1	19.26		
Residual		5	13.09	10	42
Total		12	72,491.00	12	72,491

All the a_i eliminated in the compression are not significant at the $\alpha = 0.05$ level except a_6. However, use of a_6 increases R^2 only from 0.9979 to 0.9988, so that we rate a_6 as statistically significant, but numerically unimportant. The corresponding F-value of $19.36/\{13.09/5\} = 7.39$ only slightly exceeds $F(1, 5, 0.95) = 6.61$.

The selected model is $\hat{Y} = 66.25 + 1.96\{\frac{2}{3}(X^2 - \frac{85}{4})\}X$.

Durbin–Watson $d = 2.33$, not significant.

G. First read Section 22.3. Enter 322 observations as described there. Let Z_1, Z_2 be dummy variables such that $Z_1 = 1$ for category A, 0 otherwise, and $Z_2 = 1$ for category B, 0 otherwise. We fit the model $Y = \beta_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \beta_1$ (age) $+ \epsilon$ to obtain $\hat{Y} = 80.58 + 5.73Z_1 + 7.54Z_2 + 1.08$ (age). The adjusted ANOVA table is:

ANOVA

Source	df	SS	MS	F
$a_1, a_2 \mid b_0$	2	4,398.97	2,199.49	
$b_1 \mid b_0, a_1, a_2$	1	1,000.18	1,000.18	4.52
Residual	318	70,441.50	$s^2 = 221.51$	
Lack of fit	19	2,379.08	125.21	0.55
Pure error	299	68,062.42	227.63	
Total, corrected	321	75,840.64		

Lack of fit is not significant and we reject $\beta_1 = 0$ at the $\alpha = 0.05$ level, because $4.52 >$ (value interpolated between 3.92 and 3.84 in F-table). $R^2 = (4{,}398.97 + 1{,}000.18)/75{,}840.64 = 0.0712$. This value (only 7% of the variation about the mean is explained), and the fact that the F-value 4.52 is only slightly bigger than the percentage point with which it is compared, indicate that the fitted equation is not a very useful one. Tests of $\alpha_1 = 0$, $\alpha_2 = 0$, and $\alpha_1 - \alpha_2 = 0$ confirm the fact that the two groups of men weigh about the same and are heavier than the group of women. (The appropriate t-values are 2.55, 3.82, and 0.84.) We conclude then that, although weight increases with age in these data, age is not a good predictor of it and that, although there *are* weight differences in the diet categories, they seem to be confounded with gender and may well be due to that, rather than diet.

Chapter 23

A. 1. Both methods of analysis will yield the following analysis of variance table:

ANOVA

Source of Variation	df	SS	MS	F	$F_{0.95}$
Total SS	18	24,403.750			
Mean	1	22,352.027			
Corrected total	17	2,051.723			
Steam pressure	2	963.721	481.861	11.728*	4.26
Blowing time	2	37.481	18.741	0.456	4.26
Interaction	4	680.756	170.189	4.142*	3.63
Pure error	9	369.765	41.085		

Thus both steam pressure and the interaction of steam pressure and blowing time are statistically significant.

The regression model obtained for this problem is

$$\hat{Y} = 35.239 + 4.794X_1 - 10.339X_2 - 0.206X_3 + 1.861X_4 + 5.773X_1X_3 - 2.394X_1X_4 + 6.506X_2X_3 - 3.361X_2X_4,$$

where X_1, X_2 are dummy variables for steam pressure defined as follows:

X_1	X_2	
1	0	= 10 pounds steam pressure
0	1	= 20 pounds steam pressure
−1	−1	= 30 pounds steam pressure

and X_3, X_4 are dummy variables for blowing time defined as follows:

X_3	X_4	
1	0	= blowing time 1 hour
0	1	= blowing time 2 hours
−1	−1	= blowing time 3 hours

Residual analysis indicates that the experiments have a much smaller variance at the low level of steam pressure. Since only two repeat runs are available at each set of conditions, the analysis is not necessarily invalid, but further investigation is clearly indicated. To appreciate the interaction effect, a table of mean values could be examined, or the mean responses could be plotted against blowing time for each level of steam pressure.

2.

ANOVA

Source	df	SS	MS	F	$F_{0.95}$
Total	18	417.000			
Mean	1	34.722			
Corrected total	17	382.278			
Premix speed	2	24.111	12.055	1.080	4.26
Finished mix speed	2	7.444	3.722	0.333	4.26
Interaction	4	250.223	62.556	5.602*	3.63
Pure error	9	100.500	11.167		

Thus only the interaction term is statistically significant in this experiment.

The fitted regression model obtained for this problem is

$$\hat{Y} = 1.39 - 1.39X_1 + 1.44X_2 - 0.89X_3 + 0.61X_4 + 1.39X_1X_2 + 1.89X_1X_4 + 4.56X_2X_3 - 1.44X_2X_4,$$

where X_1, X_2 are dummy variables for premix speeds, and X_3, X_4 are dummy variables for finished mix speeds as follows:

X_1	X_2		X_3	X_4	
1	0	= premix speed 1	1	0	= finished mix speed 1
0	1	= premix speed 2	0	1	= finished mix speed 2
−1	−1	= premix speed 3	−1	−1	= finished mix speed 3

Residual analysis indicates a much larger variability in the observations at the lowest level of finished mix speed. This should be investigated.

The significant interaction is most easily seen in the following table:

Table of Mean Responses

		X_2		
		−1	0	+1
X_1	−1	0.5	2.5	−3
	0	6.5	2.0	0
	+1	−5.5	1.5	8

B. 1a.

ANOVA

Source	df	SS	MS	F
Total	16	921.0000		
Regression	4	881.2500	220.3125	
b_0	1	798.0625	798.0625	
b_1	1	18.0625	18.0625	5.45*
b_2	1	60.0625	60.0625	18.13*
b_{12}	1	5.0625	5.0625	1.53 NS
Residual	12	39.7500	3.3125	

b. (1) Regression equation is significant. $F(3, 12) = 8.37$*.
(2) All except b_{12}.

c. $R^2 = 67.67\%$.

2a.

$$b_0 = \frac{798.0625}{113} = 7.0625,$$

$$b_1 = \frac{18.0625}{17} = 1.0625,$$

$$b_2 = \frac{60.0625}{31} = 1.9375,$$

$$b_{12} = \frac{5.0625}{-9} = -0.5625.$$

$\therefore \hat{Y} = 7.0625 + 1.0625X_1 + 1.9375X_2 - 0.5625X_1X_2$.

b. $s^2(\mathbf{X'CX}) = 0.6875$

$(3.3125)(\mathbf{X'CX}) = 0.6875$

$$\mathbf{X'CX} = \frac{0.6875}{3.3125} = 0.207547.$$

Variance of a single observation $= s^2(1 + \mathbf{X'CX}) = (3.3125)(1 + 0.207547)$
$= 4.0000.$

c. $\hat{Y}$ is 54 at $X_1 = 70$ and $X_2 = 150$.
$V(\hat{Y}) = 0.6875$.
$\therefore$ Confidence limits for the true mean value of Y are

$$\hat{Y} \pm t(12, 0.95)\text{se}(\hat{Y}) = 54 \pm (2.179)\sqrt{0.6875}$$
$$= 54 \pm (2.179)(0.8292)$$
$$= 54 \pm 1.8068.$$

3a. The prediction equation determined by this analysis is

$$\hat{Y} = 7.0625 + 1.0625X_1 + 1.9375X_2 - 0.5625X_1X_2.$$

b. The interaction term, X_1X_2, is not statistically significant at an α level of 0.05. Thus there is some doubt as to the validity of the assumed model. However, this doubt is based on a small number of observations, $n = 16$, and the original model was based on the knowledge of the chemist. Before considering dropping the X_1X_2 term, more experimental work should be done. This is an example of the statement: "Even though a variable is nonsignificant statistically, it should *not* be considered to have a zero effect on the result of the experiment."

C. Solution not provided.

D. Our solution uses Section 23.5. Transformed data:

Sample No.	1	2	3	4	5
$Y_{ij} - 1430$	260	120	195	295	100
	150	15	20	120	115
	315	215	80	0	135
	255	115		15	90
$J_i\overline{Y}_i$	980	465	295	430	440
J_i	4	4	3	4	4
$\overline{Y}_i = b_i$	245	116.25	98.33	107.5	110

From Eqs. (23.5.5) and (23.5.6), residual SS is 524,450 − 417,788.6 = 106,661.4. To test equality of groups we obtain, from Eqs. (23.5.10) and (23.5.6), the SS due to H_0 as (524,450 − 358,531.6) − 106,661.4 = 59,257. F = (59,257/4)/(106,661.4/14) = 1.94 < $F(4, 14, 0.95)$ = 3.11. Do not reject H_0. We conclude that the data do not indicate that the strength of the concrete is significantly affected by varying the amount of coal dust added to the sand.

Working with $Y_{ij} - 1430$ affects only the total SS and the correction factor. The between group and within group SS values are unaffected.

E. 1. We recode the factor combination *bcd* as $X_1 = -1$, $X_2 = X_3 = X_4 = 1$, putting −1 when a letter does *not* appear in the combination and 1 when it *does* appear. We can fit a model with terms in β_0, $\beta_i X_i$, and $\beta_{ij} X_i X_j$, $i \neq j = 1, 2, 3, 4$. Only X_2 and X_3 are worth retaining and the equation of choice is $\hat{Y} = 39.68 - 0.6688X_2 + 0.8937X_3$ with $R^2 = 0.7229$.

2. The data come from a four-way classification, each classification having two levels, with one observation per cell. The model of Section 23.6 would have to be extended from two-way to four-way to treat this in an ANOVA framework.

F. 1. The original setup produces a singular $\mathbf{X'X}$.

2. Any valid dummy setup can be used. Suppose we choose two dummies (Z_1, Z_2) = (−1, 1), (0, −2), and (1, 1) for groups 1, 2, and 3, respectively. This produces orthogonal **X**-columns because there are the same numbers of observations in each group. (These dummies are orthogonal polynomials; see Section 22.1.) Then $\mathbf{b}$ = (6, 2, 0) and the orthogonal SS breakup of $\mathbf{b'X'Y}$ is 324 + 24 + 0 = 348 (3 df). The residual SS = 366 − 348 = 18 (5 df). The F-statistic for testing equality of groups is thus $F(2, 5)$ = {(24 + 0)/2}/(18/5) = 3.33 < 5.79, not significant. The F-value does not depend on the dummy system chosen, as the reader can confirm by trying the problem another way.

G. For both (1) and (2), many choices of dummies can be made, for example:

$$
\begin{aligned}
(X_1, X_2) &= (1, 0) && \text{for tube 1,}\\
&= (0, 1) && \text{for tube 2,}\\
&= (0, 0) && \text{for tube 3,}\\
(X_3, X_4, X_5, X_6) &= (1, 0, 0, 0) && \text{for level 1,}\\
&\quad (0, 1, 0, 0) && \text{for level 2,}\\
&\quad (0, 0, 1, 0) && \text{for level 3,}\\
&\quad (0, 0, 0, 1) && \text{for level 4,}\\
&\quad (0, 0, 0, 0) && \text{for level 5.}
\end{aligned}
$$

Thus, if we copy the observations row by row into the $\mathbf{Y}$ vector, we get for corresponding

rows of the **X** matrix and the original **Y** vector, the following:

Subscript of X							
0	1	2	3	4	5	6	Y
1	1	0	1	0	0	0	78.4
1	0	1	1	0	0	0	84.5
1	0	0	1	0	0	0	62.1
1	1	0	0	1	0	0	10.6
1	0	1	0	1	0	0	77.5
1	0	0	0	1	0	0	94.4
1	1	0	0	0	1	0	62.7
1	0	1	0	0	1	0	4.4
1	0	0	0	0	1	0	24.8
1	1	0	0	0	0	1	92.5
1	0	1	0	0	0	1	17.5
1	0	0	0	0	0	1	0.9
1	1	0	0	0	0	0	11.6
1	0	1	0	0	0	0	0.9
1	0	0	0	0	0	0	20.4

1. We fit three models:

$$\hat{Y} = 40.52 + 10.64X_1 - 3.56X_2,$$

$$\hat{Y} = 10.97 + 64.03X_3 + 49.87X_4 + 19.67X_5 + 26.00X_6,$$

$$\hat{Y} = 8.61 + 10.64X_1 - 3.56X_2 + 64.03X_3 + 49.87X_4 + 19.67X_5 + 26.00X_6.$$

The respective regression sums of squares (given b_0) are

546 (2 df),

7672 (4 df),

8218 (6 df),

and the residual mean square is $s^2 = 1294$(8 df). It follows that the "between tubes" $F_{2,8} = \{(8218 - 7672)/2\}/1294 = 0.21$ and the "between levels" $F_{4,8} = \{(8218 - 546)/4\}/1294 = 1.48$.

2. The details are exactly parallel. Now, however, $F_{2,8} = \{(17300 - 17016)/2\}/159.2 = 0.89$ and $F_{4,8} = \{(17300 - 284)/4\}/159.2 = 26.72$. So now there are differences between levels, in this (fake) data set.

H. The four regressions explain (1) 98.7% (2) 99.1% (3) 98.5%, and (4) 99.1% of the variation about the mean $\overline{Y}$. Essentially the ranges of Y and X_1 are such that the transformations make little difference in explanatory value, at least in this set of data. All equations appear plausible ones, and the residuals look most satisfactory for the (1) fit. Equation (1) is

$$\hat{Y} = 0.3688 + 0.0340X_1 - 0.1852X_2 + 0.0998X_3 + 0.002313X_2X_3.$$

The dummy variable system suggested makes the X_2, X_3, and X_{23} columns, which carry the 3 df among the four columns DD, DS, SD, and SS, orthogonal to the 1 and X_1 columns. The b_2 and b_3 coefficients are very highly significant, the b_{23} is not significant. Additive main effects are thus indicated. Hatchabilities are best for the DD combination followed by DS, SD, and SS, and the differences are significant ones.

Chapter 24

A. $\hat{\theta} = 0.20345$, $S(\hat{\theta}) = 0.00030$; $0.179 \leq \theta \leq 0.231$.

B. $\hat{\theta} = 0.20691$, $S(\hat{\theta}) = 0.01202$; $0.190 \leq \theta \leq 0.225$.

C. $(\hat{\alpha}, \hat{\beta}) = (0.38073, 0.07949)$, $S(\hat{\alpha}, \hat{\beta}) = 0.00005$;
$S(\alpha, \beta) = 0.001$ (or 0.0009 to one more decimal place)

D. $(\hat{\alpha}, \hat{\beta}, \hat{\rho}) = (72.4326, 28.2519, 0.5968)$, $S(\hat{\alpha}, \hat{\beta}, \hat{\rho}) = 3.5688$; $S(\alpha, \beta, \rho) = 106.14$.

E. $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (5.2673, 8.5651, 294.9931)$, $S(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = 1718.2108$; $S(\alpha, \beta, \gamma) = 3400$.

F. Write the model as

$$Y = \theta + \alpha(X_1X_3 + \gamma X_1) + \beta(X_2X_3 + \gamma X_2) + \epsilon.$$

Fix γ, solve for $\hat{\theta}$, $\hat{\alpha}$, $\hat{\beta}$. Repeat for other values of γ, iterating on γ until minimum $S(\hat{\theta}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ is obtained.

G. $(115.2, 2.310, -22.022)$, $S(\hat{\boldsymbol{\theta}}) = 7.0133$. $S(\boldsymbol{\theta}) = 209.0$.

H. $(0.00376, 27{,}539)$, $S(\hat{\boldsymbol{\theta}}) = 0.00429326$. $S(\boldsymbol{\theta}) = 0.00507559$.

I. $(0.00366, 27{,}627)$, $\hat{S}(\hat{\boldsymbol{\theta}}) = 0.000754$. $S(\boldsymbol{\theta}) = 0.00204$.

J. $(3.57, 12.77, 0.63)$, $S(\hat{\boldsymbol{\theta}}) = 0.00788$. $S(\boldsymbol{\theta}) = 0.01665$.

K. $(0.480, 1.603)$, $S(\hat{\boldsymbol{\theta}}) = 7.301$. $S(\boldsymbol{\theta}) = 53.8$.
The long thin contour indicates that a large number of pairs of values (θ_1, θ_2) are almost as suitable as the actual least squares values.

L.

No.	$\hat{\theta}_1$	$\hat{\theta}_2$	$S(\hat{\boldsymbol{\theta}})$	$S(\boldsymbol{\theta})$
1	205.25	0.431	252	835.8
2	2.498	0.202	0.0262	0.0712
3	892.67	0.245	3,376.5	15,093
4	25.475	0.323	17.004	76.007
5	13.809	0.398	0.866	3.871
6	19.903	0.441	3.716	16.61
7	213.82	0.547	1,168	5,221
8	19.142	0.531	25.99	116.18
9	10,525	0.569	68,349	1,366,980

M. $\hat{\alpha} = 0.009229$, with $\text{se}(\hat{\alpha}) = 0.010725$.
$\hat{\beta} = 1.825450$, with $\text{se}(\hat{\beta}) = 0.234714$.
$s = 19.3049$, based on 41 df.
The estimates are highly correlated, with $r = 0.999$.

(Note that, if we now take natural logarithms, we obtain $\ln \hat{M} = \ln \hat{\alpha} + \hat{\beta} \ln T = -4.6854 + 1.82545 \ln T$. This compares with the linear least squares fit with coefficients $a = -5.728$, $b = 2.031$ in Exercise 13E. One could next examine the residuals from each fitted equation, the nonlinear as fitted and the linear, and make an assessment of which fit seemed to be preferable.)

N. (Partial solution)
Tree No. 1

$$\begin{aligned}
&(24.7.2) \quad \hat{\alpha} = 268.3, \quad \hat{\beta} = 0.9478, \quad \hat{k} \times 10^4 = 4.740,\\
&(24.7.5) \quad \hat{\alpha} = 154.1, \quad \hat{\beta} = 5.643, \quad \hat{k} \times 10^3 = 2.759,\\
&(24.7.8) \quad \hat{\delta} = 5.032, \quad \hat{\beta} = 5.792, \quad \hat{k} \times 10^3 = 2.814,\\
&(24.7.10) \quad \hat{\alpha} = 172.2, \quad \hat{\beta} = 2.813, \quad \hat{k} \times 10^3 = 1.626.
\end{aligned}$$

Visually, (24.7.8) looks best although this may be due to the reduced response range induced by the ln transformation. The first model (24.7.2) does not pick up well the "S-shaped" tendency in the data. With only seven observations, however, the data set is too small to permit too definite conclusions.

Tree No. 2

$$(24.7.2)\quad \hat{\alpha} = 519.3,\quad \hat{\beta} = 0.9820,\quad \hat{k} \times 10^4 = 3.208,$$

$$(24.7.5)\quad \hat{\alpha} = 218.9,\quad \hat{\beta} = 8.225,\quad \hat{k} \times 10^3 = 3.010,$$

$$(24.7.8)\quad \hat{\delta} = 5.398,\quad \hat{\beta} = 8.228,\quad \hat{k} \times 10^3 = 2.962,$$

$$(24.7.10)\quad \hat{\alpha} = 248.4,\quad \hat{\beta} = 2.645,\quad \hat{k} \times 10^3 = 1.703.$$

O.

β	2.00	2.01	2.02	2.03
α	−1.03	−1.06	−1.09	−1.12
SS	0.0275	0.0134	0.0126	0.0255

Solution is $\hat{\alpha} = -1.09$, $\hat{\beta} = 2.02$, correct to the second decimal place in β.

P. The range of γ values $0.68 \le \gamma \le 0.70$ is best, with $\hat{\alpha} = 43.2$, $\hat{\beta} = 123.2$ at $\gamma = 0.69$ and minor variations nearby.

For ϕ, $0.93 \le \phi \le 0.96$ is best with minor variations about values $\hat{\delta} = 34.0$, $\hat{\theta} = 106.0$ at $\phi = 0.94$.

Q.

$$\hat{\alpha} = 0.6588, \qquad \text{se}(\hat{\alpha}) = 0.250.$$

$$\hat{\beta} = 1.0272, \qquad \text{se}(\hat{\beta}) = 0.040.$$

$$\hat{\gamma} = 0.5184, \qquad \text{se}(\hat{\gamma}) = 0.011.$$

$$\text{corr}\,(\hat{\alpha}, \hat{\beta}) = -0.963, \qquad \text{corr}\,(\hat{\alpha}, \hat{\gamma}) = -0.268, \qquad \text{corr}\,(\hat{\beta}, \hat{\gamma}) = 0.$$

$$s = 9176.93\ (13\ \text{df}).$$

R.

Group	Equation $\hat{Y}$ =
185 kg B	$0.8822 + 2.2290\,(0.7512)^X$
275 kg B	$0.8444 + 1.7083\,(0.796)^X$
275 kg H × B	$0.8079 + 1.7701\,(0.7483)^X$
450 kg $\frac{3}{4}$ B	$0.7066 + 0.5700\,(0.7905)^X$
450 kg $\frac{3}{4}$ H	$0.6699 + 0.2448\,(0.8458)^X$

S. (Main details.) The parameter estimates are:

Sample	$\hat{\delta}$	$\hat{\beta}$	$\hat{k}$
8	4.96	78.76	0.15
14	5.08	222.36	0.21
16	4.88	103.70	0.15
22	4.79	128.24	0.17
24	5.71	23.84	0.07
32	4.88	217.93	0.20
52	5.18	50.66	0.12
54	4.95	49.17	0.13
All	4.96	89.32	0.15

Although samples 14, 24, and 32 do catch the eye, the various plots reveal an overall consistency in these data sets.

Individual readings would provide a pure error estimate of σ^2 and enable us to check its (assumed) constancy; because the *same* number of observations determines each mean, the actual parameter estimates would not change.

T. The main points are as follows.

$$\hat{\boldsymbol{\theta}} = (0.10579, 1.7007)'.$$

$$\text{Pure error SS} = 1.998 \times 10^{-4} \text{ (6 df)}.$$

$$\text{Lack of fit SS} = 1.08 \times 10^{-6} \text{ (4 df)}.$$

The model seems satisfactory. The best single run is at the largest X value practically feasible; the best pair are both at that same value.

U. We can use the initial estimates $\theta_{10} = 0.11$, $\theta_{20} = 1.7$ from the results of Exercise T. Substituting these values in the model function and (for example) putting in (X, Y) values (2, 0.0571) and (0.2, 0.0108) derived from the first and last data pairs, gives two equations from which θ_4 can be eliminated to give $0.06834 = 0.0529(2^{\theta_3}) - 0.0992(0.2^{\theta_3})$. An iterated solution is $\theta_{30} = 0.93$, whereupon either equation can be solved, or both solutions can be averaged, to give $\theta_{40} = -0.075$ approximately. The various fits give the following results.

Model with These θ's	$S(\hat{\boldsymbol{\theta}})$	df	$s = \text{se}(Y)$
$\theta_1, \theta_2, \theta_3, \theta_4$	0.0002	8	0.0050
$\theta_1, \theta_2, \theta_3(\theta_4 = 0)$	0.0002	9	0.0047
$\theta_1, \theta_2(\theta_3 = \theta_4 = 0)$	0.0002	10	0.0045

We recall from T that the pure error SS = 0.0002 (6 df). All the $S(\hat{\boldsymbol{\theta}})$ and the pure error sum of squares are about the same size, 0.0002, so the extra SS are very tiny, their small differences being reflected in the minor changes in the last column. Obviously the smallest model is satisfactory. For this, $\hat{\theta}_1 = 0.105643$ (se = 0.018), $\hat{\theta}_2 = 1.70269$ (se = 0.476).

V. Data Set 1 has the characteristics of the data in Figure 24.2a. The approximate 95% confidence contour is very large and $\hat{\theta}_1$ and $\hat{\theta}_2$ are highly correlated. Data Set 2 is like the additional data in Figure 24.2b. It provides much better estimation and a very small approximate 95% confidence contour. As an additional exercise, use all nine observations, omitting one of the duplicates at $t = 1$, and compare the three analyses.

W. By ignoring the error and forcing the curve through any of the data points, we can solve $Y = \theta_{10}\{1 - \exp(-t/4)\}$ for θ_{10}. Substitution of each data point in turn gives $\theta_{10} = 2.13, 2.22, 2.25, 2.45$. Any of these could be used; so could the average 2.26 or 2.3.

X. *Variety No. 1 Data.* If we initially ignore the error and just work with the model function we see that

$$\omega^{-\theta_1} = \theta_2 + \theta_3 X^{\theta_4}.$$

Thus, if we knew θ_1 and θ_4, the plot of $\omega^{-\theta_1}$ versus X^{θ_4} would be a straight line; see the discussion in point 4 of the last subsection of Section 24.2.

In fact, $\theta_1 = \theta_4 = 1$ gives an excellent line with intercept 0.0055 and slope 0.0016, approximately, fitted by eye. So we can take $\boldsymbol{\theta}_0 = (1, 0.0055, 0.0016, 1)$. Nonlinear estimation from there gives $\hat{\boldsymbol{\theta}} = (1.1528, 1.7793 \times 10^{-3}, 1.0189 \times 10^{-3}, 1.0046)$ with $S(\hat{\boldsymbol{\theta}}) = 0.0225461$. The approximate linearized individual confidence bands are

$$\theta_1\text{:}\quad (-2.9614, 5.2671),$$

$$\theta_2\text{:}\quad (-3.2602 \times 10^{-2}, 3.6161 \times 10^{-2}),$$

$$\theta_3\text{:}\quad (-2.1540 \times 10^{-2}, 2.3578 \times 10^{-2}),$$

$$\theta_4\text{:}\quad (-2.0443, 4.0535).$$

It is clear that the parameter estimation is poor and possible overparameterization is indicated by the off-diagonal elements of the linearized parameter estimates correlation matrix, all of which are close to ± 1. The $S(\boldsymbol{\theta})$ contour is elongated and many different final sets of estimates for the parameters, all with more or less the same $S(\hat{\boldsymbol{\theta}})$, are possible. Initial values that differ only slightly from those above will give rise to alternative $\hat{\boldsymbol{\theta}}$ values. The

plots of residuals do not reveal any abnormalities. In view of the plot that showed that values $\theta_1 = \theta_4 = 1$ seemed perfectly reasonable and the fact that the model seems overparameterized anyway, we could try to fit the simpler model

$$\ln W_i = -\ln(\alpha + \beta X_i) + \epsilon_i$$

using the same initial values (in revised notation), $\alpha_0 = 0.0055$ and $\beta_0 = 0.0016$. This leads to the solution $(\hat{\alpha}, \hat{\beta}) = (5.309 \times 10^{-3}, 1.587 \times 10^{-3})$ with $S(\hat{\alpha}, \hat{\beta}) = 0.026239$. Previously we had $S(\hat{\boldsymbol{\theta}}) = 0.0225461$ so that the increase is small considering that two parameters have been dropped. The approximate linearized confidence intervals are now

$$\alpha: \quad (4.345 \times 10^{-3}, 6.272 \times 10^{-3}),$$

$$\beta: \quad (1.587 \times 10^{-3}, 1.705 \times 10^{-3}),$$

neither of which includes zero. Overall, the fit seems adequate.

Y. To get initial values, one can do something like the following, for example. Substitute into the model function for $(X_1, X_2, Y) = (0.25,0,0.310)$ and $(5,0,1.075)$ where the Y's are roughly guessed from the data. This gives $\theta_1 = 0.310(4\theta_2 + 1) = 1.075(0.2\theta_2 + 1)$ which can be solved for $\theta_2 = 0.692$, $\theta_1 = 1.168$. Now set $(X_1, X_2, Y, \theta_1, \theta_2) = (5,0.5,0.280,1.168,0.692)$ into the model function and solve for $\theta_3 = 0.0228$. We thus obtain $\boldsymbol{\theta}_0 = (1.17,0.69,0.023)'$.

	Parameter Estimates and Linearized 95% Limits		
	Lower	$\hat{\theta}$	Upper
θ_1	1.235	1.260	1.284
θ_2	0.792	0.847	0.901
θ_3	0.0254	0.0272	0.0290

$S(\hat{\boldsymbol{\theta}}) = 0.01370$.

Correlations of parameter estimates in order 12, 13, 23: 0.866, 0.466, 0.673.

Source	SS	df	MS
Lack of fit	0.00761	15	0.000507
Pure error	0.00609	28	0.000218
Residual, $S(\hat{\boldsymbol{\theta}})$	0.01370	43	

The "approximate $F(15, 28)$" ratio is 2.33 which exceeds $F(15, 28, 0.95) = 2.04$, indicating lack of fit. However, the three groups of data for which $X_2 = 0,0.2$, and 0.5 contribute pure error sums of squaresof 0.005490 (16 df, 0.000343 per df), 0.000093 (6 df, 0.000016 per df), and 0.000508 (6 df, 0.000085 per df), respectively. The pairs of observations are considerably less variable than the groups of three and four observations, casting some doubt on what is only an approximate F-test in the best of circumstances.

Z. 1. $\hat{Q} = -2.08302$, $\hat{p} = 0.194833$, $\hat{N} = 76.8884$ with $\text{se}(\hat{Q}) = 6.113$, $\text{se}(\hat{p}) = 0.1858$, $\text{se}(\hat{N}) = 93.8601$. The correlations are 0.965 $(\hat{Q}, \hat{p})$, − 0.990 $(\hat{Q}, \hat{N})$, and −0.984 $(\hat{p}, \hat{N})$. The residual se = 2.42384 (7 df). $S(\hat{\boldsymbol{\theta}}) = 41.1252$. The negative value for $\hat{Q}$ is meaningless and the correlations are large, so it makes sense to set $Q = 0$ and refit.

2. Set $Q = 0$, then $\hat{p} = 0.278715$, $\hat{N} = 48.9678$ with $\text{se}(\hat{p}) = 0.0765$ and $\text{se}(\hat{N}) = 9.134$. The correlation between estimates is −0.733. The residual se = 2.30168 (8 df). $S(\hat{\boldsymbol{\theta}}) = 42.3819$. In this fit the residual sum of squares is only slightly higher and the residual se is lower. The $\hat{N}$ of about 49 makes sense, the se values of the coefficients are relatively small and the prediction equation $\hat{Y} = (0.278715)(48.9678)\exp\{-0.278715t\}$ produces estimates of 12, 9, 7, 5, 4, 3, 2, 2, 1, 1 for the times given—not bad in the circumstances.

Chapter 25

A. Various solutions are obtained depending on the robust method used and the tuning constants selected. Data sets with outliers tend to be better fitted by robust methods compared with least squares and the differences in the fits provide important clues to the possibility of adjusting the data.

B. The largest observation in the data set is the cause of the difference seen, because it "pulls the least squares line up toward itself." The robust fits accord the largest observation less control.

Chapter 26

A. No solution provided. Results will vary.

B. No solution provided. Results will vary.

Solutions to True/False Questions

1–101. All statements are true. They can be amended to become false if used in quizzes. (It was felt unwise to present false statements here and risk a misunderstanding.)