(iii) $\mathbf{PX} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$. □

COROLLARY If $\mathbf{X}$ has rank r $(r < p)$, then Theorem 3.1 still holds, but with p replaced by r.

Proof. Let $\mathbf{X}_1$ be an $n \times r$ matrix with r linearly independent columns and having the same column space as $\mathbf{X}$ [i.e., $\mathcal{C}(\mathbf{X}_1) = \Omega$]. Then $\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, and (i) and (ii) follow immediately. We can find a matrix $\mathbf{L}$ such that $\mathbf{X} = \mathbf{X}_1\mathbf{L}$, which implies that (cf. Exercises 3j, No. 2)

$$\mathbf{PX} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_1\mathbf{L} = \mathbf{X}_1\mathbf{L} = \mathbf{X},$$

which is (iii). □

EXERCISES 3a

1. Show that if $\mathbf{X}$ has full rank, $(Y - X\beta) = (Y - X\hat{\beta} + X(\hat{\beta} - \beta))$,

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

Cross terms cancel b/c $Y - X\hat{\beta} \perp C(X)$

and hence deduce that the left side is minimized uniquely when $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$.

2. If $\mathbf{X}$ has full rank, prove that $\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0$. *Hint*: Consider the first column of $\mathbf{X}$. By normal equations & $x_0 = 1_n$.

3. Let

$$\begin{aligned} Y_1 &= \theta + \varepsilon_1 \\ Y_2 &= 2\theta - \phi + \varepsilon_2 \\ Y_3 &= \theta + 2\phi + \varepsilon_3, \end{aligned}$$

Solve w/ $X = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}$

where $E[\varepsilon_i] = 0$ $(i = 1, 2, 3)$. Find the least squares estimates of θ and ϕ.

4. Consider the regression model

$$E[Y_i] = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2) \qquad (i = 1, 2, 3),$$

$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \to$ orthogonal so coeffs don't change.

where $x_1 = -1$, $x_2 = 0$, and $x_3 = +1$. Find the least squares estimates of β_0, β_1, and β_2. Show that the least squares estimates of β_0 and β_1 are unchanged if $\beta_2 = 0$.

5. The tension T observed in a nonextensible string required to maintain a body of unknown weight w in equilibrium on a smooth inclined plane of angle θ $(0 < \theta < \pi/2)$ is a random variable with mean $E[T] = w \sin\theta$. If for $\theta = \theta_i$ $(i = 1, 2, \ldots, n)$ the corresponding values of T are T_i $(i = 1, 2, \ldots, n)$, find the least squares estimate of w.

$X = \begin{pmatrix} \sin\theta_1 \\ \sin\theta_2 \\ \vdots \end{pmatrix}$ $T = \begin{pmatrix} T_1 \\ \vdots \end{pmatrix}$

6. If $\mathbf{X}$ has full rank, so that $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, prove that $\mathcal{C}(\mathbf{P}) = \mathcal{C}(\mathbf{X})$.

$C(P) \subseteq C(X)$ b/c $PX = X$; rank P = rank $X = p$

$W = X(X^TX)^{-1}X^TT$

 $\hat{Y} = X(\underbrace{PY}_{\hat{\beta}})$

7. For a general regression model in which **X** may or may not have full rank, show that

Recall $Y - X\hat{\beta} \perp C(X)$ $$\sum_{i=1}^{n} \hat{Y}_i(Y_i - \hat{Y}_i) = 0.$$ by minimality. since $\hat{Y} \in C(X)$, ortho.

8. Suppose that we scale the explanatory variables so that $x_{ij} = k_j w_{ij}$ for all i, j. By expressing **X** in terms of a new matrix **W**, prove that $\hat{\mathbf{Y}}$ remains unchanged under this change of scale.

For $k_j \neq 0$, $X = W \operatorname{diag}(k)$ so $\hat{\beta}_W = \hat{\beta}_X / k$

3.2 PROPERTIES OF LEAST SQUARES ESTIMATES

If we assume that the errors are unbiased (i.e., $E[\boldsymbol{\varepsilon}] = \mathbf{0}$), and the columns of **X** are linearly independent, then

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}, \end{aligned} \tag{3.10}$$

and $\hat{\boldsymbol{\beta}}$ is an unbiased estimate of $\boldsymbol{\beta}$. If we assume further that the ε_i are uncorrelated and have the same variance, that is, $\operatorname{cov}[\varepsilon_i, \varepsilon_j] = \delta_{ij}\sigma^2$, then $\operatorname{Var}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n$ and

$$\operatorname{Var}[\mathbf{Y}] = \operatorname{Var}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] = \operatorname{Var}[\boldsymbol{\varepsilon}].$$

Hence, by (1.7),

$$\begin{aligned} \operatorname{Var}[\hat{\boldsymbol{\beta}}] &= \operatorname{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \operatorname{Var}[\mathbf{Y}]\, \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{3.11}$$

The question now arises as to why we chose $\hat{\boldsymbol{\beta}}$ as our estimate of $\boldsymbol{\beta}$ and not some other estimate. We show below that for a reasonable class of estimates, $\hat{\beta}_j$ is the estimate of β_j with the smallest variance. Here $\hat{\beta}_j$ can be extracted from $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{p-1})'$ simply by premultiplying by the row vector $\mathbf{c}'$, which contains unity in the $(j+1)$th position and zeros elsewhere. It transpires that this special property of $\hat{\beta}_j$ can be generalized to the case of any linear combination $\mathbf{a}'\hat{\boldsymbol{\beta}}$ using the following theorem.

THEOREM 3.2 *Let $\hat{\boldsymbol{\theta}}$ be the least squares estimate of $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\theta} \in \Omega = \mathcal{C}(\mathbf{X})$ and $\mathbf{X}$ may not have full rank. Then among the class of linear unbiased estimates of $\mathbf{c}'\boldsymbol{\theta}$, $\mathbf{c}'\hat{\boldsymbol{\theta}}$ is the unique estimate with minimum variance. [We say that $\mathbf{c}'\hat{\boldsymbol{\theta}}$ is the best linear unbiased estimate (BLUE) of $\mathbf{c}'\boldsymbol{\theta}$.]*