7. Regression Analysis

7.1 Simple linear regression for normal-theory Gauss-Markov models.

Model 1: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

where $\epsilon_i \sim NID(0, \sigma^2)$

for $i = 1, \ldots, n$.

Matrix formulation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$\mathbf{Y} = X\beta + \epsilon$$

The OLS estimator (b.l.u.e.) for β is

$$\mathrm{b} = \underline{(X^T X)}^{-1} X^T \mathbf{Y}$$

↑ when does this exist?

Here

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}$$

$$X^T \mathbf{Y} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

$$(X^T X)^{-1}$$

$$= \frac{1}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix}$$

$$= \frac{1}{n \sum_{i=1}^n (X_i - \bar{X})^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -n\bar{X} \\ -n\bar{X} & n \end{bmatrix}$$

Then

$$\mathrm{b} = (X^T X)^{-1} X^T \mathbf{Y} =$$

$$\frac{1}{n \sum_{i=1}^n (X_i - \bar{X})^2} \begin{bmatrix} \left(\sum_{i=1}^n X_i^2\right)\left(\sum_{i=1}^n Y_i\right) - n\bar{X} \sum_{i=1}^n X_i Y_i \\ -n\bar{X} \sum_{i=1}^n Y_i + n \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

and

$$\mathrm{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \bar{Y} - b_1 \bar{X} \\ \dfrac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{bmatrix}$$

Covariance matrix

$$
\begin{aligned}
Var(\mathrm{b}) &= Var\left((X^TX)^{-1}X^T\mathrm{Y}\right) \\
&= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1} \\
&= \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2} & \frac{-\bar{X}}{\Sigma(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\Sigma(X_i-\bar{X})^2} & \frac{1}{\Sigma(X_i-\bar{X})^2} \end{bmatrix}
\end{aligned}
$$

Estimate the covariance matrix for b as

$$S_{\mathrm{b}} = \mathsf{MSE}\ (X^TX)^{-1}$$

where

$$
\begin{aligned}
\mathsf{MSE} &= \mathsf{SSE}/(n-2) \\
&= \frac{1}{n-2}\mathrm{Y}(I-P_X)\mathrm{Y}.
\end{aligned}
$$

Analysis of Variance:

$$
\begin{aligned}
&\sum_{i=1}^{n} Y_i^2 = \mathrm{Y}^T\mathrm{Y} \\
&= \mathrm{Y}^T(I-P_X+P_X-P_1+P_1)\mathrm{Y} \\
&= \mathrm{Y}^T(I-P_X)\mathrm{Y} + \mathrm{Y}^T(P_X-P_1)\mathrm{Y} + \mathrm{Y}^TP_1\mathrm{Y}
\end{aligned}
$$

- $\mathrm{Y}^T(I-P_X)\mathrm{Y}$: SSE
- $\mathrm{Y}^T(P_X-P_1)\mathrm{Y}$: "Corrected model" sum of squares; call this $R(\beta_1|\beta_0)$
- $\mathrm{Y}^TP_1\mathrm{Y}$: Correction for the "mean"; call this $R(\beta_0)$

(i) By Cochran's Theorem, these three sums of squares are multiples of independent chi-squared random variables.

(ii) By result 4.7, $\frac{1}{\sigma^2}\ \mathrm{SSE} \sim \chi^2_{(\mathrm{n}-2)}$ if the model is correctly specified.

Notation

Reduction in residual sum of squares:

$$
\begin{aligned}
&R(\beta_{k+1},\ldots,\beta_{k+q} \mid \beta_0,\beta_1,\ldots,\beta_k) \\
&\quad = \mathrm{Y}^T(I-P_{X_1})\mathrm{Y} - \mathrm{Y}^T(I-P_X)\mathrm{Y}
\end{aligned}
$$

- $\mathrm{Y}^T(I-P_{X_1})\mathrm{Y}$: sum of squared residuals for the smaller model
- $\mathrm{Y}^T(I-P_X)\mathrm{Y}$: sum of squared residuals for the larger model

Here

$$X = [\ X_1 \mid X_2\ ]$$

- X_1: columns corresponding to $\beta_0,\beta_1,\ldots,\beta_k$
- X_2: columns corresponding to $\beta_{k+1}\cdots\beta_{k+q}$

Correction for the overall mean:

$$\begin{aligned} R(\beta_0) &= \mathbf{Y}^T P_1 \mathbf{Y} \\ &= \mathbf{Y}^T (I - I + P_1)\mathbf{Y} \\ &= \mathbf{Y}^T I \mathbf{Y} - \mathbf{Y}^T (I - P_1)\mathbf{Y} \\ &= \sum_{i=1}^{n} (Y_i - 0)^2 - \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \end{aligned}$$

↗ sum of squared residuals from fitting the model $Y_i = \alpha + \epsilon_i$.

The OLS estimator for $\alpha = E(Y_i)$ is

$$\begin{aligned} \hat{\alpha} &= (1^T 1)^{-1} 1^T \mathbf{Y} \\ &= (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right) = \bar{Y} \end{aligned}$$

An alternative formula is

$$\begin{aligned} R(\beta_0) &= \mathbf{Y}^T \mathbf{P}_1 \mathbf{Y} \\ &= \mathbf{Y}^T 1 (1^T 1)^{-1} 1^T \mathbf{Y} \\ &= \left(\sum_{i=1}^{n} Y_i \right) (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right) \\ &= (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right)^2 \\ &= n \bar{Y}^2 \end{aligned}$$

with $df = rank(P_1) = rank(1) = 1$.

Reduction in the residual sum of squares for regression on X_1:

$$\begin{aligned} R(\beta_1|\beta_0) &= \mathbf{Y}^T (P_X - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T (P_X - I + I - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T (I - P_1 - (I - P_X))\mathbf{Y} \\ &= \mathbf{Y}^T (I\text{-}P_1)\mathbf{Y} - \mathbf{Y}^T (I\text{-}P_X)\mathbf{Y} \end{aligned}$$

↗ sum of squared residuals for fitting the model $Y_i = \alpha + \epsilon_i$

↗ sum of squared residuals for fitting the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

ANOVA table:

Source of variation	d.f.	Sum of Squares
Regression on X	1	$R(\beta_1\|\beta_0) = \mathbf{Y}^T (P_X - P_1)\mathbf{Y}$
Residuals	$n-2$	$\mathbf{Y}^T (I - P_X)\mathbf{Y}$
Corrected total	$n-1$	$\mathbf{Y}^T (I - P_1)\mathbf{Y}$
Correction for the mean	1	$\mathbf{Y}^T P_1 \mathbf{Y} = n\bar{Y}^2$

F-tests

From result 4.7 we have

$$\frac{1}{\sigma^2}\ R(\beta_0) = \frac{1}{\sigma^2}\ \mathrm{Y}^T P_1 \mathrm{Y} \sim \chi^2_1(\delta^2)$$

where

$$\begin{aligned}\delta^2 &= \frac{1}{\sigma^2}\ \beta^T \mathrm{X}^T P_1 \mathrm{X}\beta \\ &= \frac{1}{\sigma^2}\ \beta^T \mathrm{X}^T P_1^T P_1 X\beta \\ &= \frac{1}{\sigma^2}(P_1 X\beta)^T(P_1 X\beta) \\ &= \frac{n}{\sigma^2}(\beta_0 + B_1\bar{X})^2\end{aligned}$$

Hypothesis test:

Reject $H_0 : \beta_0 + \beta_1\bar{X} = 0$ if

$$F = \frac{R(\beta_0)}{\mathrm{MSE}} > F_{(1,n-2),\alpha}$$

Also use Result 4.7 to show that

$$\frac{1}{\sigma^2}SSE = \frac{1}{\sigma^2}\ \mathrm{Y}^T(I - P_X)\mathrm{Y} \sim \chi^2_{(n-2)}$$

Use Result 4.8 to show that

$$SSE = \frac{1}{\sigma^2}\ \mathrm{Y}^T(I - P_X)\mathrm{Y}$$

is distributed independently of

$$R(\beta_0) = \frac{1}{\sigma^2}\ \mathrm{Y}^T P_1 \mathrm{Y}\ .$$

This follows from

$$(I - P_X)P_1 = 0\ .$$

Consequently,

$$F = \frac{R(\beta_0)}{\mathrm{MSE}} \sim F_{(1,n-2)}(\delta^2)$$

and this becomes a central F-distribution when the null hypothesis is true.

Test the null hypothesis $H_0 : \beta_1 = 0$

Use

$$\begin{aligned}F &= \frac{R(\beta_1|\beta_0)/1}{\mathrm{MSE}} \\ &= \frac{[\mathrm{Y}^T(P_X - P_1)\mathrm{Y}]/[1\sigma^2]}{[\mathrm{Y}^T(I - P_X)\mathrm{Y}]/[(n-2)\sigma^2]} \\ &\sim F_{(1,n-2)}(\delta^2)\end{aligned}$$

where

$$\begin{aligned}\delta^2 &= \frac{1}{\sigma^2}\ \beta^T X^T(P_X - P_1)X\beta \\ &= \frac{1}{\sigma^2}\ \beta^T X^T(P_X - P_1)^T\underline{(P_X - P_1)X\beta}\end{aligned}$$

The null hypothesis is $H_0 : (P_X - P_1)X\beta = 0$

Here

$$
\begin{aligned}
&(P_X - P_1)X \\
&\quad = (P_X - P_1)[1|\mathrm{X}] \\
&\quad = [(P_X - P_1)1 \mid (P_X - P_1)\mathrm{X}] \\
&\quad = [P_X 1 - P_1 1 \mid P_X \mathrm{X} - P_1 \mathrm{X}] \\
&\quad = [1 - 1 \mid \mathrm{X} - \bar{X}1] \\
&\quad = \begin{bmatrix} 0 & | & X_1 - \bar{X} \\ 0 & | & X_2 - \bar{X} \\ \vdots & | & \vdots \\ 0 & | & X_n - \bar{X} \end{bmatrix}
\end{aligned}
$$

If any $X_i \neq X_j$, then we cannot have both

$$X_j - \bar{X} = 0$$

and

$$X_i = \bar{X} = 0 \ .$$

Consequently, if any $X_i \neq X_j$ then

$$(P_X - P_1)X\beta = 0$$

if and only if

$$\beta_1 = 0 \ .$$

Hence, the null hypothesis is

$$H_0 : \beta_1 = 0.$$

Note that

$$\delta^2 = \frac{1}{\sigma^2} \ \beta_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Maximize the power of the F-test for $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$ by maximizing

$$\delta^2 = \frac{1}{\sigma^2} \ \beta_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Reparameterize the model:

$$Y_i = \alpha + \beta_1(X_i - \bar{X}) + \epsilon_i$$

with $\epsilon_i \sim NID(0, \sigma^2), i = 1, \ldots, n.$

Interpretation of parameters:

$\alpha = E(Y)$ when $\mathrm{X} = \bar{\mathrm{X}}$

β_1 is the change in $E(Y)$ when X is increased by one unit.

Matrix formulation:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 - \bar{X} \\ \vdots & \vdots \\ 1 & X_n - \bar{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$Y = W\gamma + \epsilon$$

Clearly,

$$W = X \begin{bmatrix} 1 & -\bar{X} \\ 0 & 1 \end{bmatrix} = XF$$

$$X = W \begin{bmatrix} 1 & \bar{X} \\ 0 & 1 \end{bmatrix} = WG$$

For this reparameterization, the columns of W are orthogonal and

$$W^TW = \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^{n}(X_i - \bar{X})^2 \end{bmatrix}$$

$$(W^TW)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

$$W^T\mathrm{Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n}(X_i - \bar{X})Y_i \end{bmatrix}$$

Then,

$$\hat{\gamma} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = (W^TX)^{-1}W^T\mathrm{Y}$$

$$= \begin{bmatrix} \bar{Y} \\ \frac{\Sigma(X_i - \bar{X})Y_i}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

and

$$Var(\hat{\gamma}) = \sigma^2(W^TW)^{-1}$$

$$= \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

Hence, $\bar{Y}$ and $\hat{\beta}_1 = \frac{\Sigma(X_i - \bar{X})Y_i}{\Sigma(X_i - \bar{X})^2}$ are uncorrelated (independent for the normal theory Gauss-Markov model).

Analysis of Variance

The reparamterization does not change the ANOVA table.

$$P_X = X(X^TX)^{-1}X^T$$

$$= W(W^TW)^{-1}W^T = P_W$$

and

$$R(\beta_0) + R(\beta_1|\beta_0) + \text{SSE}$$

$$= \mathrm{Y}^TP_1\mathrm{Y} + \mathrm{Y}^T(P_X - P_1)\mathrm{Y} + \mathrm{Y}^T(I - P_X)\mathrm{Y}$$

$$= \mathrm{Y}^TP_1\mathrm{Y} + \mathrm{Y}^T(P_W - P_1)\mathrm{Y} + \mathrm{Y}^T(I - P_W)\mathrm{Y}$$

$$= R(\alpha) + R(\beta_1|\alpha) + SSE$$

7.2 Multiple regression analysis for the normal-theory Gauss-Markov model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_r X_{ri} + \epsilon_i$$

where

$$\epsilon_i \sim NID(0, \sigma^2) \text{ for i} = 1, \ldots, \text{n}\,.$$

Matrix formulation:

$$\mathrm{Y} = X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I)$$

where

$$X\beta = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{r1} \\ 1 & X_{12} & X_{22} & \cdots & X_{r2} \\ 1 & \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{1n} & X_{2n} & \cdots & X_{rn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}$$

$$\begin{matrix} \uparrow & \uparrow & \uparrow & & \uparrow \\ 1 & \mathrm{X}_1 & \mathrm{X}_2 & & \mathrm{X}_r \end{matrix}$$

Suppose rank$(X) = r + 1$, then

(i) the OLS estimator (b.l.u.e.) for β is

$$\mathrm{b} = (X^T X)^{-1} X^T \mathrm{Y}$$

(ii) $Var(\mathrm{b}) = \sigma^2 (X^T X)^{-1}$

(iii) $$\begin{aligned} \hat{\mathrm{Y}} &= X\mathrm{b} \\ &= X(X^T X)^{-1} X^T \mathrm{Y} \\ &= P_X \mathrm{Y} \end{aligned}$$

(iv) $\mathrm{e} = \mathrm{Y} - \hat{\mathrm{Y}} = (I - P_X)\mathrm{Y}$

(v) By result 4.7,

$$\begin{aligned} \frac{1}{\sigma^2}\,\mathrm{SSE} &= \frac{1}{\sigma^2}\,\mathrm{e}^T\mathrm{e} \\ &= \frac{1}{\sigma^2}\,\mathrm{Y}^T(I - P_X)\mathrm{Y} \\ &\sim \chi^2_{(n-r-1)} \end{aligned}$$

(vi) MSE $= \frac{\mathrm{SSE}}{n-r-1}$ is an unbiased estimator of σ^2.

ANOVA

Source of variation	d.f.	Sum of squares
Model (regression on $X_1, \ldots, X_r$)	r	$R(\beta_1, \ldots, \beta_r \mid \beta_0) = \mathrm{Y}^T(P_X - P_1)\mathrm{Y}$
Error (or residuals)	$n-r-1$	$\mathrm{Y}^T(I - P_X)\mathrm{Y}$
Corrected total	$n-1$	$\mathrm{Y}^T(I - P_1)\mathrm{Y}$
Correction for the mean	1	$R(\beta_0) = \mathrm{Y}^T P_1 \mathrm{Y} = n\bar{Y}^2$

Reduction in the residual sum of squares obtained by regression on

$$X_1, X_2, \ldots, X_r$$

is denoted by

$$\begin{aligned} R(\beta_1, \beta_2, \ldots, \beta_r \mid \beta_0) &= \mathrm{Y}^T(I - P_1)\mathrm{Y} - \mathrm{Y}^T(I - P_X)\mathrm{Y} \\ &= \mathrm{Y}^T(\mathrm{P}_X - P_1)\mathrm{Y} \end{aligned}$$

Use Cochran's theorem or results 4.7 and 4.8 to show that SSE is distributed independently of

$$R(\beta_1, \beta_2, \ldots, \beta_r | \beta_0) = SS_{\text{model}}$$

and

$$\frac{1}{\sigma^2} SSE \sim \chi^2_{(n-r-1)}$$

and that

$$\frac{1}{\sigma^2} R(\beta_1, \ldots, \beta_r | \beta_0) \sim \chi^2_{(r)}(\delta^2)$$

Then

$$F = \frac{R(\beta_1, \ldots, \beta_r | \beta_0)/r}{\mathrm{MSE}} \sim F_{(r, n-r-1)}(\delta^2)$$

where

$$\begin{aligned} \delta^2 &= \frac{1}{\sigma^2} \beta^T \mathrm{X}^T (P_X - P_1) X\beta \\ &= \frac{1}{\sigma^2} \beta^T X^T (P_X - I + I - P_1) X\beta \\ &= \frac{1}{\sigma^2} \left[\beta^T X^T (I - P_1) X\beta \right. \\ &\qquad \left. - \beta^T \mathrm{X}^T \underline{(I - P_X)} X\beta \right] \end{aligned}$$

↖ This is a matrix of zeros

$$\begin{aligned} &= \tfrac{1}{\sigma^2} \beta^T X^T (I - P_1) X\beta \\ &= \tfrac{1}{\sigma^2} \beta^T X^T (I - P_1)(I - P_1) X\beta \\ &= \tfrac{1}{\sigma^2} \left[(I - P_1) X\beta\right]^T (I - P_1) X\beta \end{aligned}$$

Note that $(I - P_1)X$ is

$$[(I-P_1)1|(I-P_1)\mathrm{X}_1|\cdots|(I-P_1)\mathrm{X}_r]$$

$$= [\ 0\ |\ \mathrm{X}_1 - \bar{X}_1 1\ |\cdots|\ \mathrm{X}_r - \bar{X}_r 1]$$

$$\Longrightarrow$$

$$(I-P_1)X\beta = \sum_{j=1}^{r} \beta_j(\mathrm{X}_j - \bar{X}_j 1)$$

Then, the noncentrality parameter is

$$\frac{1}{\sigma^2}\left[\sum_{j=1}^{r}\beta_j^2(\mathrm{X}_j - \bar{X}_j 1)^T(\mathrm{X}_j - \bar{X}_j 1) + \sum\sum_{j\neq k}\beta_j\beta_k(\mathrm{X}_j - \bar{X}_j 1)^T(\mathrm{X}_k - \bar{X}_k 1)\right]$$

$$= \frac{1}{2\sigma^2}\ \beta_*^T\left[\sum_{i=1}^{n}(\mathrm{X}_{*j} - \bar{\mathrm{X}}_*)(\mathrm{X}_{*i} - \bar{\mathrm{X}}_*)^T\right]\beta_*$$

where

$$\beta_* = \begin{bmatrix}\beta_1\\ \vdots \\ \beta_r\end{bmatrix} \quad \bar{\mathrm{X}}_* = \begin{bmatrix}\bar{X}_1\\ \vdots \\ \bar{X}_r\end{bmatrix} \quad \mathrm{X}_{*j} = \begin{bmatrix}X_{1j}\\ \vdots \\ X_{rj}\end{bmatrix}$$

If $\sum_{j=1}^{n}(\mathrm{X}_{*j} - \bar{\mathrm{X}}_*)(\mathrm{X}_{*j} - \bar{\mathrm{X}}_*)^T$ is positive definite, then the null hypothesis corresponding to $\delta^2 = 0$ is

$$H_0 : \beta_* = 0(\ \text{or}\ \beta_1 = \beta_2 = \cdots = \beta_r = 0)$$

Reject $H_0 : \beta_* = 0$ if

$$F = \frac{\mathrm{Y}^T(P_X - P_1)\mathrm{Y}/r}{\mathrm{Y}^T(I - P_X)\mathrm{Y}/(n-r-1)} > F_{(r,n-r-1),\alpha}$$

Sequential sums of squares (Type I sums of squares in PROC GLM or PROC REG in SAS).

Define

$$\begin{array}{ll} X_0 = 1 & P_0 = X_0(X_0^TX_0)^{-1}X_0^T \\ X_1 = [1|\mathrm{X}_1] & P_1 = X_1(X_1^TX_1)^{-}X_1^T \\ X_2 = [1|\mathrm{X}_1|\mathrm{X}_2] & P_2 = X_2(X_2^TX_2)^{-}X_2^T \\ \vdots & \vdots \\ X_r = [1|\mathrm{X}_1|\cdots|\mathrm{X}_r] & P_r = X_r(X_r^TX_r)^{-}X_r^T \end{array}$$

Then

$$\begin{aligned}\mathrm{Y}^T\mathrm{Y} &= \mathrm{Y}^TP_0\mathrm{Y} + \mathrm{Y}^T(P_1 - P_0)\mathrm{Y} \\ &\quad + \mathrm{Y}^T(P_2 - P_1)\mathrm{Y} + \cdots \\ &\quad + \mathrm{Y}^T(P_r - P_{r-1})\mathrm{Y} \\ &\quad + \mathrm{Y}^T(I - P_r)\mathrm{Y} \\ &= R(\beta_0) + R(\beta_1|\beta_0) + R(\beta_2|\beta_0,\beta_1) \\ &\quad + \cdots + R(\beta_r|\beta_0,\beta_1,\ldots,\beta_{r-1}) \\ &\quad + \mathrm{SSE}\end{aligned}$$

- Use Cochran's theorem to show
  - these sums of squares are distributed independently of each other.
  - Each $\frac{1}{\sigma^2}\ R(\beta_i|\beta_0,\ldots,\beta_{i-1})$ has a chi-squared distribution with one degree of freedom.
- Use Result 4.7 to show
  $\frac{1}{\sigma^2}$ SSE $\sim \chi^2_{(n-r-1)}$.

Then

$$F = \frac{R(\beta_j|\beta_0,\ldots,\beta_{j-1})/1}{\text{MSE}} \sim F_{1,n-r-1}(\delta^2)$$

where

$$\begin{aligned}\delta^2 &= \frac{1}{\sigma^2}\,\beta^T X^T (P_j - P_{j-1}) X\beta \\ &= \frac{1}{\sigma^2}\,\beta^T X^T (P_j - P_{j-1})^T (P_j - P_{j-1}) X\beta \\ &= \frac{1}{\sigma^2}\,[(P_j - P_{j-1})X\beta]^T (P_j - P_{j-1}) X\beta\end{aligned}$$

Hence, this is a test of

$$H_0 : (P_j - P_{j-1})X\beta = 0$$

vs

$$H_a : (P_j - P_{j-1})X\beta \neq 0$$

Note that

$$\begin{aligned}&(P_j - P_{j-1})X \\ &\quad = (P_j - P_{j-1})[1 \mid \mathrm{X}_1 \mid \cdots \\ &\qquad \mid \mathrm{X}_{j-1} \mid \mathrm{X}_j \mid \cdots \mid \mathrm{X}_r] \\ &\quad = [(P_j - P_{j-1})1 \mid (P_j - P_{j-1})\mathrm{X}_1 \mid \cdots \\ &\qquad \mid (P_j - P_{j-1})X_{j-1} \mid (P_j - P_{j-1})\mathrm{X}_j \mid \\ &\qquad\qquad \cdots] \\ &\quad = [O_{n\times j} \mid (P_j - P_{j-1})\mathrm{X}_j \mid \cdots \\ &\qquad\qquad \mid (P_j - P_{j-1})\mathrm{X}_r\ ]\end{aligned}$$

Then

$$\begin{aligned}&(P_j - P_{j-1})X\beta \\ &\quad = \textstyle\sum_{k=j}^{r} \beta_k (P_j - P_{j-1})\mathrm{X}_k \\ &\quad = \beta_j (P_j - P_{j-1})\mathrm{X}_j \\ &\qquad + \textstyle\sum_{k=j+1}^{r} \beta_k (P_j - P_{j-1})\mathrm{X}_k\end{aligned}$$

and the null hypothesis is

$$\begin{aligned}H_0 : 0 \;=\; &\beta_j (P_j - P_{j-1})\mathrm{X}_j \\ &+ \textstyle\sum_{k=j+1}^{r} \beta_k (P_j - P_{j-1})\mathrm{X}_k\end{aligned}$$

Type II sums of squares in SAS (these are also Type III and Type IV sums of squares for regression problems).

$R(\beta_j|\beta_0$ and all other $\beta_k's)$

$$= \mathrm{Y}^T(P_X - P_{-j})\mathrm{Y}$$

where

$$P_{-j} = X_{-j}(X_{-j}^T X_{-j})^- X_{-j}^T$$

and X_{-j} is obtained by deleting the $(j+1)$-th column of X.

From the previous discussion:

$$F = \frac{\mathrm{Y}^T(P_X - P_{-j})\mathrm{Y}/1}{\mathrm{MSE}} \sim F_{(1,n-r-1)}(\delta^2)$$

where

$$\delta^2 = \frac{1}{\sigma^2}\beta^T X^T(P_X - P_{-j})X\beta$$

$$= \frac{1}{\sigma^2}\beta_j^2 \mathrm{X}_j^T(P_X - P_{-j})\mathrm{X}_j$$

This F-test provides a test of

$H_0 : \beta_j = 0$ vs $H_A : \beta_j \neq 0$

if $(P_X - P_{-j})\mathrm{X}_j \neq 0$.

Variable	Type I Sums of squares	Type II Sums of squares
X_1	$R(\beta_1\|\beta_0)$ $=\mathrm{Y}^T(P_1 - P_0)\mathrm{Y}$	$R(\beta_1\|$ other β's) $=\mathrm{Y}^T(P_X - P_{-1})\mathrm{Y}$
X_2	$R(\beta_2\|\beta_0,\beta_1)$ $= \mathrm{Y}^T(P_2 - P_1)\mathrm{Y}$	$R(\beta_2\|$ other β's) $= \mathrm{Y}^T(P_X - P_{-2})\mathrm{Y}$
$\vdots$	$\vdots$	$\vdots$
X_r	$R(\beta_r\|\beta_0,\beta_1,\ldots,\beta_{r-1})$ $= \mathrm{Y}^T(P_r - P_{r-1})\mathrm{Y}$	$R(\beta_r\|\beta_0,\ldots,\beta_{r-1})$ $= \mathrm{Y}^T(P_X - P_{-r})\mathrm{Y}$
Residuals	SSE $= \mathrm{Y}^T(I - P_X)\mathrm{Y}$	
Corrected Total	$\mathrm{Y}^T(I - P_1)\mathrm{Y}$	

When $X_1, X_2, \ldots, X_r$ are all uncorrelated, then

(i) $R(\beta_j \mid \beta_0$ and any other β's)
$= R(\beta_j|\beta_0)$
There is only one ANOVA table.

(ii) $R(\beta_j|\beta_0) = \hat{\beta}_j^2 \sum_{i=1}^{n}(X_{ji} - \bar{X}_{j.})^2$

(iii) $F = \frac{R(\beta_j|\beta_0)}{\mathrm{MSE}} \sim F_{1,n-k-1}(\delta^2)$
where $\delta^2 = \frac{1}{\sigma^2}\beta_j^2 \sum_{i=1}^{n}(X_{ji} - \bar{X}_{j.})^2$
provides a test of
$H_0 : \beta_j = 0$ versus $H_A : \beta_j \neq 0$.

Testable Hypothesis

For any testable hypothesis, reject $H_0 : C\beta = \mathrm{d}$ in favor of the general alternative $H_A : C\beta \neq \mathrm{d}$ if

$$F = \frac{(C\mathrm{b} - \mathrm{d})^T [C(X^T X)^- C^T]^{-1} (C\mathrm{b} - \mathrm{d})/m}{\mathrm{Y}^T (I - P_X)\mathrm{Y}/(n - \mathrm{rank}(X))} > F_{(m, n-\mathrm{rank}(X)), \alpha}$$

where

m = number of rows in C
= rank(C)

and

$$\mathrm{b} = (X^T X)^- X^T \mathrm{Y}$$

Confidence interval for an estimable function $\mathrm{c}^T\beta$

$$\mathrm{c}^T\mathrm{b} \pm t_{(n-rank(X))\alpha/2} \sqrt{MSE\ \mathrm{c}^T (X^T X)^- \mathrm{c}}$$

- Use $\mathrm{c}^T = (0\ 0\ ..\ 0\ \underset{\uparrow}{1}\ 0\ ..\ 0)$
  j-th position
  to construct a confidence interval for β_{j-1}
- Use $\mathrm{c}^T = (1, x_1, x_2, \ldots, x_r)$ to construct a confidence interval for
  $E(\mathrm{Y} | X_1 = x_1, \ldots, X_r = x_r)$
  $= \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r$

Prediction Intervals

Predict a future observation at

$$X_1 = x_1,\ \ldots\ ,\ X_r = x_r$$

i.e., predict

$$Y = \underbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r}_{\text{estimate the conditional mean as } b_0 + b_1 x_1 + \cdots + b_r x_r} + \underbrace{\epsilon}_{\text{estimate this with its mean } E(\epsilon) = 0}$$

A $(1-\alpha)\times$ 100% prediction interval is

$$(\mathrm{c}^T\mathrm{b} + 0) \pm t_{(n-rank(X)),\alpha/2} \sqrt{\mathrm{MSE}\ [1 + \mathrm{c}^{\mathrm{T}} (\mathrm{X}^{\mathrm{T}}\mathrm{X})^- \mathrm{c}]}$$

where

$$\mathrm{c}^T = (1\ x_1\ \cdots\ x_r)$$

```
/*  A SAS program to perform a regression
    analysis of the effects of the
    composition of Portland cement on the
    amount of heat given off as the cement
    hardens.  Posted as  cement.sas  */

data set1;
  input run x1 x2 x3 x4 y;
/* label y = evolved heat (calories)
        x1 = tricalcium aluminate
        x2 = tricalcium silicate
        x3 = tetracalcium aluminate ferrate
        x4 = dicalcium silicate;  */
  cards;
  1  7 26  6 60  78.5
  2  1 29 15 52  74.3
  3 11 56  8 20 104.3
  4 11 31  8 47  87.6
  5  7 52  6 33  95.9
  6 11 55  9 22 109.2
  7  3 71 17  6 102.7
  8  1 31 22 44  72.5
```



```
  9  2 54 18 22  93.1
 10 21 47  4 26 115.9
 11  1 40 23 34  83.8
 12 11 66  9 12 113.2
 13 10 68  8 12 109.4
run;

proc print data=set1 uniform split='*';
  var y x1 x2 x3 x4;
  label y = 'Evolved*heat*(calories)'
     x1 = 'Percent*tricalcium*aluminate'
     x2 = 'Percent*tricalcium*silicate'
     x3 = 'Percent*tetracalcium*aluminate*ferrate'
     x4 = 'Percent*dicalcium*silicate';
  run;
```



```
/*  Regress y on all four explanatory
    variables and check residual plots
    and collinearity diagnostics */

proc reg data=set1 corr;
  model y = x1 x2 x3 x4 / p r ss1 ss2
                          covb  collin;
  output out=set2 residual=r
                     predicted=yhat;
  run;


/*  Examine smaller regression models
    corresponding to subsets of the
    explanatory variables  */


proc reg data=set1;
  model  y = x1 x2 x3 x4 /
              selection=rsquare cp aic
              sbc mse stop=4 best=6;
  run;
```



```
/*  Regress y on two of explanatory
    variables and check residual plots
    and collinearity diagnostics */

proc reg data=set1 corr;
  model y = x1 x2 / p r ss1 ss2
                        covb  collin;
  output out=set2 residual=r
                      predicted=yhat;
  run;


/*  Use the GLM procedure to identify
    all estimable functions  */

proc glm data=set1;
  model y = x1 x2 x3 x4 / ss1 ss2 e1 e2 e p;
  run;
```

Obs	Evolved heat (calories)	Percent tricalcium aluminate	Percent tricalcium silicate	Percent tetracalcium aluminate ferrate	Percent dicalcium silicate
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.2	11	66	9	12
13	109.4	10	68	8	12

Correlation

Variable	x1	x2	x3	x4	y
x1	1.0000	0.2286	-0.8241	-0.2454	0.7309
x2	0.2286	1.0000	-0.1392	-0.9730	0.8162
x3	-0.8241	-0.1392	1.0000	0.0295	-0.5348
x4	-0.2454	-0.9730	0.0295	1.0000	-0.8212
y	0.7309	0.8162	-0.5348	-0.8212	1.0000

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Model	4	2664.52051	666.13013
Error	8	47.67641	5.95955
Corrected Total	12	2712.19692	

Root MSE	2.44122	R-Square	0.9824
Dependent Mean	95.41538	Adj R-Sq	0.9736
Coeff Var	2.55852		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	63.16602	69.93378	0.90	0.3928
x1	1	1.54305	0.74331	2.08	0.0716
x2	1	0.50200	0.72237	0.69	0.5068
x3	1	0.09419	0.75323	0.13	0.9036
x4	1	-0.15152	0.70766	-0.21	0.8358

Variable	DF	Type I SS	Type II SS
Intercept	1	118353	4.86191
x1	1	1448,75413	25.68225
x2	1	1205.70283	2.87801
x3	1	9.79033	0.09319
x4	1	0.27323	0.27323

Collinearity Diagnostics

Number	Eigenvalue	Condition Index
1	4.11970	1.00000
2	0.55389	2.72721
3	0.28870	3.77753
4	0.03764	10.46207
5	0.00006614	249.57825

Collinearity Diagnostics

	Proportion of Variation				
	Intercept	x1	x2	x3	x4
1	0.000005	0.00037	0.00002	0.00021	0.00036
2	8.812E-8	0.01004	0.00001	0.00266	0.00010
3	3.060E-7	0.000581	0.00032	0.00159	0.00168
4	0.000127	0.05745	0.00278	0.04569	0.00088
5	0.99987	0.93157	0.99687	0.94985	0.99730

Obs	Dep Var y	Predicted Value	Std Error Mean Predict	Student Residual
1	78.5000	78.4929	1.8109	0.00432
2	74.3000	72.8005	1.4092	0.752
3	104.3000	105.9744	1.8543	-1.054
4	87.6000	89.3333	1.3265	-0.846
5	95.9000	95.6360	1.4598	0.135
6	109.2000	105.2635	0.8602	1.723
7	102.7000	104.1289	1.4791	-0.736
8	72.5000	75.6760	1.5604	-1.692
9	93.1000	91.7218	1.3244	0.672
10	115.9000	115.6010	2.0431	0.224
11	83.8000	81.8034	1.5924	1.079
12	113.2000	112.3007	1.2519	0.429
13	109.4000	111.6675	1.3454	-1.113

Obs	-2-1 0 1 2	Cook's D
1	| | |	0.000
2	| *| |	0.057
3	| **| |	0.303
4	| *| |	0.060
5	| | |	0.002
6	| |*** |	0.084
7	| *| |	0.063
8	| ***| |	0.395
9	| |* |	0.038
10	| | |	0.023
11	| |** |	0.172
12	| | |	0.013
13	| **| |	0.108

The REG Procedure
Model: MODEL1
R-Square Selection Method
Regression Models for Dependent Variable: y

Number in Model	R-Square	AIC	SBC	Variables in Model
1	0.6744	58.8383	59.96815	x4
1	0.6661	59.1672	60.29712	x2
1	0.5342	63.4964	64.62630	x1
1	0.2860	69.0481	70.17804	x3
2	0.9787	25.3830	27.07785	x1 x2
2	0.9726	28.6828	30.37766	x1 x4
2	0.9353	39.8308	41.52565	x3 x4
2	0.8470	51.0247	52.71951	x2 x3
2	0.6799	60.6172	62.31201	x2 x4
2	0.5484	65.0933	66.78816	x1 x3
3	0.9824	24.9187	27.17852	x1 x2 x4
3	0.9823	24.9676	27.22742	x1 x2 x3
3	0.9814	25.6553	27.91511	x1 x3 x4
3	0.9730	30.4953	32.75514	x2 x3 x4
4	0.9824	26.8933	29.71808	x1 x2 x3 x4

This output was produced by the e option in the model statement of the GLM procedure. It indicates that all five regression parameters are estimable.

The GLM Procedure

General Form of Estimable Functions

Effect	Coefficients
Intercept	L1
x1	L2
x2	L3
x3	L4
x4	L5

This output was produced by the e1 option in the model statement of the GLM procedure. It describes the null hypotheses that are tested with the sequential Type I sums of squares.

Type I Estimable Functions

	Coefficients			
Effect	x1	x2	x3	x4
Intercept	0	0	0	0
x1	L2	0	0	0
x2	0.6047*L2	L3	0	0
x3	-0.8974*L2	0.0213*L3	L4	0
x4	-0.6984*L2	-1.0406*L3	-1.0281*L4	L5

Type II Estimable Functions

Effect	x1	x2	x3	x4
	----Coefficients----			
Intercept	0	0	0	0
x1	L2	0	0	0
x2	0	L3	0	0
x3	0	0	L4	0
x4	0	0	0	L5

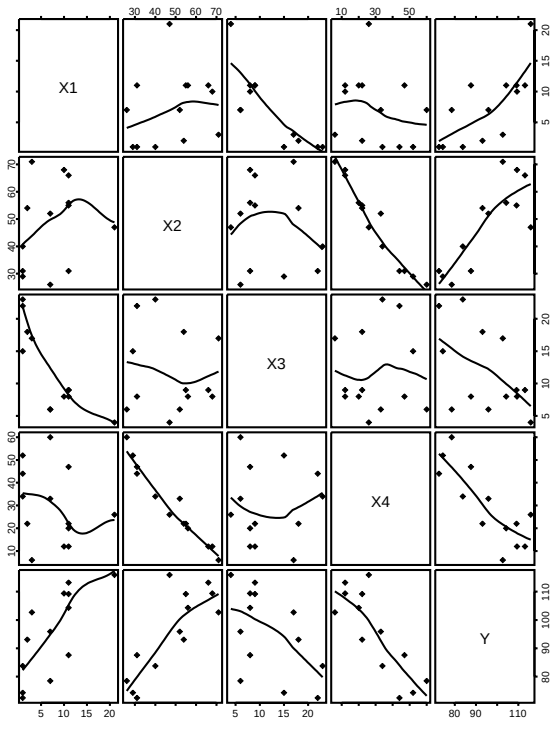

```
> # The commands are posted as:   cement.spl
>
> # The data file is stored under the name
> # cement.txt.  It has variable names on the
> # first line.  We will enter the data into
> # a data frame.

> cement<-read.table("c:\\cement.txt",header=T)

> cement

   run X1 X2 X3 X4     Y
 1   1  7 26  6 60  78.5
 2   2  1 29 15 52  74.3
 3   3 11 56  8 20 104.3
 4   4 11 31  8 47  87.6
 5   5  7 52  6 33  95.9
 6   6 11 55  9 22 109.2
 7   7  3 71 17  6 102.7
 8   8  1 31 22 44  72.5
 9   9  2 54 18 22  93.1
```



```
10  10 21 47  4 26 115.9
11  11  1 40 23 34  83.8
12  12 11 66  9 12 113.2
13  13 10 68  8 12 109.4


> #  Compute correlations and round the results
> #  to four significant digits

> round(cor(cement[-1]),4)

        X1      X2      X3      X4       Y
X1  1.0000  0.2286 -0.8241 -0.2454  0.7309
X2  0.2286  1.0000 -0.1392 -0.9730  0.8162
X3 -0.8241 -0.1392  1.0000  0.0295 -0.5348
X4 -0.2454 -0.9730  0.0295  1.0000 -0.8212
 Y  0.7309  0.8162 -0.5348 -0.8212  1.0000
```



```
> #  Create a scatterplot matrix with smooth
> #  curves.  Unix users  should first use
> #  motif( )  to open a graphics wundow

> points.lines <- function(x, y)
+   {
+   points(x, y)
+   lines(loess.smooth(x, y, 0.90))
+   }

>  par(din=c(7,7),pch=18,mkh=.15,cex=1.2,lwd=3)
>  pairs(cement[ ,-1], panel=points.lines)
```




```
> #  Use the lm( )  function to
> #  fit a  linear regression model

> cement.out <- lm(Y~X1+X2+X3+X4, cement)
> summary(cement.out)

Call: lm(formula = Y ~ X1+X2+X3+X4, data=cement)
Residuals:
    Min     1Q Median    3Q   Max
 -3.176 -1.674  0.264 1.378 3.936

Coefficients:
                      Std.
             Value    Error   t value  Pr(>|t|)
(Intercept)  63.1660  69.9338   0.9032   0.3928
         X1   1.5431   0.7433   2.0759   0.0716
         X2   0.5020   0.7224   0.6949   0.5068
         X3   0.0942   0.7532   0.1250   0.9036
         X4  -0.1515   0.7077  -0.2141   0.8358

Residual standard error: 2.441 on 8 d.f.
Multiple R-Squared: 0.9824

F-statistic: 111.8 on 4 and 8 degrees of freedom,
   the p-value is 4.707e-007
```



```
Correlation of Coefficients:
   (Intercept)      X1      X2      X3
X1 -0.9678
X2 -0.9978      0.9510
X3 -0.9769      0.9861  0.9624
X4 -0.9983      0.9568  0.9979  0.9659


> anova(cement.out)

Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
      Df Sum of Sq  Mean Sq  F Value   Pr(F)
   X1  1  1448.754 1448.754 243.0978  0.0000
   X2  1  1205.703 1205.703 202.3144  0.0000
   X3  1     9.790    9.790   1.6428  0.2358
   X4  1     0.273    0.273   0.0458  0.8358
Resid  8    47.676    5.960
```



```
> #  Create a function to evaluate an orthogonal
> #  projection matrix.  Then create a function
> #  to compute type II sums of squares.
> #  This uses the ginverse( )  function


> #=======================================
> # project( )
> #--------------
> # calculate orthogonal projection matrix
> #=======================================
> project <- function(X)
+     {X%*%ginverse(crossprod(X))%*%t(X)}
> #=======================================
```

```
> #======================================
> # typeII.SS( )
> #------------------
> # calculate Type II sum of squares
> #
> # input  lmout = object made by the
> #                  lm( ) function
> #            y = dependent variable
> #======================================
>
> typeII.SS <- function(lmout,y)
+ {
+         #  generate the model matrix
+    model <- model.matrix(lmout)
+
+            #  create list of parameter names
+    par.name <- dimnames(model)[[2]]
+
+            # compute number of parameters
+    n.par <- dim(model)[2]
+
+            # Compute residual mean square
+    SS.res <- deviance(lmout)
+  df2    <- lmout$df.resid
+  MS.res <- SS.res/df2
```



```
+  result <- NULL        # store results
+
+                        # Compute Type II SS
+  for (i in 1:n.par) {
+   A <- project(model)-project(model[,-i])
+   SS.II  <- t(y) %*% A %*% y
+   df1    <- qr(project(model))$rank -
+               qr(project(model[ ,-i]))$rank
+   MS.II  <- SS.II/df1
+   F.stat <- MS.II/MS.res
+   p.val  <- 1-pf(F.stat,df1,df2)
+   temp <- cbind(df1,SS.II,MS.II,F.stat,p.val)
+   result <- rbind(result,temp)
+   }
+
+result<-rbind(result,c(df2,SS.res,MS.res,NA,NA))
+dimnames(result)<-list(c(par.name,"Residual"),
+c("Df","Sum of Sq","Mean Sq","F Value","Pr(F)"))
+ cat("Analysis of Variance
+   (TypeII Sum of Squares) \n")
+  round(result,6)
+ }
> #=======================================
```



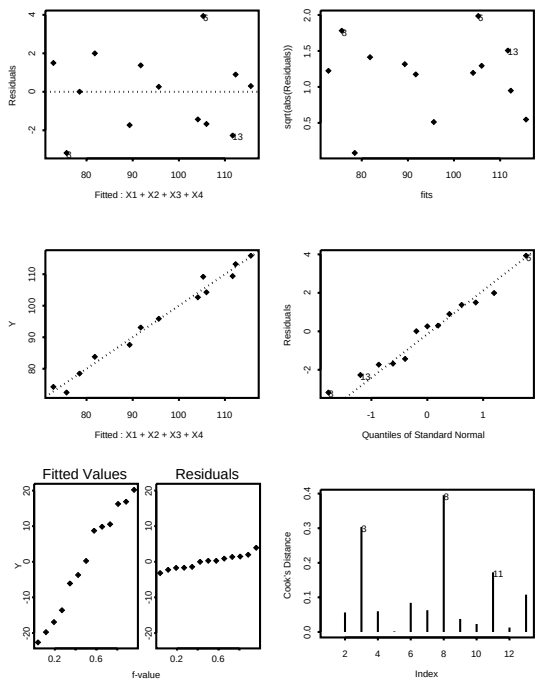
```
> typeII.SS(cement.out, cement$Y)


Analysis of Variance (TypeII Sum of Squares)
        Df Sum of Sq   Mean Sq  F Value    Pr(F)
(Inter)  1  4.861907  4.861907 0.815818 0.392790
     X1  1 25.682254 25.682254 4.309427 0.071568
     X2  1  2.878010  2.878010 0.482924 0.506779
     X3  1  0.093191  0.093191 0.015637 0.903570
     X4  1  0.273229  0.273229 0.045847 0.835810
  Resid  8 47.676412  5.959551       NA       NA
```



```
> # Venables and Ripley have supplied functions
> # studres( ) and stdres( ) to compute
> # studentized and standardized residuals.
> # Use the library( ) function to attach the
> # MASS library before using these functions.

> library(MASS)
> cement.res <- cbind(cement$Y,
+                     cement.out$fitted,
+                     cement.out$resid,
+                     studres(cement.out),
+                     stdres(cement.out))

> dimnames(cement.res) <- list(cement$run,
+    c("Response","Predicted","Residual",
+                 "Stud. Res.","Std. Res."))

> round(cement.res,4)
```

	Response	Predicted	Residual	Stud. Res.	Std. Res.
1	78.5	78.4929	0.0071	0.0040	0.0043
2	74.3	72.8005	1.4995	0.7299	0.7522
3	104.3	105.9744	-1.6744	-1.0630	-1.0545
4	87.6	89.3333	-1.7333	-0.8291	-0.8458
5	95.9	95.6360	0.2640	0.1264	0.1349
6	109.2	105.2635	3.9365	2.0324	1.7230
7	102.7	104.1289	-1.4289	-0.7128	-0.7358
8	72.5	75.6760	-3.1760	-1.9745	-1.6917
9	93.1	91.7218	1.3782	0.6472	0.6721
10	115.9	115.6010	0.2990	0.2100	0.2237
11	83.8	81.8034	1.9966	1.0919	1.0790
12	113.2	112.3007	0.8993	0.4061	0.4291
13	109.4	111.6675	-2.2675	-1.1326	-1.1131


```
> # Produce plots for model diagnostics
> # including Cook's D.

> par(mfrow=c(3,2))
> plot(cement.out)


> # Search for a simpler model

> cement.stp <- step(cement.out,
+     scope=list(upper = ~X1 + X2 + X3 + X4,
+     lower = ~ 1), trace=F)
```


```
> # Search for a simpler model

> cement.stp <- step(cement.out,
+     scope=list(upper = ~X1 + X2 + X3 + X4,
+     lower = ~ 1), trace=F)

> cement.stp$anova

Stepwise Model Path
Analysis of Deviance Table

Initial Model:
Y ~ X1 + X2 + X3 + X4

Final Model:
Y ~ X1 + X2


  Step Df Deviance Resid. Df Resid. Dev      AIC
1                          8   47.67641 107.2719
2 - X3  1 0.093191          9   47.76960  95.4460
3 - X4  1 9.970363         10   57.73997  93.4973
```