THE DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

PRACTICAL II SOLUTIONS

II.1 Let x denote the number of years of formal education and let Y denote an individual's income at age 30. Assume that simple linear regression is applicable and consider this data:

Formal education (years)	Income ($000)
8	8
12	15
14	16
16	20
16	25
20	40

a) Find **y, X** and $\boldsymbol{\theta}$

$$\mathbf{y}=\begin{bmatrix}8\\15\\16\\20\\25\\40\end{bmatrix},\quad \mathbf{X}=\begin{bmatrix}1 & 8\\1 & 12\\1 & 14\\1 & 16\\1 & 16\\1 & 20\end{bmatrix}\quad\text{and}\quad \boldsymbol{\theta}=\begin{bmatrix}\theta_0\\\theta_2\end{bmatrix}$$

b) Find $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$ and $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}6 & 86\\86 & 1316\end{bmatrix},\quad \mathbf{X}'\mathbf{y}=\begin{bmatrix}124\\1988\end{bmatrix}\ \text{and}$$

$$\begin{aligned}\left(\mathbf{X}'\mathbf{X}\right)^{-1}&=\frac{1}{6\times1316-86\times86}\begin{bmatrix}1316 & -86\\-86 & 6\end{bmatrix}\\&=\frac{1}{500}\begin{bmatrix}1316 & -86\\-86 & 6\end{bmatrix}\\&=\begin{bmatrix}2.632 & -0.172\\-0.172 & 0.012\end{bmatrix}\end{aligned}$$

c) Find the least squares estimates for $\boldsymbol{\theta}$ by finding $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$

$$\hat{\boldsymbol{\theta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}=\begin{bmatrix}2.632 & -0.172\\-0.172 & 0.012\end{bmatrix}\begin{bmatrix}124\\1988\end{bmatrix}=\begin{bmatrix}-15.568\\2.528\end{bmatrix}$$

d) Estimate the average salary of individuals who have had 15 years of formal education

For x=15, $\widehat{E[Y]} = \begin{bmatrix} 1 & 15 \end{bmatrix} \begin{bmatrix} -15.568 \\ 2.528 \end{bmatrix} = 22.352$

II.2 Find $E[\mathbf{a}'\mathbf{Y}]$ for

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}, \quad \boldsymbol{\psi} = \begin{bmatrix} 3 \\ 7 \\ 1 \end{bmatrix} \text{ and } \mathbf{a} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

$$E[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'E[\mathbf{Y}] = \mathbf{a}'\boldsymbol{\psi} = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 1 \end{bmatrix} = 12$$

II.3 Show that the quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j$ for

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}$$

$$\begin{aligned}
\mathbf{y}'\mathbf{A}\mathbf{y} &= \begin{bmatrix} y_1 & y_2 & \cdots & y_k \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \\
&= \begin{bmatrix} y_1 & y_2 & \cdots & y_k \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{k} a_{1j} y_j \\ \sum_{j=1}^{k} a_{2j} y_j \\ \vdots \\ \sum_{j=1}^{k} a_{kj} y_j \end{bmatrix} \\
&= y_1 \sum_{j=1}^{k} a_{1j} y_j + y_2 \sum_{j=1}^{k} a_{2j} y_j + \ldots + y_k \sum_{j=1}^{k} a_{kj} y_j \\
&= \sum_{i=1}^{k} y_i \sum_{j=1}^{k} a_{ij} y_j \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} y_i a_{ij} y_j \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j
\end{aligned}$$

II.4 Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} 2 & 4 \\ 1 & 6 \end{bmatrix}$$

Find $\mathbf{y}'\mathbf{A}\mathbf{y}$ via the addition formula in the previous exercise and by direct matrix multiplication. Why do you think $\mathbf{y}'\mathbf{A}\mathbf{y}$ is called a quadratic form?

$$\begin{aligned}\mathbf{y}'\mathbf{A}\mathbf{y} &= \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j \\ &= 2y_1y_1 + 4y_1y_2 + 1y_2y_1 + 6y_2y_2 \\ &= 2y_1^2 + 5y_1y_2 + 6y_2^2\end{aligned}$$

$$\begin{aligned}\mathbf{y}'\mathbf{A}\mathbf{y} &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 2y_1 + 4y_2 \\ 1y_1 + 6y_2 \end{bmatrix} \\ &= 2y_1^2 + 5y_1y_2 + 6y_2^2\end{aligned}$$

It is called a quadratic form because all the terms involve second-order terms in y, that is, either y_i^2 *or* $y_i y_j$

II.5 Verify that $\mathbf{V} = E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]$ is equivalent to the following expression for $\mathbf{V}$ by obtaining an expression for the *ij*th element of $E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]$.

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1i} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2i} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{1i} & \sigma_{2i} & \cdots & \sigma_i^2 & \cdots & \sigma_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{in} & \cdots & \sigma_n^2 \end{bmatrix}$$

The ijth element of $\mathbf{V} = E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]$ *is the expectation of the product of the ith and jth elements of* $\left(\mathbf{Y} - E[\mathbf{Y}]\right)$. *The ith element of* $\left(\mathbf{Y} - E[\mathbf{Y}]\right)$ *is* $\left(\mathbf{Y}_i - E[\mathbf{Y}_i]\right)$ *so that the ijth element of* $\mathbf{V} = E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]$ *is* $E\left[\left(\mathbf{Y}_i - E[\mathbf{Y}_i]\right)\left(\mathbf{Y}_j - E[\mathbf{Y}_j]\right)\right]$. *By definition this is* σ_{ij} *which for* $i = j$ *is* σ_i^2. *The two expressions are equivalent.*

II.6 Prove that the least squares estimator of $\boldsymbol{\theta}$, given by $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, is unbiased. Also prove that $\operatorname{var}\left[\hat{\boldsymbol{\theta}}\right] = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$.

To prove that $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$, need to show that $E\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$.
Let $\mathbf{L} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Using theorem II.3, we have

$$E\left[\hat{\boldsymbol{\theta}}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right] = E[\mathbf{L}\mathbf{Y}] = \mathbf{L}E[\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}]$$

Now $E[\mathbf{Y}] = \boldsymbol{\theta}$ so that

$$E\left[\hat{\boldsymbol{\theta}}\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\theta} = \boldsymbol{\theta}$$

Similarly, using theorem II.3,

$$\operatorname{var}\left[\hat{\boldsymbol{\theta}}\right] = \operatorname{var}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right] = \operatorname{var}[\mathbf{L}\mathbf{Y}] = \mathbf{L}\operatorname{var}[\mathbf{Y}]\mathbf{L}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{var}[\mathbf{Y}]\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}'$$

Now $\operatorname{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ so that

$$\begin{aligned}
\operatorname{var}\left[\hat{\boldsymbol{\theta}}\right] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{var}[\mathbf{Y}]\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}' \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

II.7 Use theorem I.6 to prove that $E[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'E[\mathbf{Y}]$. (Hint: write down the summation formula for $\mathbf{a}'\mathbf{Y}$.)

In summation notation, $\mathbf{a}'\mathbf{Y} = \sum_{i=1}^{n} a_i Y_i$ so that $E[\mathbf{a}'\mathbf{Y}] = E\left[\sum_{i=1}^{n} a_i Y_i\right]$. Using theorems I.5/I.12 with $t = n$, $c_i = a_i$ and $u_i(Y_1, Y_2, \ldots, Y_n) = Y_i$

$$\begin{aligned}
E[\mathbf{a}'\mathbf{Y}] &= \sum_{i=1}^{n} a_i E[Y_i] \\
&= \mathbf{a}'E[\mathbf{Y}]
\end{aligned}$$

Also prove that $\operatorname{var}[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'\operatorname{var}[\mathbf{Y}]\mathbf{a}$ using the following steps:

a) Use the definition of the variance of a random variable to obtain an expression for the variance of the random variable $\mathbf{a}'\mathbf{Y}$.

Now, being a scalar function of random variables, $\mathbf{a}'\mathbf{Y}$ *is itself a random variable. From definition I.5,* $\operatorname{var}[\mathbf{a}'\mathbf{Y}] = E\left[\left(\mathbf{a}'\mathbf{Y} - E[\mathbf{a}'\mathbf{Y}]\right)^2\right] = E\left[\left\{\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\right\}^2\right]$.

b) Show that $\operatorname{var}[\mathbf{a}'\mathbf{Y}] = E\left[\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\mathbf{a}\right]$.

As $\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)$ *is a scalar,* $\left\{\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\right\}' = \mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)$ *and*

$$\begin{aligned}
\operatorname{var}[\mathbf{a}'\mathbf{Y}] &= E\left[\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\right] \\
&= E\left[\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left\{\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\right\}'\right] \\
&= E\left[\mathbf{a}'\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\mathbf{a}\right]
\end{aligned}$$

c) Let S_{ij} be the element from the *i*th row and *j*th column of $\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'$ and use the result from b) to obtain an expression for $var[\mathbf{a}'\mathbf{Y}]$ in terms of these elements.

$$\begin{aligned}
\operatorname{var}[\mathbf{a}'\mathbf{Y}] &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j S_{ij}\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left[S_{ij}\right]
\end{aligned}$$

d) Obtain the required result.

$$\begin{aligned}
\operatorname{var}[\mathbf{a}'\mathbf{Y}] &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j S_{ij}\right] \\
&= \mathbf{a}'E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]\mathbf{a} \\
&= \mathbf{a}'\operatorname{var}[\mathbf{Y}]\mathbf{a}
\end{aligned}$$

II.8 Let $\mathbf{Y}$ be a normally distributed random vector representing a random sample with $E[\mathbf{Y}] = \mathbf{X\theta}$ and $\text{var}[\mathbf{Y}] = \mathbf{V}_{\mathbf{Y}} = \sigma^2\mathbf{I}_n$ where $\mathbf{X}$ is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters and $n \geq q$. Prove that the maximum likelihood estimator for σ^2 is given by

$$\tilde{\sigma}_n^2 = \frac{\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)'\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)}{n} = \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{n}$$

The maximum likelihood estimate of σ^2 is given by $\partial\ell/\partial\sigma^2 = 0$ as follows:

$$\begin{aligned}
\frac{\partial\ell}{\partial\sigma^2} &= \frac{\partial\left\{-\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\theta}\right)'\left(\mathbf{y}-\mathbf{X\theta}\right)\right\}}{\partial\sigma^2} \\
&= -\left\{\left(\frac{n}{2}\right)\left(\frac{2\pi}{2\pi\sigma^2}\right) - \left(\frac{1}{2\left[\sigma^2\right]^2}\right)\left(\mathbf{y}-\mathbf{X\theta}\right)'\left(\mathbf{y}-\mathbf{X\theta}\right)\right\} \\
&= 0
\end{aligned}$$

which implies

$$\left(\frac{n}{2\sigma^2}\right) = \left(\frac{1}{2\left[\sigma^2\right]^2}\right)\left(\mathbf{y}-\mathbf{X\theta}\right)'\left(\mathbf{y}-\mathbf{X\theta}\right)$$

so that the estimate is given by

$$\tilde{\sigma}^2 = \frac{\left(\mathbf{y}-\mathbf{X\theta}\right)'\left(\mathbf{y}-\mathbf{X\theta}\right)}{n}$$

so that the estimator is

$$\tilde{\sigma}_n^2 = \frac{\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)'\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)}{n} = \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{n}$$

Is this estimator unbiased?

Theorem II.6 tells us that the estimator is biased.

II.9 Show that the matrices $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{R}_\mathbf{X} = \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)$ are symmetric and idempotent. (Note: the inverse of a symmetric matrix is also symmetric.)

$$\mathbf{P}'_\mathbf{X} = \left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}' = \mathbf{X}\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}'\mathbf{X}'$$

and $(\mathbf{X}'\mathbf{X})' = \mathbf{X}'\mathbf{X}$ *so that* $\mathbf{X}'\mathbf{X}$ *is symmetric and so* $(\mathbf{X}'\mathbf{X})^{-1}$ *must also be symmetric. Hence,*

$$\mathbf{P}'_\mathbf{X} = \mathbf{X}\left\{(\mathbf{X}'\mathbf{X})^{-1}\right\}'\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}_\mathbf{X}$$

and $\mathbf{P}_\mathbf{X}$ *is symmetric.*

$$\mathbf{P}^2_\mathbf{X} = \left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}\left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}_\mathbf{X}$$

and $\mathbf{P}_\mathbf{X}$ *is idempotent.*

$$\mathbf{R}'_\mathbf{X} = \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' = \mathbf{I}'_n - \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' = \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) = \mathbf{R}_\mathbf{X}$$

and $\mathbf{R}_\mathbf{X}$ *is symmetric.*

$$\begin{aligned}
\mathbf{R}^2_\mathbf{X} &= \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \mathbf{I}_n - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
&= \mathbf{I}_n - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
&= \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \mathbf{R}_\mathbf{X}
\end{aligned}$$

and $\mathbf{R}_\mathbf{X}$ *is idempotent.*

II.10 In exercise II.1 we considered the following data.

Formal education (years)	Income ($000)
8	8
12	15
14	16
16	20
16	25
20	40

We found $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} -15.568 \\ 2.528 \end{bmatrix}$.

a) Compute the fitted values and residuals

Formal education (years)	Income ($000)	Fitted values	Residuals
8	8	4.656	3.344
12	15	14.768	0.232
14	16	19.824	–3.824
16	20	24.880	–4.880
16	25	24.880	0.120
20	40	34.992	5.008

b) Find the Total, Regression and Residual sums of squares.

Total	3170
Regression	3095.232
Residual	74.768

c) Construct the ANOVA table for testing H_0: $\theta = \mathbf{0}$ versus H_a: $\theta \neq \mathbf{0}$.

Source	**DF**	**SS**	**MSq**	**F**
Regression	2	3095.232	1547.616	82.7956
Residual	4	74.768	18.692	
Total	6	3170.000		

d) What is the value of the correction factor for this example?

$$\left(\sum_{i=1}^{n} Y_i\right)^2 \Big/ n = (8+15+16+20+25+40)/6 = 124^2/6 = 2562.667$$

e) Construct the ANOVA table for testing that the slope is zero given that the intercept is in the model.

Source	**DF**	**SS**	**MSq**	**F**
Regression	1	532.653	532.653	28.4963
Residual	4	74.768	18.692	
Total (corrected)	5	607.333		

II.11 In theorem II.21, it was asserted that $\mathbf{P}_{\mathbf{X}_1/\mathbf{X}_2}\mathbf{R}_{\mathbf{X}} = \mathbf{0}$. Prove this result.

First recall that $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ *is idempotent and that* $\mathbf{X}_2'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}_2'$. *Then*

$$\begin{aligned}
\mathbf{P}_{\mathbf{X}_1|\mathbf{X}_2}\mathbf{R}_{\mathbf{X}} &= \left\{\mathbf{P}_{\mathrm{X}} - \mathbf{P}_{\mathrm{X}_2}\right\}\left\{\mathbf{I}_n - \mathbf{P}_{\mathrm{X}}\right\} \\
&= -\mathbf{P}_{\mathrm{X}_2}\left\{\mathbf{I}_n - \mathbf{P}_{\mathrm{X}}\right\} \text{ as } \mathbf{P}_{\mathrm{X}}\left\{\mathbf{I}_n - \mathbf{P}_{\mathrm{X}}\right\} = \mathbf{0} \\
&= -\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\left\{\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right\} \\
&= -\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2' + \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \\
&= -\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2' + \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2' \\
&= \mathbf{0}
\end{aligned}$$

II.12 By considering the form of the **X** matrix for a model that includes only the intercept term, show that the Regression sum of squares for such a model is

$$\left(\sum_{i=1}^{n} Y_i\right)^2 \Big/ n$$

Notation: $\mathbf{J}_n$ denotes the $\mathrm{n} \times \mathrm{n}$ matrix of ones.

The ***X*** *matrix for such a model is* $\mathbf{1}_n$, *an n-vector of ones. In general, the Regression sum of squares is given by* $\mathbf{Y}'\mathbf{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{Y}'\left\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}\mathbf{Y}$ *so that for the model under consideration*

$$\begin{aligned}
\mathbf{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
&= \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'
\end{aligned}$$

Now $\mathbf{1}_n'\mathbf{1}_n = n$ *and* $\mathbf{1}_n\mathbf{1}_n' = \mathbf{J}_n$ *where* $\mathbf{J}_n$ *is the* $\mathrm{n} \times \mathrm{n}$ *matrix of ones. Hence,*

$$\begin{aligned}
\mathbf{P}_{\mathbf{X}} &= \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n' \\
&= \frac{1}{n}\mathbf{J}_n
\end{aligned}$$

and so the regression sum of squares in this case is

$$\begin{aligned}
\mathbf{Y'P_XY} &= \frac{1}{n}\mathbf{Y'J}_n\mathbf{Y} \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} Y_i \quad \sum_{i=1}^{n} Y_i \quad \cdots \quad \sum_{i=1}^{n} Y_i\right]\mathbf{Y} \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i Y_1 + \sum_{i=1}^{n} Y_i Y_2 + \ldots + \sum_{i=1}^{n} Y_i Y_n\right) \\
&= \frac{1}{n}\left(Y_1\sum_{i=1}^{n} Y_i + Y_2\sum_{i=1}^{n} Y_i + \ldots + Y_n\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)\sum_{i=1}^{n} Y_i \\
&= \left(\sum_{i=1}^{n} Y_i\right)^2 \Big/ n
\end{aligned}$$

II.13

a) What is the formula for the sample variance of the response variable?

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

b) Prove that $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathbf{y'y} - \left(\sum_{i=1}^{n} y_i\right)^2 \Big/ n$ where $\bar{y} = \sum_{i=1}^{n} y_i$. Verify this result for example II.1 for which $\mathbf{y'y} - \left(\sum_{i=1}^{n} y_i\right)^2 \Big/ n = 334.800$.

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}\left(y_i^2 - 2y_i\bar{y} + \bar{y}^2\right) \\
&= \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} 2y_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2 \\
&= \mathbf{y'y} - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2 \\
&= \mathbf{y'y} - 2\bar{y}n\bar{y} + n\bar{y}^2 \\
&= \mathbf{y'y} - n\bar{y}^2 \\
&= \mathbf{y'y} - \left(\sum_{i=1}^{n} y_i\right)^2 \Big/ n
\end{aligned}$$

y	*deviation*
50	*-0.8*
40	*-10.8*
52	*1.2*
47	*-3.8*
65	*14.2*
$\bar{y} = 254/5$	*SS= 334.8*
$= 50.8$	

c) Which formula do you think would be the best to use for computing the corrected total sum of squares in terms of numerical accuracy?

The formula involving the correction factor is the most accurate as it involves sum and squaring the original observations and a single difference. The formula involving deviations requires n differences and if there is little variation in the observations $\left(s^2 \approx 0\right)$ *the differences will be numerically unstable.*