CHAPTER 21

More Geometry of Least Squares

The basic geometry of least squares appears in the foregoing chapter. Here we take things a little further by considering what happens geometrically when we test linear hypotheses of the form H_0: $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ in the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. (The alternative hypothesis is always that H_0 is false.) We suppose that $\mathbf{A}$ is a q by p matrix ($q < p$) of full rank so that the rows of $\mathbf{A}$ are linearly independent; $\boldsymbol{\beta}$ is p by 1 and $\mathbf{c}$ is a q by 1 vector of constants; $\mathbf{X}$ is n by p, and assumed to be of full rank p.

21.1. THE GEOMETRY OF A NULL HYPOTHESIS: A SIMPLE EXAMPLE

We first consider the simple example of fitting a straight line $Y = \beta_0 + \beta_1 X + \epsilon$ using a set of n data points represented by two n by 1 vectors $\mathbf{Y}$ and $\mathbf{X}_1$. In matrix terms $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, we thus have the model function

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{1}, \mathbf{X}_1)\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = \beta_0\mathbf{1} + \beta_1\mathbf{X}_1. \tag{21.1.1}$$

The estimation space Ω is a plane spanned by $\mathbf{1}$ and $\mathbf{X}_1$. Suppose $\mathbf{A} = (2, -1)$ and $\mathbf{c} = 4$ so that $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ implies $2\beta_0 - \beta_1 = 4$. Obviously $p = 2$ and $q = 1$. We substitute in (21.1.1) for β_1 to obtain for the model under $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$,

$$\mathbf{X}\boldsymbol{\beta} = \beta_0(\mathbf{1} + 2\mathbf{X}_1) - 4\mathbf{X}_1. \tag{21.1.2}$$

The estimation space ω for this model function is a straight line swept out by combining the constant vector $-4\mathbf{X}_1$ with the variable length vector $\beta_0(\mathbf{1} + 2\mathbf{X}_1)$, as shown in Figure 21.1. The three black dots show points for which $\beta_0 = 0, 1$, and 1.5 on ω. Clearly ω is part of Ω. The space $\Omega - \omega$ is spanned by any set of vectors in Ω that are all orthogonal to ω. For our example, $p = 2$ and $q = 1$, so there is only one such vector. If we write, in (21.1.2),

$$\mathbf{u} = \beta_0\mathbf{1} + (2\beta_0 - 4)\mathbf{X}_1$$

for the vector that spans ω, an obviously orthogonal vector is

$$(\mathbf{u}'\mathbf{X}_1)\mathbf{1} - (\mathbf{u}'\mathbf{1})\mathbf{X}_1, \tag{21.1.3}$$

which spans $\Omega - \omega$.

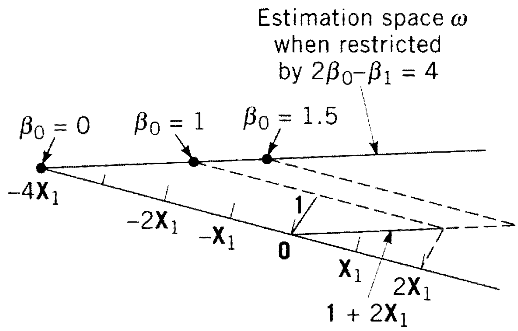

Figure 21.1. The estimation space Ω is the plane spanned by vectors $\mathbf{1}$ and $\mathbf{X}_1$. When restricted by $2\beta_0 - \beta_1 = 4$, the reduced estimation space ω is a straight line parallel to $\mathbf{1} + 2\mathbf{X}_1$ but displaced a distance equal to the length of $-4\mathbf{X}_1$.

21.2. GENERAL CASE $H_0: \mathbf{A}\boldsymbol{\beta} = \mathbf{c}$: THE PROJECTION ALGEBRA

The constant $\mathbf{c}$ is essentially an "origin choice" on ω. For purposes of defining the spaces ω and $\Omega - \omega$, we can temporarily get rid of it. Suppose $\boldsymbol{\beta}^*$ is *any* numerical choice that satisfies $\mathbf{A}\boldsymbol{\beta}^* = \mathbf{c}$. We can rewrite the model as

$$\begin{aligned} \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^* &= \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + \boldsymbol{\epsilon} \\ &= \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \end{aligned} \tag{21.2.1}$$

and for the new parameter vector $\boldsymbol{\theta}$,

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\beta} - \mathbf{A}\boldsymbol{\beta}^* = \mathbf{A}\boldsymbol{\beta} - \mathbf{c} = \mathbf{0}. \tag{21.2.2}$$

We now rewrite $\mathbf{A}\boldsymbol{\theta} = \mathbf{0}$ as

$$\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\theta}) = \mathbf{0}. \tag{21.2.3}$$

This makes it obvious that all $\mathbf{X}\boldsymbol{\theta}$ points in the $(p - q)$-dimensional space ω are orthogonal to the *columns* of the n by q matrix $\mathbf{U} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'$.

This implies that the q-dimensional $\Omega - \omega$ space is defined by the columns of $\boldsymbol{U}$, and so a unique projection matrix for $\Omega - \omega$ is given by

$$\mathbf{P}_1 = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{21.2.4}$$

Because

$$\mathbf{P} = \mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{21.2.5}$$

is the unique projection matrix for Ω, the projection matrix for ω is

$$\mathbf{P}_\omega = \mathbf{P} - \mathbf{P}_1. \tag{21.2.6}$$

We now project $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*$ via (21.2.6) to give

$$\mathbf{P}_\omega\mathbf{Y} - \mathbf{P}_\omega\mathbf{X}\boldsymbol{\beta}^* = \mathbf{P}\mathbf{Y} - \mathbf{P}\mathbf{X}\boldsymbol{\beta}^* - \mathbf{P}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*). \tag{21.2.7}$$

We note that:

(i) $\mathbf{P}_\omega\mathbf{Y} = \mathbf{X}\mathbf{b}_H$, where $\mathbf{b}_H$ is the least squares estimate of $\boldsymbol{\beta}$ in the restricted space ω.

(ii) $\mathbf{P}_\omega\mathbf{X}\boldsymbol{\beta}^* = \mathbf{X}\boldsymbol{\beta}^* = \mathbf{c}$, because the projection into ω of a vector already in ω (namely, $\mathbf{X}\boldsymbol{\beta}^*$) leaves it untouched.

(iii) $\mathbf{PY} = \mathbf{Xb}$, where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the usual (unrestricted) least squares estimator.

(iv) $\mathbf{PX}\boldsymbol{\beta}^* = \mathbf{X}\boldsymbol{\beta}^* = \mathbf{c}$; the argument is similar to (ii).

(v) $\mathbf{P}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{Ab} - \mathbf{c}).$ (21.2.8)

Putting the pieces back into (21.2.7), canceling two $\mathbf{c}$'s, and multiplying through by $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ to "cancel" $\mathbf{X}$ throughout, gives the restricted (by $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$) least squares estimate vector

$$\mathbf{b}_H = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{Ab} - \mathbf{c}). \tag{21.2.9}$$

The form of this is $\mathbf{b}$ adjusted by an amount that depends on $\mathbf{X}$, $\mathbf{A}$, and how far off $\mathbf{Ab}$ is from $\mathbf{c}$.

Properties

All three of the projection matrices are symmetric and idempotent. Note also that

$$\begin{aligned} &(a)\ \mathbf{PP}_\omega = \mathbf{P}_\omega = \mathbf{P}_\omega\mathbf{P} \\ &(b)\ \mathbf{PP}_1 = \mathbf{P}_1 = \mathbf{P}_1\mathbf{P} \\ &(c)\ \mathbf{P}_\omega\mathbf{P}_1 = \mathbf{0} = \mathbf{P}_1\mathbf{P}_\omega \end{aligned} \tag{21.2.10}$$

Geometrically, (a) means that a vector $\mathbf{Y}$ projected first into ω and then into Ω stays in ω, or that, if projected first into Ω and then into ω, finishes up in ω; (b) is a similar result. Part (c), which can be proved by writing $\mathbf{P}_1\mathbf{P}_\omega = \mathbf{P}_1(\mathbf{P} - \mathbf{P}_1) = \mathbf{P}_1\mathbf{P} - \mathbf{P}_1^2 = \mathbf{P}_1 - \mathbf{P}_1 = \mathbf{0}$, means that the split of Ω into the two subspaces, ω created by $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$, and $\Omega - \omega$, is an orthogonal split.

21.3 GEOMETRIC ILLUSTRATIONS

Figure 21.2 shows the case $n \geq 3$, $p = 2$, $q = 1$. The base plane of the figure is Ω defined by the two vectors in $\mathbf{X}$ (which are not specifically shown, but define the plane). The space ω is a straight line (shown) and the space $\Omega - \omega$ is a perpendicular straight line (not shown). The vertical dimension of the figure represents the other $(n - 2)$ dimensions. The points $\hat{\mathbf{Y}} = \mathbf{Xb}$ and $\hat{\mathbf{Y}}_H = \mathbf{Xb}_H$ are the unrestricted and restricted least squares points on Ω and ω, respectively. Note that we also show a general point $\mathbf{X}\boldsymbol{\beta}_H$ on ω. The sum of squares due to the hypothesis, $\text{SS}(H_0)$, is the squared distance between $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_H$. Via Pythagoras's theorem,

$$\text{SS}(H_0) = (\mathbf{Y} - \hat{\mathbf{Y}}_H)'(\mathbf{Y} - \hat{\mathbf{Y}}_H) - (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}), \tag{21.3.1}$$

that is, the difference between the two residual sums of squares. Also, if $\mathbf{c} = \mathbf{0}$ (so that ω includes the origin $\mathbf{O}$, and *not* as in Figure 21.1) we can write, alternatively,

$$\text{SS}(H_0) = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_H'\hat{\mathbf{Y}}_H. \tag{21.3.2}$$

[*Note:* If in the example of Section 21.1, $\mathbf{c}$ were zero, than ω would consist of the line $(\mathbf{1} + 2\mathbf{X}_1)$ and would contain the origin.]

Figure 21.3 shows the case $n > 3$, $p = 3$, $q = 1$. The point $\hat{\mathbf{Y}} = \mathbf{Xb}$ lies in the three-dimensional Ω space, and $\hat{\mathbf{Y}}_H = \mathbf{Xb}_H$ is in the base plane ω. The lines from these two points back to $\mathbf{Y}$ (not seen) are orthogonal to their associated respective spaces,

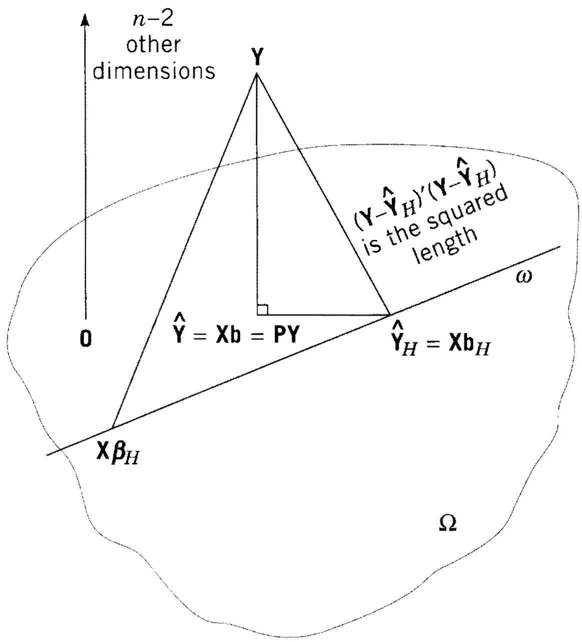

Figure 21.2. Case $n \geq 3$, $p = 2$, $q = 1$. Geometry related to $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$.

although this cannot be visualized directly in the figure. Again, the line joining $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_H$ (which is in $\Omega - \omega$) is orthogonal to ω. The two representations of $\mathrm{SS}(H_0)$ apply as before.

21.4. THE F-TEST FOR H_0, GEOMETRICALLY

The F-test (see Section 9.1) is carried out on the ratio

$$F = \{\mathrm{SS}(H_0)/q\}/s^2, \tag{21.4.1}$$

where s^2 is the residual from the full model. The appropriate degrees of freedom are

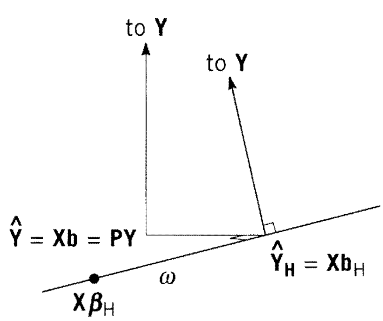

Figure 21.3. Case $n > 3$, $p = 3$, $q = 1$. Geometry related to $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$.

$\{q, n - p\}$. Figure 21.4 is a simplified version of Figure 21.3, and the letters A, B, C, D, E, and F in Figure 21.4a denote lengths. The dimensions in which these vectors lie, in the general case, are shown in Figure 21.4b. Thus

$$F = \{C^2/q\}/\{B^2/(n - p)\} \tag{21.4.2}$$

is a per degree of freedom comparison of the squared lengths C^2 and B^2. The hypothesis $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ would *not* be rejected if F were small, and would be rejected if the ratio were large.

All the routine F-tests in regression can be set up via the Ω, ω framework. For example:

1. In the pure error/lack of fit test of Section 20.7, the roles of $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_H$ in Figure 21.4a are played by $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$. See Figures 20.14, 20.15, and 21.4c.

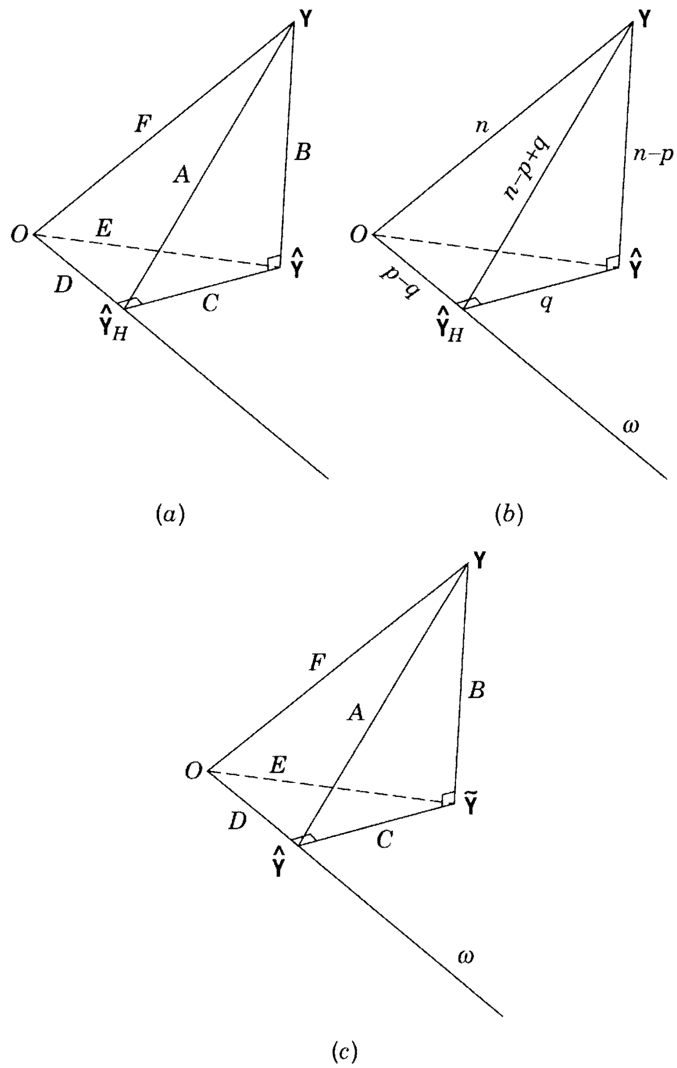

Figure 21.4. Geometry of the F-test for $H_0: \mathbf{A}\boldsymbol{\beta} = \mathbf{c}$: ($a$) lengths of vectors whose squares provide the sums of squares; (b) the dimensions (degrees of freedom of the spaces) in which those vectors lie, in the general case; and (c) an amended version of (a) appropriate for the pure error test of Section 20.7.

2. In the test for overall regression, $H_0:\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$, $\hat{\mathbf{Y}}_H$ is given by $\overline{Y}\mathbf{1}$, and $q = p - 1$ in Figure 21.4*b*. See Section 21.5 and Figure 21.5.

21.5. THE GEOMETRY OF R^2

Figure 21.5 has the same general appearance as Figure 21.4. It is in fact a special case where the initial model is

$$Y = \beta_0 + \beta_1 X_1 + \cdots \beta_{p-1} X_{p-1} + \epsilon \tag{21.5.1}$$

and the hypothesis to be tested is that of "no regression," interpreted as $H_0:\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$. Thus our restriction $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ becomes

$$[\mathbf{0}, \mathbf{I}_{p-1}]\boldsymbol{\beta} = \mathbf{0}. \tag{21.5.2}$$

The reduced model is just $Y = \beta_0 + \epsilon$, or $\mathbf{Y} = \mathbf{1}\beta_0 + \boldsymbol{\epsilon}$, so that ω is defined by the n by 1 vector $\mathbf{1}$. Thus $\hat{\mathbf{Y}}_H = \overline{Y}\mathbf{1}$. The R^2 statistic is defined as

$$\begin{aligned} R^2 &= \frac{\Sigma(\hat{Y}_i - \overline{Y})^2}{\Sigma(Y_i - \overline{Y})^2} = \frac{(\hat{\mathbf{Y}} - \overline{Y}\mathbf{1})'(\hat{\mathbf{Y}} - \overline{Y}\mathbf{1})}{(\mathbf{Y} - \overline{Y}\mathbf{1})'(\mathbf{Y} - \overline{Y}\mathbf{1})} \\ &= \frac{G^2}{K^2} \end{aligned}$$

in Figure 21.5. Special cases are $R^2 = 1$, which results when $B = 0$ (zero residual vector), and $R^2 = 0$, occurring when $G = 0$, that is, when $\hat{\mathbf{Y}} = \overline{Y}\mathbf{1}$ and there is no regression in excess of $\hat{Y}_i = \overline{Y}$.

21.6. CHANGE IN R^2 FOR MODELS NESTED VIA $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, NOT INVOLVING β_0

Figure 21.6 shows $\hat{\mathbf{Y}}$, and $\hat{\mathbf{Y}}_H$ developed through imposing the full rank hypothesis $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{A}$ is q by p and does not involve β_0. We have

$$R^2 = G^2/K^2 \quad \text{and} \quad R_H^2 = H^2/K^2. \tag{21.6.1}$$

So

$$R^2 - R_H^2 = (G^2 - H^2)/K^2 = C^2/K^2, \tag{21.6.2}$$

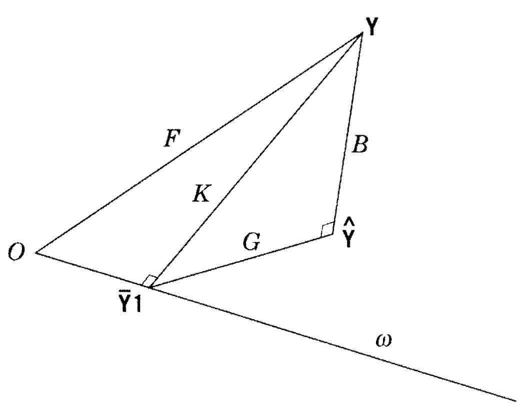

Figure 21.5. The geometry of $R^2 = G^2/K^2$.

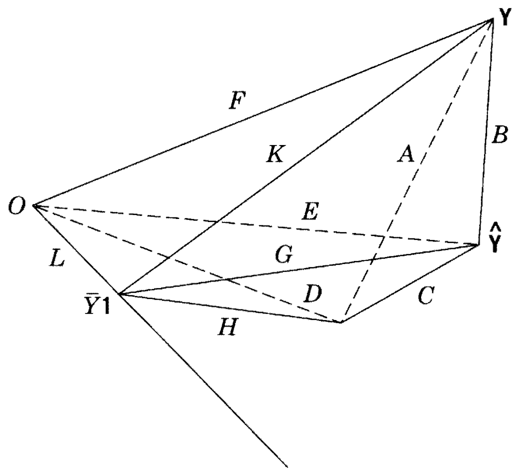

Figure 21.6. The geometry of changes in R^2 involving models nested via $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, not involving β_0.

because the lines with lengths H and C lie in orthogonal spaces spanned by P_ω and $P_\Omega - P_\omega$. Now

$$C^2 = A^2 - B^2 \tag{21.6.3}$$

is the sum of squares due to the hypothesis $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ of this section, and is tested via the F-statistic

$$F = \frac{C^2/q}{B^2/(n-p)} = \frac{n-p}{q}\cdot\frac{C^2}{B^2} \tag{21.6.4}$$

and

$$B^2 = K^2 - G^2 = K^2(1 - R^2). \tag{21.6.5}$$

Thus, from (21.6.2), (21.6.4), and (21.6.5),

$$F = \frac{n-p}{q}\cdot\frac{R^2 - R_H^2}{1 - R^2}. \tag{21.6.6}$$

This shows how the F-statistic for testing such an $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ is related to the difference in the R^2 statistics. A special case is when the points $\bar{Y}\mathbf{1}$ and $\hat{\mathbf{Y}}_H$ coincide as in Section 21.5, where $\mathbf{A} = (\mathbf{0}, \mathbf{I}_{p-1})$. We then have $R_H^2 = 0$ so that now

$$F = \frac{n-p}{q}\cdot\frac{R^2}{1 - R^2} \tag{21.6.7}$$

linking the F for testing $H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ with R^2 for the full model. If we rewrite this as

$$R^2 = \frac{qF/(n-p)}{\{qF/(n-p)\} + 1},$$

we revert to Eq. (5.3.3) with $\nu_1 = q = p - 1$ and $\nu_2 = n - p$.

21.7. MULTIPLE REGRESSION WITH TWO PREDICTOR VARIABLES AS A SEQUENCE OF STRAIGHT LINE REGRESSIONS

The stepwise selection procedure discussed in Chapter 15 involves the addition of one variable at a time to an existing equation. In this section we discuss, algebraically and geometrically, how a composite equation can be built up through a series of simple straight line regressions. Although this is not the best practical way of obtaining the final equation, it is instructive to consider how it is done. We illustrate using the steam data with the two variables X_8 and X_6. The equation obtained from the joint regression is given in Section 6.2 as

$$\hat{Y} = 9.1266 - 0.0724X_8 + 0.2029X_6.$$

Another way of obtaining this solution is as follows:

1. Regress Y on X_8. This straight line regression was performed in Chapter 1, and the resulting equation was

$$\hat{Y} = 13.6230 - 0.0798X_8.$$

This fitted equation predicts 71.44% of the variation about the mean. Adding a new variable, say, X_6 (the number of operating days), to the prediction equation might improve the prediction significantly.

In order to accomplish this, we desire to relate the number of operating days to the amount of unexplained variation in the data after the atmospheric temperature effect has been removed. However, if the atmospheric temperature variations are in any way related to the variability shown in the number of operating days, we must correct for this first. Thus we need to determine the relationship between the unexplained variation in the amount of steam used after the effect of atmospheric temperature has been removed, and the remaining variation in the number of operating days after the effect of atmospheric temperature has been removed from it.

2. Regress X_6 on X_8; calculate residuals $X_{6i} - \hat{X}_{6i}$, $i = 1, 2, \ldots, n$. A plot of X_6 against X_8 is shown in Figure 21.7. The fitted equation is

$$\hat{X}_6 = 22.1685 - 0.0367X_8.$$

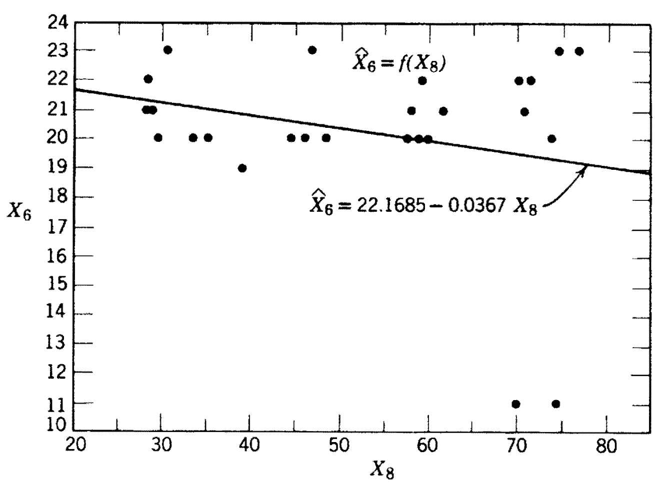

Figure 21.7. The least squares fit of X_6 on X_8.

TABLE 21.1. Residuals: $X_{6i} - \hat{X}_{6i}$

Observation Number i	X_{6i}	$\hat{X}_{6i}$	$X_{6i} - \hat{X}_{6i}$	Observation Number i	X_{6i}	$\hat{X}_{6i}$	$X_{6i} - \hat{X}_{6i}$
1	20	20.87	−0.87	14	19	20.73	−1.73
2	20	21.08	−1.08	15	23	20.45	2.55
3	23	21.04	1.96	16	20	20.39	−0.39
4	20	20.01	−0.01	17	22	19.99	2.01
5	21	19.92	1.08	18	22	19.60	2.40
6	22	19.55	2.45	19	11	19.60	−8.60
7	11	19.44	−8.44	20	23	19.44	3.56
8	23	19.36	3.64	21	20	19.53	0.47
9	21	19.58	1.42	22	21	20.04	0.96
10	20	20.06	−0.06	23	20	20.53	−0.53
11	20	20.47	−0.47	24	20	20.94	−0.94
12	21	21.11	−0.11	25	22	21.12	0.88
13	21	21.14	−0.14				

Fitted values and residuals are shown in Table 21.1. We note that there are two residuals −8.44 and −8.60 that have absolute values considerably greater than the other residuals. They arise from months in which the number of operating days was unusually small, 11 in each case. We can, of course, take the attitude that these are "outliers" and that months with so few operating days should not even be considered in the analysis. However, if we wish to obtain a satisfactory prediction equation that will be valid for *all* months, irrespective of the number of operating days, then it is important to take account of these particular results and develop an equation that makes use of the information they contain. As can be seen from the data and from Figure 21.7 and Table 21.2, if these particular months were ignored, the apparent effect of the number of operating days on the response would be small. This would *not* be because the variable did not affect the response but because the variation actually observed in the variable was so slight that the variable could not exert any appreciable effect on the response. If a variable appears to have a significant effect on the response in one analysis but not in a second, it may well be that it varied over

TABLE 21.2. Deviations of $\hat{Y}_i = f(X_8)$ and $\hat{X}_{6i} = f(X_8)$ from Y_i and X_{6i}, Respectively

Observation Number i	$Y_i - \hat{Y}_i$	$X_{6i} - \hat{X}_{6i}$	Observation Number i	$Y_i - \hat{Y}_i$	$X_{6i} - \hat{X}_{6i}$
1	0.17	−0.87	14	−0.93	−1.73
2	−0.12	−1.08	15	1.05	2.55
3	1.34	1.96	16	−0.17	−0.39
4	−0.53	−0.01	17	1.20	2.01
5	0.55	1.08	18	0.08	2.40
6	0.80	2.45	19	−1.20	−8.60
7	−1.32	−8.44	20	1.20	3.56
8	1.00	3.64	21	−0.19	0.47
9	−0.16	1.42	22	−0.51	0.96
10	0.11	−0.06	23	−1.20	−0.53
11	−1.68	−0.47	24	−0.60	−0.94
12	0.87	−0.11	25	−0.26	0.88
13	0.50	−0.14			

a wider range in the first set of data than in the second. This, incidentally, is one of the drawbacks of using plant data "as it comes." Quite often the normal operating range of a variable is so slight that no effect on response is revealed, even when the variable does, over larger ranges of operation, have an appreciable effect. Thus designed experiments, which assign levels wider than normal operating ranges, often reveal effects that had not been noticed previously.

3. We now regress $Y - \hat{Y}$ against $X_6 - \hat{X}_6$ by fitting the model

$$(Y_i - \hat{Y}_i) = \beta(X_{6i} - \hat{X}_{6i}) + \epsilon_i.$$

No "β_0" term is required in this first-order model since we are using two sets of residuals whose sums are zero, and thus the line must pass through the origin. (If we did put a β_0 term in, we should find $b_0 = 0$, in any case.) For convenience the two sets of residuals used as data are extracted from Tables 1.2 and 21.1 and are given in Table 21.2. A plot of these residuals is shown in Figure 21.8. The fitted equation takes the form

$$(\widehat{Y - \hat{Y}}) = 0.2015(X_6 - \hat{X}_6).$$

Within the parentheses we can substitute for $\hat{Y}$ and $\hat{X}_6$ as functions of X_8, and the large caret on the left-hand side can then be attached to Y to represent the overall fitted value $\hat{Y} = \hat{Y}(X_6, X_8)$ as follows:

$$[\hat{Y} - (13.6230 - 0.0798X_8)] = 0.2015[X_6 - (22.1685 - 0.0367X_8)]$$

or

$$\hat{Y} = 9.1560 - 0.0724X_8 + 0.2015X_6.$$

The previous result was

$$\hat{Y} = 9.1266 - 0.0724X_8 + 0.2029X_6.$$

In theory these two results are identical; practically, as we can see, discrepancies have occurred due to rounding errors. Ignoring rounding errors for the moment, we shall

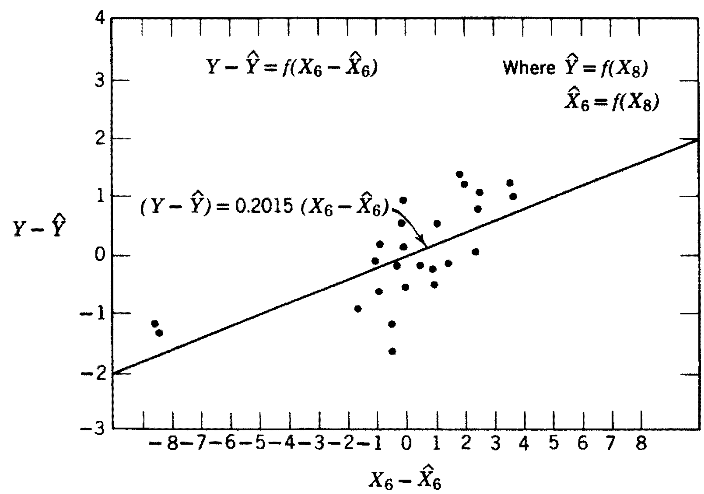

Figure 21.8. A plot of the residuals in Table 21.2.

now show, geometrically, through a simple example, why the two methods should provide us with identical results.

Geometrical Interpretation

Consider an example in which we have $n = 3$ observations of the response Y, namely, Y_1, Y_2, and Y_3 taken at the three sets of conditions: (X_1, Z_1), (X_2, Z_2), (X_3, Z_3). We can plot in three dimensions on axes labeled 1, 2, and 3, with origin at O, the points $Y \equiv (Y_1, Y_2, Y_3), X \equiv (X_1, X_2, X_3)$, and $Z \equiv (Z_1, Z_2, Z_3)$. The geometrical interpretation of regression is as follows. To regress Y on X we drop a perpendicular YP onto OX. The coordinates of the point P are the fitted values $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_3$. The length OP^2 is the sum of squares due to the regression, OY^2 is the total sum of squares, and YP^2 is the residual sum of squares. By Pythagoras, $OP^2 + YP^2 = OY^2$, which provides the analysis of variance breakup of the sums of squares (see Figure 21.9).

If we complete the parallelogram, which has OY as diagonal and OP and PY as sides, we obtain the parallelogram $OP'YP$ as shown. Then the coordinates of P' are the values of the residuals from the regression of variable Y on variable X. In vector terms we could write

$$\overrightarrow{OP} + \overrightarrow{OP}' = \overrightarrow{OY},$$

or, in "statistical" vector notation,

$$\hat{\mathbf{Y}} + (\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{Y}.$$

This result is true in general for n dimensions. (The only reason we take $n = 3$ is so we can provide a diagram.)

Suppose we wish to regress variable Y on variables X and Z simultaneously. The lines OX and OZ define a plane in three dimensions. We drop a perpendicular YT onto this plane. Then the coordinates of the point T are the fitted values $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_3$ for *this* regression. OT^2 is the regression sum of squares, YT^2 is the residual sum of squares, and OY^2 is the total sum of squares. Again, by Pythagoras, $OY^2 = OT^2 + YT^2$, which, again, gives the sum of squares breakup we see in the analysis of variance table. Completion of the parallelogram $OT'YT$ with diagonal OY and sides OT and TY provides OT', the vector of residuals of this regression, and the coordinates of T' give the residuals $\{(Y_1 - \hat{Y}_1), (Y_2 - \hat{Y}_2), (Y_3 - \hat{Y}_3)\}$ of the regression of Y on X and Z simultaneously. Again, in vector notation,

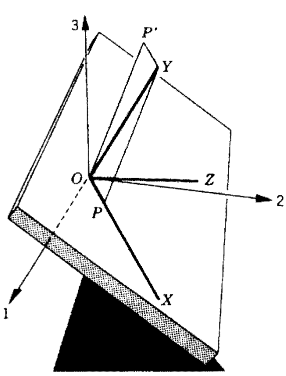

Figure 21.9. Geometrical interpretation of the regression of Y on X.

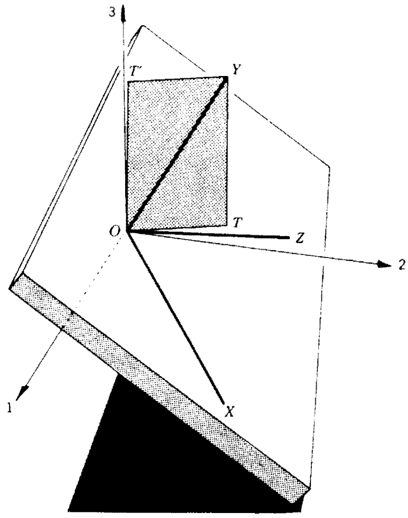

Figure 21.10. Geometrical interpretation of the regression of Y on X and Z.

$$\overrightarrow{OT} + \overrightarrow{OT}' = \overrightarrow{OY}$$

or, in "statistical" vector notation,

$$\hat{\mathbf{Y}} + (\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{Y}$$

for this regression (see Figure 21.10).

As we saw in the numerical example above, the same final residuals should arise (ignoring rounding) if we do the regressions (1) Y on X, and (2) Z on X, and then regress the residuals of (1) on the residuals of (2). That this is true can be seen geometrically as follows. Figure 21.11 shows three parallelograms in three-dimensional space.

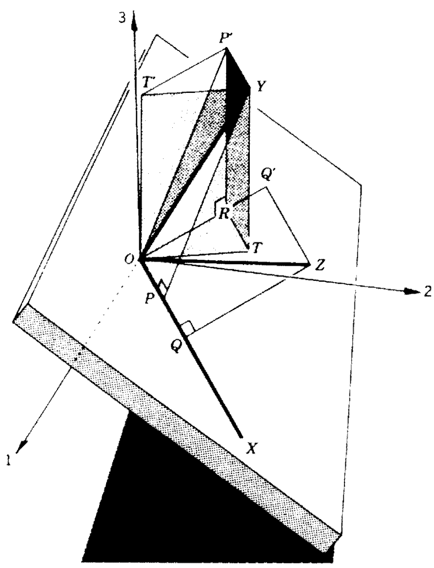

Figure 21.11. The regression of Y on X and Z can also be viewed as a two-step procedure as described in the text.

1. $OP'YP$ from the regression of Y on X.
2. $OQ'ZQ$ from the regression of Z on X.
3. $OT'YT$ from the regression of Y on X and Z simultaneously.

Now the regression of the residuals of (1) onto the residuals of (2) is achieved by dropping the perpendicular from P' onto OQ'. Suppose the point of impact is R. Then a line through O parallel to RP' and of length RP' will be the residual vector of the two-step regression of Y on X and Z. However, the points O, Q', Z, P, Q, X, and T all lie in the plane π defined by OZ and OX. Thus so does the point R. Since $OP'YP$ is a parallelogram, and $P'R$ and YT are perpendicular to plane π, $P'R = YT$ in length. Since $TY = OT'$, it follows that $OT' = RP'$. But OT', RP', and TY are all parallel and perpendicular to plane π. Hence $OT'P'R$ is a parallelogram from which it follows that $\overrightarrow{OT}'$ is the vector of residuals from the two-step regression. Since it originally resulted from the regression of Y on Z and X together, the two methods must be equivalent. Thus we can see that the planar regression of Y on X and Z together can be regarded as the totality of successive straight line regressions of:

1. Y on X,
2. Z on X, and
3. Residuals of (1) on the residuals of (2).

The same result is obtained if the roles of Z and X are interchanged. All linear regressions can be broken down into a series of simple regressions in this way.

EXERCISES FOR CHAPTER 21

A. Suppose

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = \begin{array}{c} \begin{matrix} 1 & X_1 & X_2 \end{matrix} \\ \left[\begin{array}{cc|c} 1 & -3 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 3 & 1 \end{array}\right] \end{array}$$

Let ω be the space defined by the $(1, X_1)$ columns and Ω be the space defined by the $(1, X_1, X_2)$ columns.

1. Evaluate $\mathbf{P}_\Omega$, $\mathbf{P}_\omega$ and their difference. Give the dimensions of the spaces spanned by their columns.
2. Show, through your example data, that the general result

$$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{P}_{\omega^\perp \cap \Omega},$$

where $\omega^\perp$ is the complement of ω with respect to the full four-dimensional space E_4, is true.

3. Give a basis for (i.e., a set of vectors that span) $R(P_{\omega^\perp \cap \Omega})$.
4. What are the eigenvalues of $\mathbf{P}_\Omega - \mathbf{P}_\omega$? (Write them down without detailed calculations if you wish, but explain how you did this.) What theorem does your answer confirm?

B. Use Eq. (21.2.9) to fit the model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$, subject to $\beta_0 - 2\beta_1 = 4$, to the seven data points:

X	Y
-1	15, 18
0	18, 21, 22
1	28, 30

C. In *The American Statistician*, Volume 45, No. 4, November 1991, pp. 300–301, A. K. Shah notes, in a numerical example based on Exercise 15H with $Y = X_6$ and $X = X_5$, that if he regresses Y versus X, he gets $R^2 = 0.0782$, while if he regresses $Y - X$ versus X, he gets $R^2 = 0.6995$. He also notes that he gets the same residual SS in both cases and that the fitted relationship between Y and X is exactly the same, namely, $\hat{Y} = 4.6884 - 0.2360X$.

Look at this geometrically and draw diagrams showing how this sort of thing can happen. Show also that we can have a "reverse" case, where R^2 will decrease, not increase.

D. Show (via the method of Lagrange's undetermined multipliers) that Eq. (21.2.9) can also be obtained by minimizing the sum of squares function $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$.

E. See (21.2.10). Confirm that the projection matrices $\mathbf{P}$, $\mathbf{P}_1 = \mathbf{P}_{\Omega-\omega}$, and $\mathbf{P}_\omega$ are all symmetric and idempotent. Also confirm that $\mathbf{P}_\omega$ and $\mathbf{P}_1$ are orthogonal.

F. Show, for Section 21.5, when $(\mathbf{0}, \mathbf{I}_{p-1})\boldsymbol{\beta} = \mathbf{0}$ determines $\boldsymbol{\omega}$, that $\boldsymbol{\omega}$ is spanned by $\mathbf{1}$. Do this via the (unnecessarily complicated for this example) method of Eqs. (21.2.4) to (21.2.6), and using the formula for the partitioned inverse of $\mathbf{X}'\mathbf{X}$, when $\mathbf{X} = [\mathbf{1}, \mathbf{X}_1]$, where here $\mathbf{X}_1$ is n by $(p - 1)$. Refer to Appendix 5A.

G. (*Sources:* The data below are quoted from p. 165 of *Linear Models* by S. R. Searle, published in 1971 by John Wiley & Sons. Searle's source is a larger set of data given by W. T. Federer in *Experimental Design*, published in 1955 by MacMillan, p. 92. The "Searle data" also feature in these papers: "A matrix identity and its applications to equivalent hypotheses in linear models," by N. N. Chan and K.-H. Li, *Communications in Statistics, Theory and Methods*, **24**, 1995, 2769–2777; and "Nontestable hypotheses in linear models," by S. R. Searle, W. H. Swallow, and C. E. McCulloch, *SIAM Journal of Algebraic and Discrete Methods*, **5**, 1984, 486–496.) Consider the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 101 \\ 105 \\ 94 \\ 84 \\ 88 \\ 32 \end{bmatrix}.$$

It is desired to test the hypothesis $H_0: \beta_1 = 7$, $\beta_2 = 4$, versus the alternative "not so." Is this possible? If not, what *is* it possible to test?

H. Suppose we have five observations of Y at five coded X-values, $-2, -1, 0, 1, 2$. Consider the problem of fitting the quadratic equation $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$ and testing the null hypothesis $H_0: \beta_2 = 0$ versus $H_1: \beta_2 \neq 0$. The null hypothesis divides the estimation space Ω (whose projection matrix is $\mathbf{P}$) into the subspaces: $\boldsymbol{\omega}$ (defined by H_0 and with projection matrix $\mathbf{P}_\omega$) and $\boldsymbol{\Omega} - \boldsymbol{\omega}$ (with projection matrix $\mathbf{P}_1$).

1. Find $\mathbf{P}$.
2. Find $\mathbf{P}_\omega$.
3. Find $\mathbf{P}_1$.
4. Confirm that equations (21.2.10) are true for this example.
5. Evaluate $\mathbf{P}_1\mathbf{x}$, where $\mathbf{x} = (-2, -1, 0, 1, 2)'$, and then explain why your answer is obvious.