STAT 571A — Advanced Statistical Regression Analysis

Chapter 5 NOTES

Matrix Approach to SLR Analysis

Introduction to Matrices

By appealing to matrix notation & matrix algebra, linear regression models become far more compact.

Recall that a matrix is an r×c rectangular array:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rc} \end{bmatrix}$$

(where order is important!)

Shorthand notation: $\mathbf{A} = [a_{ij}]$

Types of Matrices/Vectors

For $A_{r\times c} = [a_{ij}]$:

- if r = c, $A_{r\times r}$ is a **Square Matrix**
- if c = 1, $A_{r\times 1}$ is a **Column Vector**
- if r = 1, $A_{1\times c}$ is a **Row Vector**

The **transpose** of a matrix switches the rows and columns. Notation: $A' = [a_{ji}]$. E.g., if

$$A = \begin{bmatrix} 2 & 5 \\ 10 & 7 \\ 3 & 4 \end{bmatrix} \text{ then } A' = \begin{bmatrix} 2 & 10 & 3 \\ 5 & 7 & 4 \end{bmatrix}$$

More on Matrices

- Two matrices $\mathbf{A}$ and $\mathbf{B}$ are **equal** if
  (a) they have the same dimensions (r×c), and
  (b) $a_{ij} = b_{ij}$.
- If a matrix is made up of random variables Y_{ij}, then the expected value of $\mathbf{Y} = [Y_{ij}]$ is taken elementwise:

$$E\{\mathbf{Y}\} = [\ E\{Y_{ij}\}\ ]$$

§5.2: Matrix Form of SLR Model

We add/subtract matrices and vectors elementwise (if they have the same dimensions): $c\mathbf{A} + d\mathbf{B} = [\, ca_{ij} + db_{ij} \,]$

With this, a matrix formulation of the SLR model takes

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad E\{\mathbf{Y}\} = \begin{bmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

such that $\mathbf{Y} = E\{\mathbf{Y}\} + \boldsymbol{\epsilon}$.

§5.3: Matrix Multiplication

Multiplication for matrices is trickier: it's an operation first across rows, then down columns.

For $\mathbf{A}_{r \times c} = [a_{ij}]$ and $\mathbf{B}_{c \times d} = [b_{ij}]$, the $r \times d$ product $\mathbf{AB}$ is $(\mathbf{AB})_{r \times d} = [\ \sum_{k=1}^{c} a_{ik} b_{kj}\ ]$.

(and, $\mathbf{BA}$ doesn't exist unless $r=d$).

See various examples on pp. 182-184.

The $X'X$ Matrix

Let $X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ be the **design matrix** of the SLR. We will find the following matrix to be useful:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}.$$

Matrix Products

Also,

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

and

$$\mathbf{Y'Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum Y_i^2$$

Special Matrices

If $A_{r\times r}$ is square such that $a_{ij} = 0$ for <u>all</u> $i \neq j$, we call it a diagonal matrix.

Notation: $A_{r\times r} = \text{diag}\{a_{11}, a_{22}, \ldots, a_{rr}\}$

If $A_{r\times r}$ is square and diagonal such that $a_{ii} = 1$ for <u>all</u> i, we denote it as

$$I_r = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

and call it the Identity Matrix.

Special Matrices (cont'd)

- For the identity matrix $\mathbf{I}$,

$$\mathbf{A}_{r\times c}\mathbf{I}_c = \mathbf{I}_r\mathbf{A}_{r\times c} = \mathbf{A}_{r\times c},$$

  if the products exist ("conformable").

- Some other special matrices/vectors:

$$\mathbf{1}_{r\times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{0}_{r\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{J}_{r\times r} = \mathbf{1}\mathbf{1}' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

§5.5: Rank of a Matrix

Suppose a matrix $\mathbf{A}$ is partitioned into its constituent columns: $\mathbf{A} = [\mathbf{C}_1\ \mathbf{C}_2\ \cdots\ \mathbf{C}_c]$, where $\mathbf{C}_m$ is the *m*th col. vector.

Now, if there exist constants $k_m \neq 0$ such that the vector sum $\sum_{m=1}^{c} k_m \mathbf{C}_m = 0$, we say the columns of $\mathbf{A}$ are linearly dependent.

If not, the col's are linearly independent.

(See example on. p. 188.)

Rank (cont'd)

- <u>Def'n</u>: The **Rank** of a matrix is the (maximum) number of linearly independent columns the matrix possesses. Notation: rank($\mathbf{A}$)
- (Can do this with linearly independent rows, instead. The rank will not change.)
- Notice that rank($\mathbf{A}_{r\times c}$) $\leq$ min{r, c}

§5.6: Inverse of a Matrix

- The Inverse of a square r×r matrix $\mathbf{A}$ is another r×r matrix $\mathbf{A}^{-1}$ that satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_r$$

- $\mathbf{A}^{-1}$ can only exist if rank($\mathbf{A}_{r \times r}$) = r.
- Direct calculation of $\mathbf{A}^{-1}$ is usually tedious. See pp. 190-191.
- A special case: for $\mathbf{A} = \mathrm{diag}\{a_{11},\ldots,a_{rr}\}$, $\mathbf{A}^{-1} = \mathrm{diag}\{1/a_{11},\ldots,1/a_{rr}\}$.

Inverses (cont'd)

A useful, special case of an inverse matrix:

for

$$X'X = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix},$$

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix}.$$

Inverses (cont'd)

- Since matrix multiplication is order-sensitive, we **can't** be cavalier with the algebra.
- E.g., suppose $\mathbf{A}^{-1}$ exists and we have the relationship $\mathbf{AY} = \mathbf{C}$.
- Then (pre-multiply by $\mathbf{A}^{-1}$): $\mathbf{A}^{-1}\mathbf{AY} = \mathbf{A}^{-1}\mathbf{C}$. But since $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ we find $\mathbf{Y} = \mathbf{A}^{-1}\mathbf{C}$.
- But, $\mathbf{CA}^{-1}$ doesn't necess. make sense!

§5.7: Matrix Relations

(Assuming all products/inverses exist:)

$$A + B = B + A$$

$$(A + B) + C = A + (B + C)$$

$$(AB)C = A(BC)$$

$$C(A + B) = CA + CB$$

$$k(A + B) = kA + kB$$

$$(A')' = A$$

$$(A + B)' = A' + B'$$

$$(AB)' = B'A'$$

$$(ABC)' = C'B'A'$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$(A^{-1})^{-1} = A$$

$$(A')^{-1} = (A^{-1})'$$

§5.8: Covariance Matrix

We saw that for $\mathbf{Y}' = [\,Y_1 \;\; Y_2 \;\; \ldots \;\; Y_n\,]$,

the mean vector is $E\{\mathbf{Y}\} = [\ E\{Y_i\}\ ]$.

We also have

the variance-covariance matrix of $\mathbf{Y}$

as

$$\sigma^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \cdots & \sigma\{Y_1,Y_n\} \\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \cdots & \sigma\{Y_2,Y_n\} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma\{Y_n,Y_1\} & \sigma\{Y_n,Y_2\} & \cdots & \sigma^2\{Y_n\} \end{bmatrix}$$

Covariance Matrix (cont'd)

If $\mathbf{Y}' = [\ Y_1\ Y_2\ \ldots\ Y_n\ ]_{1\times n}$ is random but $\mathbf{A}_{n\times n}$ has only fixed (nonrandom) elements, then

- $E\{\mathbf{A}\} = \mathbf{A}$
- $E\{\mathbf{AY}\} = \mathbf{A}E\{\mathbf{Y}\}$
- and $\boldsymbol{\sigma}^2\{\mathbf{AY}\} = \mathbf{A}(\boldsymbol{\sigma}^2\{\mathbf{Y}\})\mathbf{A}'$

Multivariate Normal Dist'n

With this notation, we can extend the bivariate normal dist'n from §2.11 into a p-dimensional **multivariate normal dist'n**.

The p×1 mean vector is

$$E\{\mathbf{Y}\} = \boldsymbol{\mu} = [\mu_1 \ \dots \ \mu_p]'.$$

The p×p covariance matrix is

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

Notation: $\mathbf{Y} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

§5.9: Matrix Formulation of SLR

Putting all this together, the SLR model has a compact matrix formulation.

Recall $\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $E\{\mathbf{Y}\} = \begin{bmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{bmatrix}$, $\boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

and $\mathbf{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$. **Now, let** $\boldsymbol{\beta}_{2\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$.

SLR Formulation (cont'd)

With these, the SLR model in matrix terms is simply

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

where $\boldsymbol{\epsilon} \sim N_n(\mathbf{0},\sigma^2\mathbf{I})$.

(Compact, eh?)

Notice that since $E\{\boldsymbol{\epsilon}\} = \mathbf{0}$, $E\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$.

§5.10: LS Estimation

Moreover, the LS normal equations can be written simply as

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{Y},$$

for $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$.

The solution for $\mathbf{b}$ is clearly

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

whenever $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

§5.11: Hat Matrix

We can also write the fitted values in matrix form. Let $\hat{\mathrm{Y}} = [\hat{\mathrm{Y}}_1 \ldots \hat{\mathrm{Y}}_n]'$ so that $\hat{\mathrm{Y}} = \mathrm{Xb}$. But, this is

$$\hat{\mathrm{Y}} = \mathrm{Xb} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y} = \mathrm{HY}$$

for $\mathrm{H} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$.

We call H the **hat matrix**.

(This reminds us that $\hat{\mathrm{Y}}$ is a linear combination of the Y_i's.)

Note that $\mathrm{H}^2 = \mathrm{H}$ ("idempotent") and $\mathrm{H}' = \mathrm{H}$ (symmetric).

Residual Vector

From this, the residual vector is

$$\mathbf{e} = [e_i] = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

Here again, $(\mathbf{I} - \mathbf{H})^2 = \mathbf{I} - \mathbf{H}$ and $(\mathbf{I} - \mathbf{H})' = \mathbf{I} - \mathbf{H}$.

We can show $\boldsymbol{\sigma}^2\{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\sigma^2$.

Estimate this matrix via

$$\mathbf{s}^2\{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\times MSE.$$

§5.12: ANOVA

The ANOVA components may also be written in matrix form. Recall $\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$.

Then

$$\text{SSTO} = \sum(Y_i - \bar{Y})^2 = \mathbf{Y}'\mathbf{Y} - \tfrac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

$$= \mathbf{Y}'\mathbf{I}\mathbf{Y} - \mathbf{Y}'\left(\tfrac{1}{n}\mathbf{J}\right)\mathbf{Y} = (\mathbf{Y}'\mathbf{I} - \mathbf{Y}'\tfrac{1}{n}\mathbf{J})\mathbf{Y}$$

$$= \mathbf{Y}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{Y}$$

The SSE is a little trickier... →

ANOVA decomposition (cont'd)

We know SSE = $\sum(Y_i - \hat{Y}_i)^2 = \mathbf{Y'Y} - \mathbf{b'X'Y}$

But, recall that $\mathbf{b'} = [(\mathbf{X'X})^{-1}\mathbf{X'Y}]'$

$$= \mathbf{Y'(X')'[(X'X)^{-1}]'}$$

$$= \mathbf{Y'X[(X'X)']^{-1}}$$

$$= \mathbf{Y'X[X'X]^{-1}}$$

Now use this $\mathbf{b'}$ in the expression for SSE:

SSE $= \mathbf{Y'Y} - \mathbf{b'X'Y} = \mathbf{Y'Y} - (\mathbf{Y'X[X'X]^{-1}})\mathbf{X'Y}$

$$= \mathbf{Y'IY} - \mathbf{Y'X[X'X]^{-1}X'Y}$$

cont'd →

ANOVA decomposition (cont'd)

$$\text{SSE} = \cdots = \mathbf{Y'IY} - \mathbf{Y'X[X'X]^{-1}X'Y}$$
$$= (\mathbf{Y'I} - \mathbf{Y'X[X'X]^{-1}X'})\mathbf{Y}$$
$$= \mathbf{Y'}(\mathbf{I} - \mathbf{X[X'X]^{-1}X'})\mathbf{Y}$$

But notice that $\mathbf{X[X'X]^{-1}X'} = \mathbf{H}$ (the 'hat matrix'), so we find $\text{SSE} = \mathbf{Y'}(\mathbf{I} - \mathbf{H})\mathbf{Y}$

Lastly,

SSR = (by subtraction) = SSTO – SSE

$$= \cdots = \mathbf{b'X'Y} - \tfrac{1}{n}\mathbf{Y'JY} = \mathbf{Y'}(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\mathbf{Y}$$

Quadratic Forms

Notice that each SS expresses in the form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ for some symmetric matrix $\mathbf{A}$. This is a <u>number</u> and is given a special name: a quadratic form:

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_iY_j.$$

$\mathbf{A}$ is the "matrix of the quadratic form."

It is common for SS terms to appear as quadratic forms.

§5.13: Estimation

With $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ we can show (pp. 207-208) that $\sigma^2\{\mathbf{b}\} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$. Estimate this via

$$s^2\{\mathbf{b}\} = MSE(\mathbf{X}'\mathbf{X})^{-1}.$$

To estimate $E\{Y_h\} = \beta_0 + \beta_1 X_h$, let $\mathbf{X}_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix}$.

Then $E\{Y_h\} = \mathbf{X}_h'\boldsymbol{\beta}$. The point estimator becomes $\hat{Y}_h = \mathbf{X}_h'\mathbf{b}$, with

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h$$

and estimated std. error

$$s\{\hat{Y}_h\} = \sqrt{MSE \times \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h}\ .$$

Prediction

Or, for predicting a new observation, $\hat{Y}_{h(new)}$, at $X_{h(new)}$, let

$$\mathbf{X}_{h(new)} = \begin{bmatrix} 1 \\ X_{h(new)} \end{bmatrix}.$$

Then

$$\hat{Y}_{h(new)} = \mathbf{X}'_{h(new)}\mathbf{b},$$

with

$$s\{pred\} = \sqrt{MSE\left(1 + \mathbf{X}'_{h(new)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_{h(new)}\right)}.$$

Matrix Calculations in R

- In R, matrix operations are straightforward (once you know the syntax).
- An n×1 vector Y is entered as

```
> Y = c(y1, y2, ..., yn)
```

- An n×2 matrix X is entered as

```
> X = matrix( c(x11, x21, ..., xn1, x12,
                x22, ..., xn2),
              ncol=2, byrow=F )
```

- The n×1 vector of ones (for known n) is

```
> OneVec = rep(1,n)
```

Matrix Calculations in R (cont'd)

- $\mathbf{X}'$, the transpose of $\mathbf{X}$, is simply

```
> t(X)
```

- Multiply two (conformable) matrices or vectors with the operator `%*%`. So, $\mathbf{X}'\mathbf{Y}$ is

```
> t(X)%*%Y
```

  It's also: `crossprod(X,Y)`.

- Invert a (square!) matrix $\mathbf{A}$ with `solve(A)`. So, e.g., $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is

```
> solve(t(X)%*%X)%*%t(X)%*%Y
```