Chapter 7

7.1 $$\hat{\beta} = (Z'Z)^{-1}Z'y = \frac{1}{120}\begin{bmatrix} 120 & -10 \\ -10 & 1 \end{bmatrix}\begin{bmatrix} 72 \\ 872 \end{bmatrix} = \frac{1}{15}\begin{bmatrix} -10 \\ 19 \end{bmatrix} = \begin{bmatrix} -.667 \\ 1.267 \end{bmatrix}$$

$$\hat{y} = Z\hat{\beta} = \frac{1}{15}\begin{bmatrix} 180 \\ 85 \\ 123 \\ 351 \\ 199 \\ 142 \end{bmatrix} = \begin{bmatrix} 12.000 \\ 5.667 \\ 8.200 \\ 23.400 \\ 13.267 \\ 9.467 \end{bmatrix} ; \quad \hat{\varepsilon} = y - \hat{y} = \begin{bmatrix} 15 \\ 9 \\ 3 \\ 25 \\ 9 \\ 13 \end{bmatrix} - \begin{bmatrix} 12.000 \\ 5.667 \\ 8.200 \\ 23.400 \\ 13.267 \\ 9.467 \end{bmatrix} = \begin{bmatrix} 3.000 \\ 3.333 \\ -5.200 \\ 1.600 \\ -6.267 \\ 3.533 \end{bmatrix}$$

Residual sum of squares: $\hat{\varepsilon}'\hat{\varepsilon} = 101.467$

Fitted equation: $\hat{y} = -.667 + 1.267\, z_1$

7.2 Standardized variables

z_1	z_2	y
-.292	-1.088	.391
-1.166	-.726	-.391
-.817	-.726	-1.174
1.283	.363	1.695
-.117	.726	-.652
1.108	1.451	.130

Fitted equation:

$$\hat{y} = 1.33 z_1 - .79 z_2$$

Also, prior to standardizing the variables, $\bar{z}_1 = 11.667$, $\bar{z}_2 = 5.000$ and $\bar{y} = 12.000$; $\sqrt{s_{z_1z_1}} = 5.716$, $\sqrt{s_{z_2z_2}} = 2.757$ and $\sqrt{s_{yy}} = 7.667$.

The fitted equation for the original variables is

$$\frac{\hat{y} - 12}{7.667} = 1.33\left(\frac{z_1 - 11.667}{5.716}\right) - .79\left(\frac{z_2 - 5}{2.757}\right)$$

$$\hat{y} = .43 + 1.78 z_1 - 2.19 z_2$$

7.3 Follow hint and note that $\hat{\varepsilon}^* = Y^* - \hat{Y}^* = V^{-1/2}Y - V^{-1/2}Z\hat{\beta}_W$ and $(n-r-1)\sigma^2 = \hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*$ is distributed as χ^2_{n-r-1}.

7.4 a) $V = I$ so $\hat{\underset{\sim}{\beta}}_W = (\underset{\sim}{z}'\underset{\sim}{z})^{-1}\underset{\sim}{z}'\underset{\sim}{y} = (\sum_{j=1}^{n} z_j y_j)/(\sum_{j=1}^{n} z_j^2)$.

b) V^{-1} is diagonal with $j^{\underline{th}}$ diagonal element $1/z_j$ so

$$\hat{\underset{\sim}{\beta}}_W = (\underset{\sim}{z}'V^{-1}\underset{\sim}{z})^{-1}\ \underset{\sim}{z}'V^{-1}\underset{\sim}{y} = (\sum_{j=1}^{n} y_j)/(\sum_{j=1}^{n} z_j)$$

c) V^{-1} is diagonal with $j^{\underline{th}}$ diagonal element $1/z_j^2$ so

$$\hat{\underset{\sim}{\beta}}_W = (\underset{\sim}{z}'V^{-1}\underset{\sim}{z})^{-1}\underset{\sim}{z}'V^{-1}\underset{\sim}{y} = (\sum_{j=1}^{n} (y_j/z_j))/n$$

7.5 Solution follows from Hint.

7.6 a) First note that $\Lambda^- = \text{diag}[\lambda_1^{-1},\dots,\lambda_{r_1+1}^{-1},0,\dots 0]$ is a generalized inverse of Λ since

$$\Lambda\Lambda^- = \begin{bmatrix} I_{r_1+1} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{so} \quad \Lambda\Lambda^-\Lambda = \begin{bmatrix} \lambda_1 & & & & & 0 \\ & \ddots & & & & \\ & & \lambda_{r_1+1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{bmatrix} = \Lambda$$

Since

$$Z'Z = \sum_{i=1}^{p} \lambda_i \underset{\sim}{e}_i \underset{\sim}{e}_i' = P\Lambda P'$$

$$(Z'Z)^- = \sum_{i=1}^{r_1+1} \lambda_i^{-1}\ \underset{\sim}{e}_i \underset{\sim}{e}_i' = P\Lambda^- P'$$

with $PP' = P'P = I_p$, we check that the defining relation holds

$$(Z'Z)(Z'Z)^-(Z'Z) = P\Lambda\underbrace{P'(P}\Lambda^-\underbrace{P')P}\Lambda P'$$

$$= P\Lambda\Lambda^-\ \Lambda P'$$

$$= P\Lambda P' = Z'Z$$

b) By the hint, if $Z\hat{\underset{\sim}{\beta}}$ is the projection, $\underset{\sim}{0} = Z'(\underset{\sim}{y} - Z\hat{\underset{\sim}{\beta}})$ or $Z'Z\hat{\underset{\sim}{\beta}} = Z'\underset{\sim}{y}$. In c), we show that $Z\hat{\underset{\sim}{\beta}}$ is the projection of $\underset{\sim}{y}$.

c) Consider $\underset{\sim}{q}_i = \lambda_i^{-1/2} Z\underset{\sim}{e}_i$ for $i = 1,2,\ldots,r_1+1$. Then

$$Z(Z'Z)^{-}Z' = Z\left(\sum_{i=1}^{r_1+1} \lambda_i^{-1} \underset{\sim}{e}_i\underset{\sim}{e}_i'\right)Z' = \sum_{i=1}^{r_1+1} \underset{\sim}{q}_i\underset{\sim}{q}_i'$$

The $\{\underset{\sim}{q}_i\}$ are r_1+1 mutually perpendicular unit length vectors that span the space of all linear combinations of the columns of Z. The projection of $\underset{\sim}{y}$ is then (see Result 2A.2 and Definition 2A.12)

$$\sum_{i=1}^{r_1+1} (\underset{\sim}{q}_i'\underset{\sim}{y})\underset{\sim}{q}_i = \sum_{i=1}^{r_1+1} \underset{\sim}{q}_i(\underset{\sim}{q}_i'\underset{\sim}{y}) = \left(\sum_{i=1}^{r_1+1} \underset{\sim}{q}_i\underset{\sim}{q}_i'\right)\underset{\sim}{y} = Z(Z'Z)^{-} Z'\underset{\sim}{y}$$

d) See Hint.

7.7 Write $\underset{\sim}{\beta} = \left[\begin{array}{c}\underset{\sim}{\beta}_{(1)} \\ \hline \underset{\sim}{\beta}_{(2)}\end{array}\right]$ and $Z = \left[\begin{array}{c|c} Z_1 & Z_2 \end{array}\right]$.

Recall from Result 7.4 that $\hat{\underset{\sim}{\beta}} = \left[\begin{array}{c}\hat{\underset{\sim}{\beta}}_{(1)} \\ \hline \hat{\underset{\sim}{\beta}}_{(2)}\end{array}\right] = (Z'Z)^{-1}Z'\underset{\sim}{y}$ is distributed as $N_{r+1}(\underset{\sim}{\beta},\sigma^2(Z'Z)^{-1})$ independently of $n\hat{\sigma}^2 = (n-r-1)s^2$ which is distributed as $\sigma^2 \chi^2_{n-r-1}$. From the Hint, $(\hat{\underset{\sim}{\beta}}_{(2)}-\underset{\sim}{\beta}_{(2)})'(C^{22})^{-1}(\hat{\underset{\sim}{\beta}}_{(2)}-\underset{\sim}{\beta}_{(2)})$ is $\sigma^2\chi^2_{r-q}$ and this is distributed independently of s^2. (The latter follows because the full random vector $\hat{\underset{\sim}{\beta}}$ is distributed independently of s^2). The result follows from the definition of a F random variable as the ratio of two independent χ^2 random variables divided by their degrees of freedom.

7.8 (a) $H^2 = Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z' = Z(Z'Z)^{-1}Z' = H$.

(b) Since $I - H$ is an idempotent matrix, it is positive semidefinite. Let a be an $n \times 1$ unit vector with j th element 1. Then $0 \leq a'(I-H)a = (1 - h_{jj})$. That is, $h_{jj} \leq 1$. On the other hand, $(Z'Z)^{-1}$ is positive definite. Hence $h_{jj} = b_j'(Z'Z)^{-1}b_j > 0$ where b_j is the j th row of Z.
$\sum_{i=1}^{r+1} h_{jj} = tr(Z(Z'Z)^{-1}Z') = tr((Z'Z)^{-1}Z'Z) = tr(I_{r+1}) = r+1$.

(c) Using

$$(Z'Z)^{-1} = \frac{1}{n\sum_{i=1}^n (z_j - \bar{z})^2}\begin{bmatrix} \sum_{i=1}^n z_i^2 & -\sum_{i=1}^n z_i \\ -\sum_{i=1}^n z_i & n \end{bmatrix},$$

we obtain

$$\begin{aligned} h_{jj} &= (1\ z_j)(Z'Z)^{-1}\begin{pmatrix} 1 \\ z_j \end{pmatrix} \\ &= \frac{1}{n\sum_{i=1}^n (z_j - \bar{z})^2}\left(\sum_{i=1}^n z_i^2 - 2z_j\sum_{i=1}^n z_i + nz_j^2\right) \\ &= \frac{1}{n} + \frac{(z_j - \bar{z})^2}{\sum_{i=1}^n (z_j - \bar{z})^2} \end{aligned}$$

7.9

$$Z' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}; \quad (Z'Z)^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}$$

$$\hat{\underset{\sim}{\beta}}_{(1)} = (Z'Z)^{-1}Z'\underset{\sim}{y}_{(1)} = \begin{bmatrix} 3 \\ -.9 \end{bmatrix}; \quad \hat{\underset{\sim}{\beta}}_{(2)} = (Z'Z)^{-1}Z'\underset{\sim}{y}_{(2)} = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$$

$$\hat{\beta} = \left[\hat{\underset{\sim}{\beta}}_{(1)} \;\vdots\; \hat{\underset{\sim}{\beta}}_{(2)}\right] = \begin{bmatrix} 3 & 0 \\ -.9 & 1.5 \end{bmatrix}$$

Hence

$$\hat{\Upsilon} = Z\hat{\beta} = \begin{bmatrix} 4.8 & -3.0 \\ 3.9 & -1.5 \\ 3.0 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3.0 \end{bmatrix};$$

$$\hat{\varepsilon} = \Upsilon - \hat{\Upsilon} = \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 4.8 & -3.0 \\ 3.9 & -1.5 \\ 3.0 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3.0 \end{bmatrix} = \begin{bmatrix} .2 & 0 \\ -.9 & .5 \\ 1.0 & -1.0 \\ -.1 & .5 \\ -.2 & 0 \end{bmatrix}$$

$$\Upsilon'\Upsilon = \hat{\Upsilon}'\hat{\Upsilon} + \hat{\varepsilon}'\hat{\varepsilon}$$

$$\begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix} = \begin{bmatrix} 53.1 & -13.5 \\ -13.5 & 22.5 \end{bmatrix} + \begin{bmatrix} 1.9 & -1.5 \\ -1.5 & 1.5 \end{bmatrix}$$

7.10 **a)** Using Result 7.7, the 95% confidence interval for the mean reponse is given by

$$[1,\ .5]\begin{bmatrix}3.0\\-.9\end{bmatrix} \pm 3.18\sqrt{[1,\ .5]\begin{bmatrix}.2 & 0\\0 & .1\end{bmatrix}\begin{bmatrix}1\\.5\end{bmatrix}\left(\frac{1.9}{3}\right)} \quad \text{or}$$

(1.35, 3.75).

b) Using Result 7.8, the 95% prediction interval for the actual Y is given by

$$[1,\ -.5]\begin{bmatrix}3.0\\-.9\end{bmatrix} \pm 3.18\sqrt{\left\{1 + [1,\ .5]\begin{bmatrix}.2 & 0\\0 & .1\end{bmatrix}\begin{bmatrix}1\\.5\end{bmatrix}\right\}\left(\frac{1.9}{3}\right)} \quad \text{or}$$

(-.25, 5.35) .

c) Using (7-42) a 95% prediction ellipse for the actual Y's is given by

$$[y_{01} - 2.55,\ y_{02} - .75]\begin{bmatrix}7.5 & 7.5\\7.5 & 9.5\end{bmatrix}\begin{bmatrix}y_{01} - 2.55\\y_{02} - .75\end{bmatrix}$$

$$\le (1 + .225)\left(\frac{(2)(3)}{2}\right)(19) = 69.825$$

7.11 The proof follows the proof of Result 7.10 with Σ^{-1} replaced by **A**.

$$(\underset{\sim}{Y}-ZB)'(\underset{\sim}{Y}-Z'B) = \sum_{j=1}^{n} (\underset{\sim}{Y}_j-B\underset{\sim}{z}_j)(\underset{\sim}{Y}_j-B\underset{\sim}{z}_j)'$$

and

$$\Sigma_{j=1}^{n} d_j^2(B) = tr[A^{-1}(\underset{\sim}{Y}-ZB)'(\underset{\sim}{Y}-ZB)] .$$

Next,

$$(\underset{\sim}{Y}-ZB)'(\underset{\sim}{Y}-ZB) = (\underset{\sim}{Y}-Z\hat{\underset{\sim}{\beta}}+Z\hat{\underset{\sim}{\beta}}-ZB)'(Y-Z\hat{\underset{\sim}{\beta}}+Z\hat{\beta}-ZB) = \hat{\varepsilon}'\hat{\varepsilon} + (\hat{\underset{\sim}{\beta}}-B)'Z'Z(\hat{\underset{\sim}{\beta}}-B)]$$

so

$$\Sigma_{j=1}^{n} d_j^2(B) = tr[A^{-1}\hat{\varepsilon}'\hat{\varepsilon}] + tr[A^{-1}(\hat{\underset{\sim}{\beta}}-B)'Z'Z(\hat{\underset{\sim}{\beta}}-B)]$$

The first term does not depend on the choice of B. Using Result 2A.12(c)

$$\begin{aligned} tr[A^{-1}(\hat{\underset{\sim}{\beta}}-B)'Z'Z(\hat{\underset{\sim}{\beta}}-B) &= tr[(\hat{\underset{\sim}{\beta}}-B)'Z'Z(\hat{\underset{\sim}{\beta}}-B)A] \\ &= tr[Z'Z(\hat{\underset{\sim}{\beta}}-B)A(\hat{\underset{\sim}{\beta}}-B)'] \\ &= tr[Z(\hat{\underset{\sim}{\beta}}-B)A(\hat{\underset{\sim}{\beta}}-B)'Z'] \\ &\geq \underset{\sim}{c}'A\underset{\sim}{c} > 0 \end{aligned}$$

where $\underset{\sim}{c}$ is any non-zero row of $\underset{\sim}{Z}(\hat{\underset{\sim}{\beta}}-B)$. Unless $B = \hat{\underset{\sim}{\beta}}$, $Z(\hat{\underset{\sim}{\beta}}-B)$ will have a non-zero row. Thus $\hat{\underset{\sim}{\beta}}$ is the best choice for any positive definite A.

7.12 **(a)** best linear predictor $= -4 + 2Z_1 - Z_2$

(b) mean square error $= \sigma_{yy} - \underset{\sim}{\sigma}'_{zy} \Sigma_{zz}^{-1} \underset{\sim}{\sigma}_{zy} = 4$

(c) $$\rho_{Y(x)} = \sqrt{\frac{\underset{\sim}{\sigma}'_{zy} \Sigma_{zz}^{-1} \underset{\sim}{\sigma}_{zy}}{\sigma_{yy}}} = \frac{\sqrt{5}}{3} = .745$$

(d) Following equation (7-56), we partition Σ as

$$\Sigma = \left[\begin{array}{cc|c} 9 & 3 & 1 \\ 3 & 2 & 1 \\ \hline 1 & 1 & 1 \end{array}\right]$$

and determine covariance of $\begin{bmatrix} Y \\ Z_1 \end{bmatrix}$ given z_2 to be

$$\begin{bmatrix} 9 & 3 \\ 3 & 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} (1)^{-1} [1, 1] = \begin{bmatrix} 8 & 2 \\ 2 & 1 \end{bmatrix} .$$ Therefore

$$\rho_{YZ_1 \cdot Z_2} = \frac{2}{\sqrt{8}\sqrt{1}} = \frac{\sqrt{2}}{2} = .707$$

7.13 **(a)** By Result 7.13, $\hat{\underset{\sim}{\beta}} = S_{zz}^{-1} \underset{\sim}{s}_{zy} = \begin{bmatrix} 3.73 \\ 5.57 \end{bmatrix}$

(b) Let $\underset{\sim}{Z}'_{(2)} = [Z_2, Z_3]$ $$R_{z_1(z_2 z_3)} = \sqrt{\frac{\underset{\sim}{s}'_{z_{(2)} z_1} S_{z_{(2)} z_{(2)}}^{-1} \underset{\sim}{s}_{z_{(2)} z_1}}{s_{z_1 z_1}}}$$

$$= \sqrt{\frac{3452.33}{5691.34}} = .78$$

(c) Partition $\underset{\sim}{Z} = \left[\begin{array}{c} \underset{\sim}{Z}_{(1)} \\ \hline Z_3 \end{array}\right]$ so

$$S = \left[\begin{array}{cc|c} 5691.34 & & \\ 600.51 & 126.05 & \\ \hline 217.25 & 23.37 & 23.11 \end{array}\right] = \left[\begin{array}{c|c} S_{z_{(1)}z_{(1)}} & \underset{\sim}{s}'_{z_3 z_{(1)}} \\ \hline \underset{\sim}{s}_{z_3 z_{(1)}} & s_{z_3 z_3} \end{array}\right]$$

and

$$S_{z_{(1)}z_{(1)}} - \underset{\sim}{s}'_{z_3 z_{(1)}} \, s^{-1}_{z_3 z_3} \, \underset{\sim}{s}_{z_3 z_{(1)}} = \begin{bmatrix} 3649.04 & 380.82 \\ 380.82 & 102.42 \end{bmatrix}$$

Thus

$$r_{z_1 z_2 \cdot z_3} = \frac{380.82}{\sqrt{3649.04}\,\sqrt{102.42}} = .62$$

7.14 **(a)** The large positive correlation between a manager's experience and achieved rate of return on portfolio indicates an apparent advantage for managers with experience. The negative correlation between attitude toward risk and achieved rate of return indicates an apparent advantage for conservative managers.

(b) From (7-57)

$$r_{yz_1 \cdot z_2} = \frac{s_{yz_1 \cdot z_2}}{\sqrt{s_{yy \cdot z_2}}\,\sqrt{s_{z_1 z_1 \cdot z_2}}} = \frac{s_{yz_1} - \dfrac{s_{yz_2}\, s_{z_1 z_2}}{s_{z_2 z_2}}}{\sqrt{s_{yy} - \dfrac{s^2_{yz_2}}{s_{z_2 z_2}}}\,\sqrt{s_{z_1 z_1} - \dfrac{s^2_{z_1 z_2}}{s_{z_2 z_2}}}}$$

$$= \frac{r_{yz_1} - r_{yz_2}\, r_{z_1 z_2}}{\sqrt{1 - r^2_{yz_2}}\,\sqrt{1 - r^2_{z_1 z_2}}} = .31$$

Removing "years of experience" from consideration, we now have a positive correlation between "attitude toward risk" and "achieved return". After adjusting for years of experience, there is an apparent advantage to managers who take risks.

7.15 (a) MINITAB computer output gives: $\hat{y} = 11{,}870 + 2634z_1 + 45.2z_2$; residual sum of squares = 204995012 with 17 degrees of freedom. Thus $s = 3473$. Now for example, the estimated standard deviation of $\hat{\beta}_0$ is $\sqrt{1.9961s^2} = 4906$. Similar calculations give the estimated standard deviations of $\hat{\beta}_1$ and $\hat{\beta}_2$.

(b) An analysis of the residuals indicate there are no apparent model inadequacies.

(c) The 95% prediction interval is (\$51,228; \$66,239)

(d) Using (7-14), $F = \dfrac{(45.2)(.0067)^{-1}(45.2)}{12058533} = .025$

Since $F_{1,17}(.05) = 4.45$ we cannot reject $H_0: \beta_2 = 0$. It appears as if Z_2 is not needed in the model provided Z_1 is included in the model.

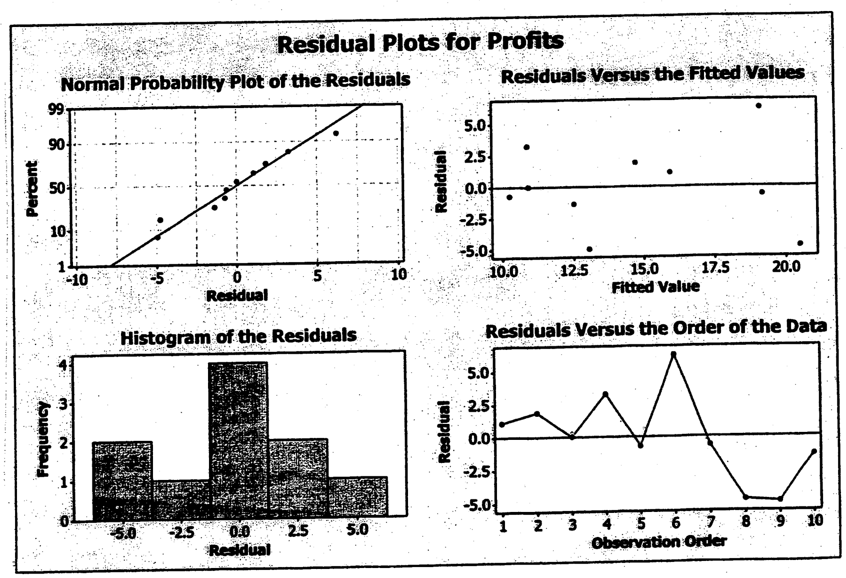
7.16

Predictors	$p = r + 1$	C_p
Z_1	2	1.025
Z_2	2	12.24
Z_1, Z_2	3	3

7.17 (a) Minitab output for the regression of profits on sales and assets follows.

```
Profits = 0.01 + 0.0681 Sales + 0.00577 Assets

Predictor       Coef    SE Coef     T      P
Constant       0.013      7.641  0.00  0.999
Sales        0.06806    0.02785  2.44  0.045
Assets      0.005768   0.004946  1.17  0.282

S = 3.86282   R-Sq = 55.7%   R-Sq(adj) = 43.0%

Analysis of Variance

Source          DF      SS     MS     F      P
Regression       2  131.26  65.63  4.40  0.058
Residual Error   7  104.45  14.92
Total            9  235.71
```

(b) Given the small sample size, the residual plots below are consistent with the usual regression assumptions. The leverages do not indicate any unusual observations. All leverages are less than $3p/n=3(3)/10=.9$.



Obs	1	2	3	4	5	6	7	8	9	10
Lev	.6257	.1011	.2433	.2222	.2513	.2746	.2785	.3642	.2029	.4362

(c) With sales = 100 and assets = 500, a 95% prediction interval for profits is: (−1.55, 20.95).

(d) The t-value for testing $H_0:\beta_2=0$ is $t=1.17$ with a p value of .282. We cannot reject H_0 at any reasonable significance level. The model should be refit after dropping assets as a predictor variable. That is, consider the simple linear regression model relating profits to sales.

7.18 (a) The calculations for the C_p plot are given below. Note that p is the number of model parameters including the intercept.

p (predictor)	2 (sales)	2 (assets)	3 (sales, assets)
C_p	2.4	7.0	3.0

(b) The AIC values are shown below.

p (predictor)	2 (sales)	2 (assets)	3 (sales, assets)
AIC	29.24	33.63	29.46

7.19 (a) The "best" regression equation involving ln(y) and $z_1, z_2, \ldots, z_5$ is

$$\widehat{\ln}(y) = 2.756 - .322z_2 + .114z_4$$

with $s = 1.058$ and $R^2 = .60$. It may be possible to find a better model using first and second order predictor variable terms.

(b) A plot of the residuals versus the predicted values indicates no apparent problems. A Q-Q plot of the residuals is a bit wavy but the sample size is not large. Perhaps a transformation other than the logarithmic transformation would produce a better model.

7.20 Eigenvalues of the correlation matrix of the predictor variables z_1, $z_2,\ldots,z_5$ are 1.4465, 1.1435, .8940, .8545, .6615. The corresponding eigenvectors give the coefficients of z_1, $z_2,\ldots,z_5$ in the principle component. For example, the first principal component, written in terms of standardized predictor variables, is

$$\hat{x}_1 = .6064z_1^* - .3901z_2^* - .6357z_3^* - .2755z_4^* - .0045z_5^* .$$

A regression of $\ell n(y)$ on the first principle component gives

$$\hat{\ell n}(y) = 1.7371 - .0701\hat{x}_1$$

with $s = .701$ and $R^2 = .015$.

A regression of $\ell n(y)$ on the fourth principle component produces the best of the one principle component predictor variable regressions. In this case $\hat{\ell n}(y) = 1.7371 + .3604\hat{x}_4$ and $s = .618$ and $R^2 = .235$.

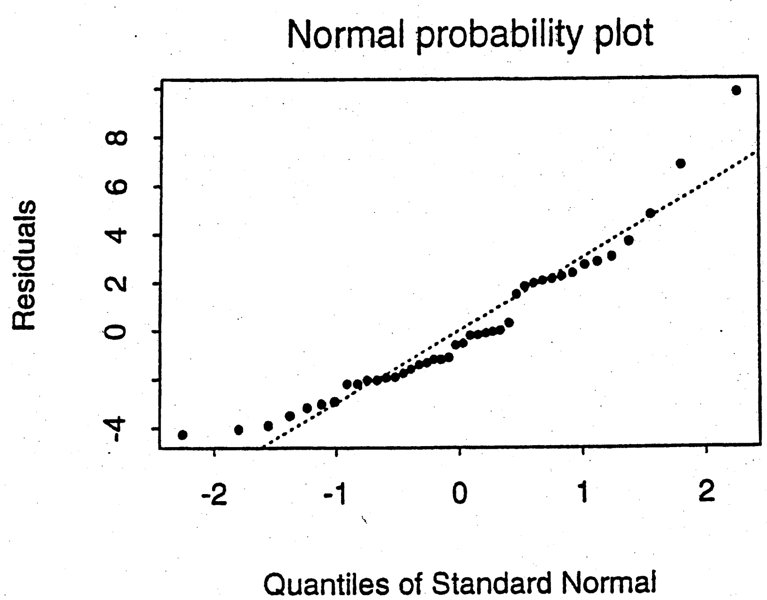
7.21 This data set doesn't appear to yield a regression relationship which explains a large proportion of the variation in the responses.

(a) (i) One reader, starting with a full quadratic model in the predictors z_1 and z_2, suggested the fitted regression equation:

$$\hat{y}_1 = -7.3808 + .5281z_2 - .0038z_2^2$$

with $s = 3.05$ and $R^2 = .22$. (Can you do better than this?)

(ii) A plot of the residuals versus the fitted values suggests the response may not have constant variance. Also a Q-Q plot of the residuals has the following general appearance:



Therefore the normality assumption may also be suspect. Perhaps a better regression can be obtained after the responses have been transformed or re-expressed in a different metric.

(iii) Using the results in (a)(i), a 95% prediction interval of $z_1 = 10$ (not needed) and $z_2 = 80$ is

$$10.84 \pm 2.02\sqrt{7.47} \quad \text{or} \quad (5.32, 16.37).$$

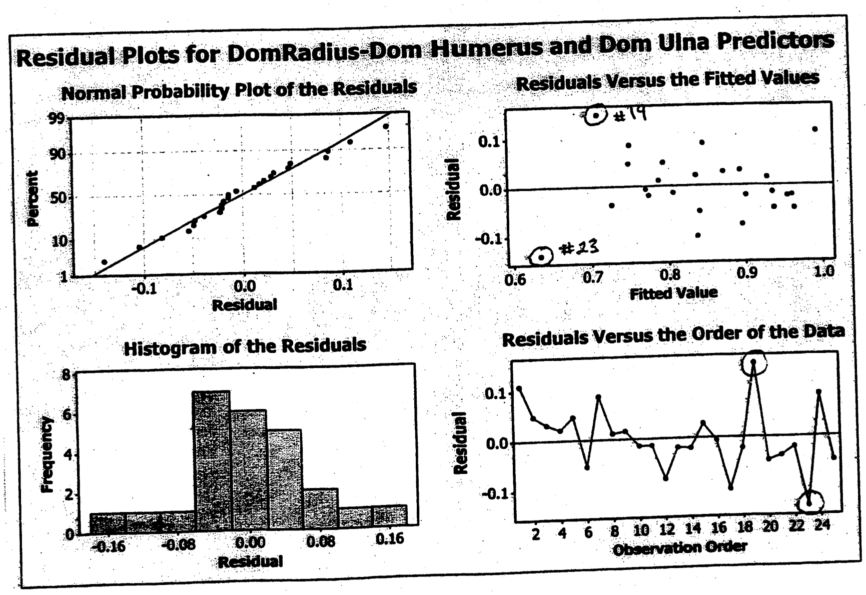
7.22 (a) The full regression model relating the dominant radius bone to the four predictor variables is shown below along with the "best" model after eliminating non-significant predictors. A residual analysis for the best model indicates there is no reason to doubt the standard regression assumptions although observations 19 and 23 have large standardized residuals.

(i)

```
The regression equation is
DomRadius = 0.103 + 0.276 DomHumerus - 0.165 Humerus + 0.357 DomUlna
            + 0.407 Ulna


Predictor        Coef   SE Coef      T      P
Constant       0.1027    0.1064   0.97  0.346
DomHumerus     0.2756    0.1147   2.40  0.026
Humerus       -0.1652    0.1381  -1.20  0.246
DomUlna        0.3566    0.1985   1.80  0.088
Ulna           0.4068    0.2174   1.87  0.076


S = 0.0663502   R-Sq = 71.8%   R-Sq(adj) = 66.1%
```

```
The regression equation is
DomRadius = 0.164 + 0.162 DomHumerus + 0.552 DomUlna


Predictor        Coef   SE Coef     T      P
Constant       0.1637    0.1035  1.58  0.128
DomHumerus    0.16249   0.05940  2.74  0.012
DomUlna        0.5519    0.1566  3.53  0.002


S = 0.0687763   R-Sq = 66.7%   R-Sq(adj) = 63.6%

Analysis of Variance

Source          DF       SS       MS      F      P
Regression       2  0.20797  0.10399  21.98  0.000
Residual Error  22  0.10406  0.00473
Total           24  0.31204
```

(ii)



(b) The full regression model relating the radius bone to the four predictor variables is shown below. This fitted model along with the fitted model for the dominant radius bone using four predictors shown in part (a) (i) and the error sum of squares and cross products matrix constitute the multivariate multiple regression model. It appears as if a multivariate regression model with only one or two predictors will represent the data well. Using Result 7.11, a multivariate regression model with predictors dominant ulna and ulna may be reasonable. The results for these predictors follow.

```
The regression equation is
Radius = 0.114 - 0.0110 DomHumerus + 0.152 Humerus + 0.198 DomUlna + 0.462 Ulna

Predictor       Coef   SE Coef      T       P
Constant     0.11423   0.08971   1.27   0.217
DomHumerus  -0.01103   0.09676  -0.11   0.910
Humerus       0.1520    0.1165   1.31   0.207
DomUlna       0.1976    0.1674   1.18   0.252
Ulna          0.4625    0.1833   2.52   0.020

S = 0.0559501   R-Sq = 77.2%   R-Sq(adj) = 72.6%
```

Error sum of squares and cross products matrix: $\begin{bmatrix} .088047 & .050120 \\ .050120 & .062608 \end{bmatrix}$

```
The regression equation is
DomRadius = 0.223 + 0.564 DomUlna + 0.321 Ulna

Predictor     Coef   SE Coef      T       P
Constant    0.2235    0.1120   2.00   0.059
DomUlna     0.5645    0.2108   2.68   0.014
Ulna        0.3209    0.2202   1.46   0.159

S = 0.0760309   R-Sq = 59.2%   R-Sq(adj) = 55.5%

Analysis of Variance

Source          DF        SS        MS      F       P
Regression       2  0.184863  0.092431  15.99   0.000
Residual Error  22  0.127175  0.005781
Total           24  0.312038
```

```
The regression equation is
Radius = 0.178 + 0.322 DomUlna + 0.595 Ulna

Predictor      Coef   SE Coef      T       P   VIF
Constant    0.17846   0.08931   2.00   0.058
DomUlna      0.3220    0.1680   1.92   0.068   2.1
Ulna         0.5953    0.1755   3.39   0.003   2.1

S = 0.0606160   R-Sq = 70.5%   R-Sq(adj) = 67.8%

Analysis of Variance

Source          DF        SS         MS      F      P
Regression       2  0.193195   0.096597  26.29  0.00(
Residual Error  22  0.080835   0.003674
Total           24  0.274029
```

Error sum of squares and cross products matrix: $\begin{bmatrix} .127175 & .064903 \\ .064903 & .080835 \end{bmatrix}$

7.23. (a) Regression analysis using the response Y_1 = SalePr.

Summary of Backward Elimination Procedure for Dependent Variable X2

Step	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	X9	7	0.0041	0.5826	7.6697	0.6697	0.4161
2	X3	6	0.0043	0.5782	6.3735	0.7073	0.4033
3	X5	5	0.0127	0.5655	6.4341	2.0795	0.1538

Dependent Variable: X2 SalePr

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	16462859.832	3292571.9663	18.224	0.0001
Error	70	12647164.839	180673.78342		
C Total	75	29110024.671			

Root MSE 425.05739 R-square 0.5655

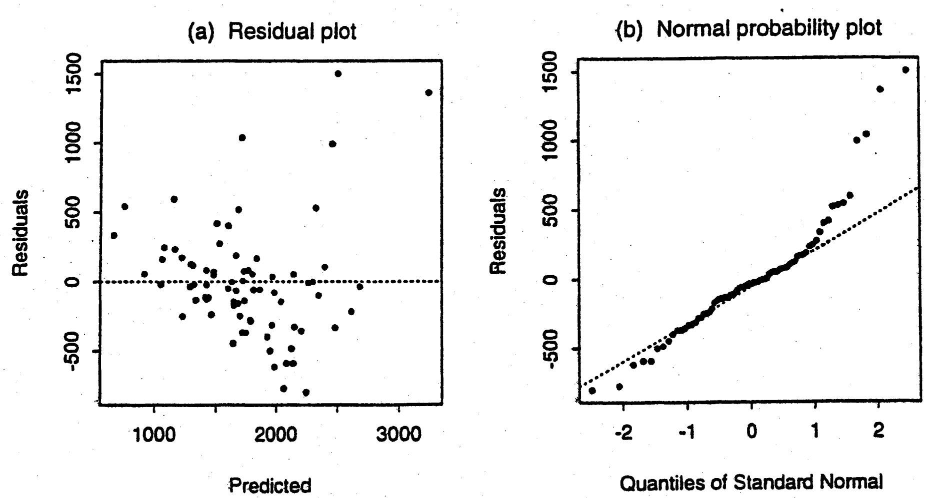
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-5605.823664	1929.3986440	-2.905	0.0049
X1	1	-77.633612	22.29880197	-3.482	0.0009
X4	1	-2.332721	0.75490590	-3.090	0.0029
X6	1	389.364490	89.17300145	4.366	0.0001
X7	1	1749.420733	701.21819165	2.495	0.0150
X8	1	133.177529	46.66673277	2.854	0.0057

The 95% prediction interval for SalePr for z_0 is

$$z_0'\hat{\beta} \pm t_{70}(0.025)\sqrt{(425.06)^2(1 + z_0'(\mathrm{Z}'\mathrm{Z})^{-1}z_0)}.$$

SalePr=f(Breed, FtFrBody, Frame, BkFat, SaleHt)



(b) Regression analysis using the response $Y_1 = \ln(\text{SalePr})$.

Summary of Backward Elimination Procedure for Dependent Variable LOGX2

Step	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	X3	7	0.0033	0.6368	7.6121	0.6121	0.4368
2	X7	6	0.0057	0.6311	6.6655	1.0594	0.3070
3	X9	5	0.0122	0.6189	6.9445	2.2902	0.1348
4	X4	4	0.0081	0.6108	6.4537	1.4890	0.2265

Dependent Variable: LOGX2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	4.02968	1.00742	27.854	0.0001
Error	71	2.56794	0.03617		
C Total	75	6.59762			

Root MSE 0.19018 R-square 0.6108

Parameter Estimates

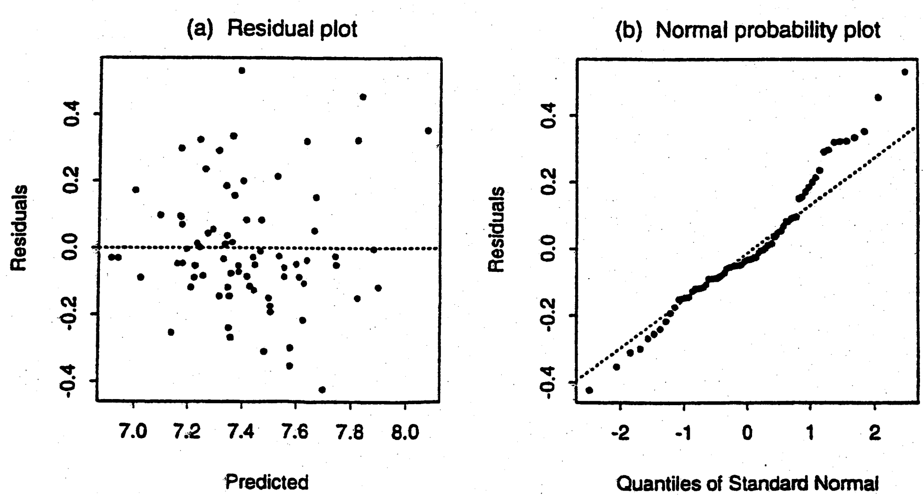
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	5.235773	0.91286786	5.736	0.0001
X1	1	-0.049418	0.00846029	-5.841	0.0001
X5	1	-0.027613	0.00827438	-3.337	0.0013
X6	1	0.183611	0.03992448	4.599	0.0001
X8	1	0.058996	0.01927655	3.060	0.0031

The 95% prediction interval for $\ln(\text{SalePr})$ for z_0 is

$$z_0'\hat{\beta} \pm t_{70}(0.025)\sqrt{(0.1902)^2(1 + z_0'(Z'Z)^{-1}z_0)}.$$

The few outliers among these latter residuals are not so pronounced.

ln(SalePr)=f(Breed, PrctFFB, Frame, SaleHt)

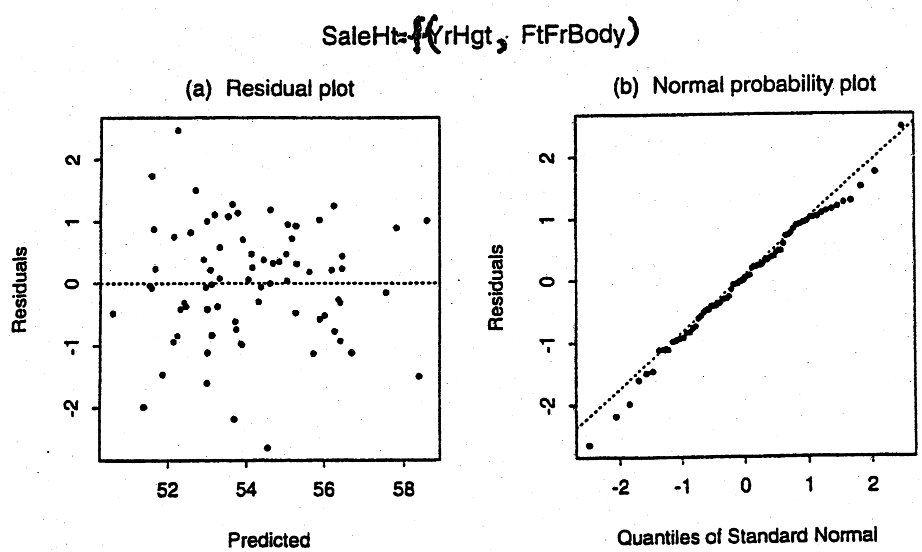


7.24. (a) Regression analysis using the response Y_1 = SaleHt and the predictors Z_1 = YrHgt and Z_2 = FtFrBody.

```
Dependent Variable: X8          SaleHt
Analysis of Variance
                         Sum of          Mean
Source          DF      Squares        Square       F Value       Prob>F
Model            2    235.74533     117.87267       131.165       0.0001
Error           73     65.60204       0.89866
C Total         75    301.34737

    Root MSE        0.94798      R-square         0.7823
Parameter Estimates
                    Parameter        Standard    T for H0:
Variable  DF         Estimate           Error  Parameter=0     Prob > |T|
INTERCEP   1         7.846281      3.36221288        2.334         0.0224
X3         1         0.802235      0.08088562        9.918         0.0001
X4         1         0.005773      0.00151072        3.821         0.0003
```

The 95% prediction interval for SaleHt for $z_0' = (1, 50.5, 970)$ is

$$53.96 \pm t_{73}(0.025)\sqrt{0.8987(1.0148)} = (52.06, 55.86).$$



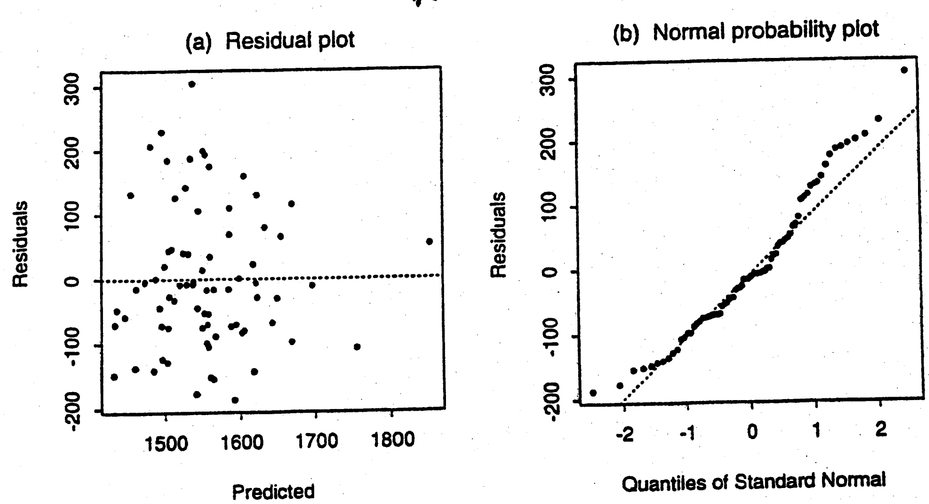
(b) Regression analysis using the response Y_2 = SaleWt and the predictors Z_1 = YrHgt and Z_2 = FtFrBody.

```
Dependent Variable: X9          SaleWt
Analysis of Variance
                          Sum of          Mean
Source          DF       Squares        Square      F Value      Prob>F
Model            2  390456.63614  195228.31807       16.319      0.0001
Error           73  873342.99544   11963.60268
C Total         75  1263799.6316

    Root MSE       109.37826      R-square         0.3090
Parameter Estimates
                    Parameter         Standard     T for H0:
Variable   DF        Estimate            Error   Parameter=0     Prob > |T|
INTERCEP    1      675.316794     387.93499836         1.741         0.0859
X3          1        2.719286       9.33265244         0.291         0.7716
X4          1        0.745610       0.17430765         4.278         0.0001
```

The 95% prediction interval for SaleWt for $z_0' = (1, 50.5, 970)$ is

$$1535.9 \pm t_{73}(0.025)\sqrt{11963.6(1.0148)} = (1316.3, 1755.5).$$

SaleWt=f(YrHgt, FtFrBody)



Multivariate regression analysis using the responses Y_1 = SaleHt and Y_2 = SaleWt and the predictors Z_1 = YrHgt and Z_2 = FtFrBody.

```
Multivariate Test:  HO: YrHgt = 0
Multivariate Statistics and Exact F Statistics
S=1     M=0     N=35
```

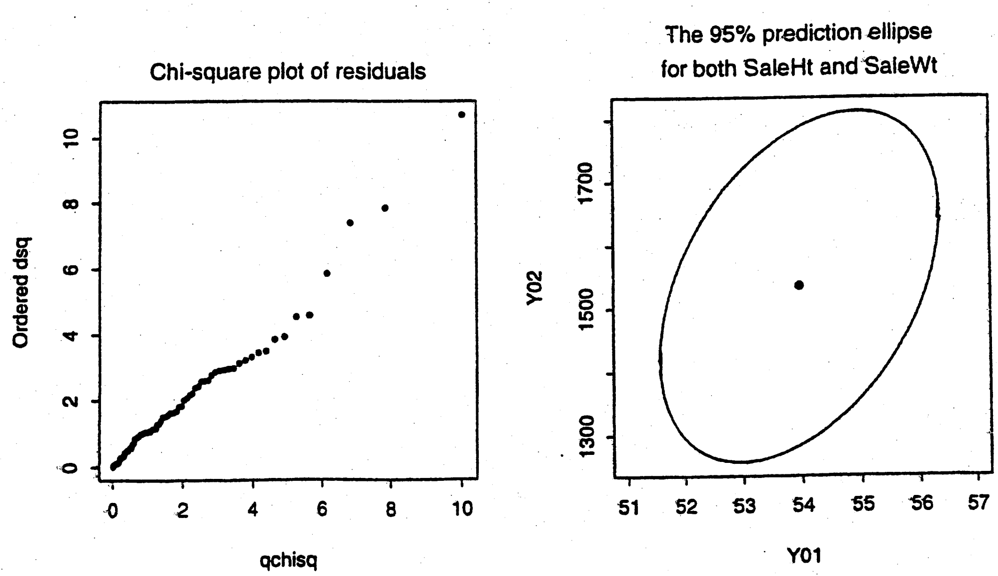
Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.38524567	57.4469	2	72	0.0001
Pillai's Trace	0.61475433	57.4469	2	72	0.0001
Hotelling-Lawley Trace	1.59574625	57.4469	2	72	0.0001
Roy's Greatest Root	1.59574625	57.4469	2	72	0.0001

```
Multivariate Test:  HO: FtFrBody = 0
Multivariate Statistics and Exact F Statistics
S=1     M=0     N=35
```

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.75813396	11.4850	2	72	0.0001
Pillai's Trace	0.24186604	11.4850	2	72	0.0001
Hotelling-Lawley Trace	0.31902811	11.4850	2	72	0.0001
Roy's Greatest Root	0.31902811	11.4850	2	72	0.0001

The theory requires using x_3 (YrHgt) to predict both SaleHt and SaleWt, even though this term could be dropped in the prediction equation for SaleWt. The 95% prediction ellipse for both SaleHt and SaleWt for $z_0' = (1, 50.5, 970)$ is

$$1.3498(Y_{01} - 53.96)^2 + 0.0001(Y_{02} - 1535.9)^2 - 0.0098(Y_{01} - 53.96)(Y_{02} - 1535.9)$$
$$= 1.0148\frac{2(73)}{72}F_{2,72}(0.05) = 6.4282.$$



7.25. (a) Regression analysis using the first response Y_1. The backward elimination procedure gives $Y_1 = \beta_{01} + \beta_{11}Z_1 + \beta_{21}Z_2$. All variables left in the model are significant at the 0.05 level. (It is possible to drop the intercept but we retain it.)

Dependent Variable: Y1 TOT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	5905583.8728	2952791.9364	22.962	0.0001
Error	14	1800356.3625	128596.88303		
C Total	16	7705940.2353			

Root MSE 358.60408 R-square 0.7664

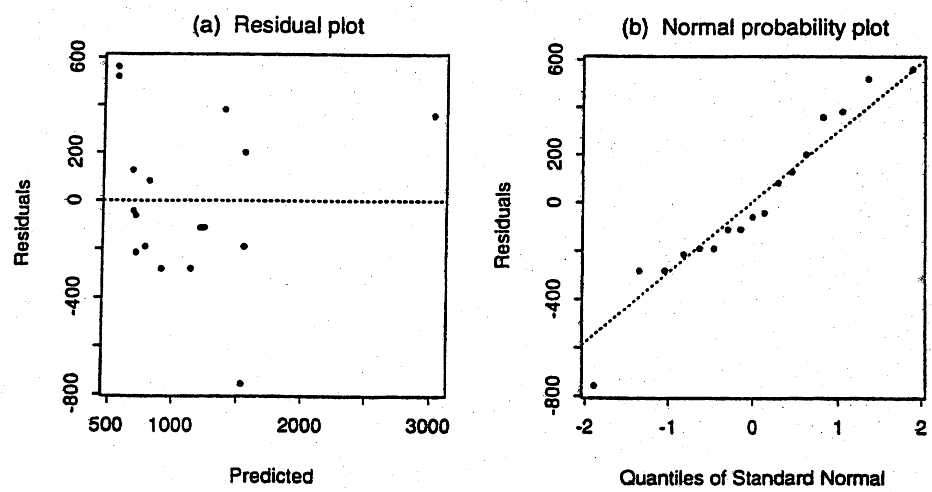
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	56.720053	206.70336862	0.274	0.7878
Z1	1	507.073084	193.79082471	2.617	0.0203
Z2	1	0.328962	0.04977501	6.609	0.0001

The 95% prediction interval for Y_1 = TOT for $z_0' = (1, 1, 1200)$ is

$$958.5 \pm t_{14}(0.025)\sqrt{128596.9(1.0941)} = (154.0, 1763.1).$$

TOT=f(GEN, AMT)



(b) Regression analysis using the second response Y_2. The backward elimination procedure gives $Y_2 = \beta_{02} + \beta_{12}Z_1 + \beta_{22}Z_2$. All variables left in the model are significant at the 0.05 level.

Dependent Variable: Y2 AMI

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	5989720.5384	2994860.2692	25.871	0.0001
Error	14	1620657.344	115761.23886		
C Total	16	7610377.8824			

Root MSE 340.23703 R-square 0.7870

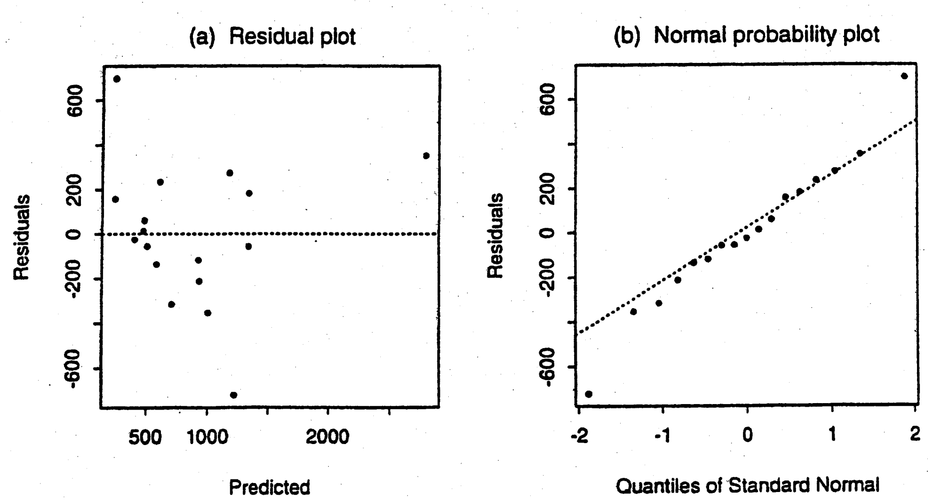
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-241.347910	196.11640164	-1.231	0.2387
Z1	1	606.309666	183.86521452	3.298	0.0053
Z2	1	0.324255	0.04722563	6.866	0.0001

The 95% prediction interval for $Y_2 = \text{AMI}$ for $z_0' = (1, 1, 1200)$ is

$$754.1 \pm t_{14}(0.025)\sqrt{115761.2(1.0941)} \;=\; (-9.234, 1517.4).$$

AMI=f(GEN, AMT)



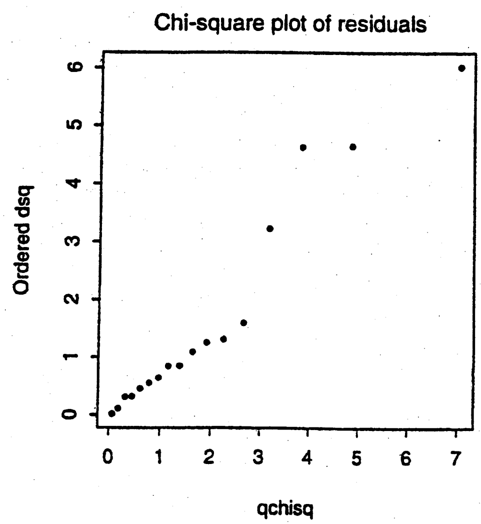
(c) Multivariate regression analysis using Y_1 and Y_2.

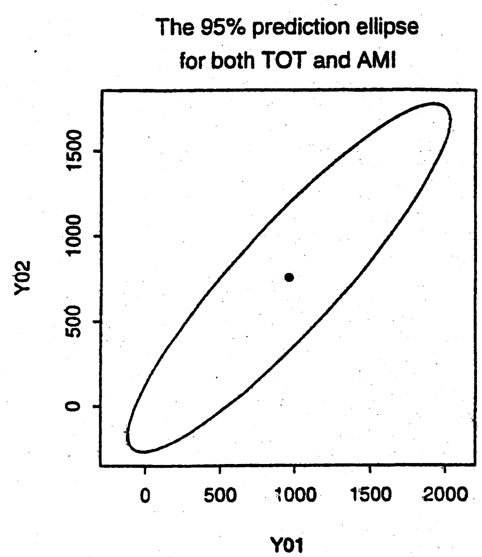
```
Multivariate Test:  HO: PR=0, DIAP=0, QRS=0
Multivariate Statistics and F Approximations
S=2     M=0     N=4
```

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.44050214	1.6890	6	20	0.1755
Pillai's Trace	0.60385990	1.5859	6	22	0.1983
Hotelling-Lawley Trace	1.16942861	1.7541	6	18	0.1657
Roy's Greatest Root	1.07581808	3.9447	3	11	0.0391

Based on Wilks' Lambda, the three variables Z_3, Z_4 and Z_5 are not significant. The 95% prediction ellipse for both TOT and AMI for $z_0' = (1, 1, 1200)$ is

$$\begin{aligned} &4.305 \times 10^{-5}(Y_{01} - 958.5)^2 + 4.782 \times 10^{-5}(Y_{02} - 754.1)^2 \\ &- 8.214 \times 10^{-5}(Y_{01} - 958.5)(Y_{02} - 754.1) \;=\; 1.0941\frac{2(14)}{13}F_{2,13}(0.05) \;=\; 8.968. \end{aligned}$$





7.26 (a) (i) The table below summarizes the results of the "best" individual regressions. Each predictor variable is significant at the 5% level.

Fitted model	R^2	s
$\hat{y}_1 = -70.1 + .0593z_2 + .0555z_3 + 82.53z_4$	73.6%	1.5192
$\hat{y}_2 = -21.6 - .9640z_1 + 27.04z_4$ $\hat{y}_2 = -20.92 + .0117z_3 + 26.12z_4$	76.5% 75.4%	.3530 .3616
$\hat{y}_3 = -43.8 + .0288z_2 + .0282z_3 + 44.59z_4$	80.7%	.6595
$\hat{y}_4 = -17.0 + .0224z_2 + .0120z_3 + 15.77z_4$	75.7%	.3504

(ii) Observations with large standardized residuals (outliers) include #51, #52 and #56. Observations with high leverage include #57, #58, #60 and #61. Apart from the outliers, the residuals plots look good.

(iii) 95% prediction interval for Y_3 is: (1.077, 4.239)

(b) (i) Using all four predictor variables, the estimated coefficient matrix and estimated error covariance matrix are

$$B = \begin{bmatrix} -74.232 & -24.015 & -45.763 & -17.727 \\ -3.120 & -1.185 & -1.486 & -.550 \\ .098 & .009 & .047 & .029 \\ .049 & .008 & .025 & .011 \\ 85.076 & 28.755 & 45.798 & 16.220 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 2.244 & .398 & .914 & .511 \\ .398 & .118 & .193 & .089 \\ .914 & .193 & .419 & .210 \\ .511 & .089 & .210 & .122 \end{bmatrix}$$

A multivariate regression model using only the three predictors z_2, z_3 and z_4 will adequately represent the data.

(ii) The same outliers and leverage points indicated in (a) (ii) are present. Otherwise the residual analysis suggests the usual regression assumptions are reasonable.

(iii) The simultaneous prediction interval for Y_3 will be wider than the individual interval in (a) (iii).

7.27 The table below summarizes the results of the "best" individual regressions. Each predictor variable is significant at the 5% level. (The levels of Severity are coded: Low=1, High=2; the levels of Complexity are coded: Simple=1, Complex=2; the levels of Exper are coded: Novice=1, Guru=2, Experienced=3.) There are no significant interaction terms in either model.

Fitted model	R^2	s
$\widehat{Assessment} = -1.834 + 1.270 Severity + 3.003 Complexity$	74.1%	.9853
$\widehat{Implementation} = -4.919 + 3.477 Severity + 5.827 Complexity$	71.9%	2.1364

For the multivariate regression with the two predictor variables Severity and Complexity, the estimated coefficient matrix and estimated error covariance matrix are

$$\mathrm{B} = \begin{bmatrix} -1.834 & -4.919 \\ 1.270 & 3.477 \\ 3.003 & 5.827 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} .9707 & 1.9162 \\ 1.9162 & 4.5643 \end{bmatrix}$$

A residual analysis suggests there is no reason to doubt the standard regression assumptions.