Matrix Approach to Linear Regression

Dr. Frank Wood

Random Vectors and Matrices

- Let's say we have a vector consisting of three random variables

$$\underset{3\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

The expectation of a random vector is defined

$$\underset{3\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}$$

Expectation of a Random Matrix

- The expectation of a random matrix is defined similarly

$$\underset{n\times p}{\mathbf{E}\{\mathbf{Y}\}} = [E\{Y_{ij}\}] \qquad i = 1, \ldots, n; j = 1, \ldots, p$$

Covariance Matrix of a Random Vector

- The collection of variances and covariances of and between the elements of a random vector can be collection into a matrix called the covariance matrix

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_3\} \\ \sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\} \end{bmatrix}$$

remember

$$\sigma\{Y_2, Y_1\} = \sigma\{Y_1, Y_2\}$$

so the covariance matrix is symmetric

Derivation of Covariance Matrix

- In vector terms the covariance matrix is defined by

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \mathbf{E}\{[\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}][\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}]'\}$$

because

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \mathbf{E}\left\{ \begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ Y_3 - E\{Y_3\} \end{bmatrix} [Y_1 - E\{Y_1\} \quad Y_2 - E\{Y_2\} \quad Y_3 - E\{Y_3\}] \right\}$$

verify first entry

Regression Example

- Take a regression example with n=3 with constant error terms $\sigma^2\{\epsilon_i\} = \sigma^2$ and are uncorrelated so that $\sigma^2\{\epsilon_i, \epsilon_j\} = 0$ for all $i \neq j$
- The covariance matrix for the random vector ϵ is

$$\underset{3\times 3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

which can be written as

$$\underset{3\times 3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \sigma^2 \underset{3\times 3}{\mathbf{I}}$$

Basic Results

- If A is a constant matrix and Y is a random matrix then

$$\mathbf{W} = \mathbf{A}\mathbf{Y}$$

is a random matrix

$$\mathbf{E}\{\mathbf{A}\} = \mathbf{A}$$

$$\mathbf{E}\{\mathbf{W}\} = \mathbf{E}\{\mathbf{A}\mathbf{Y}\} = \mathbf{A}\mathbf{E}\{\mathbf{Y}\}$$

$$\sigma^2\{\mathbf{W}\} = \sigma^2\{\mathbf{A}\mathbf{Y}\} = \mathbf{A}\sigma^2\{\mathbf{Y}\}\mathbf{A}'$$

Multivariate Normal Density

- Let Y be a vector of p observations

$$\underset{p\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}$$

- Let μ be a vector of p means for each of the p observations

$$\underset{p\times 1}{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Multivariate Normal Density

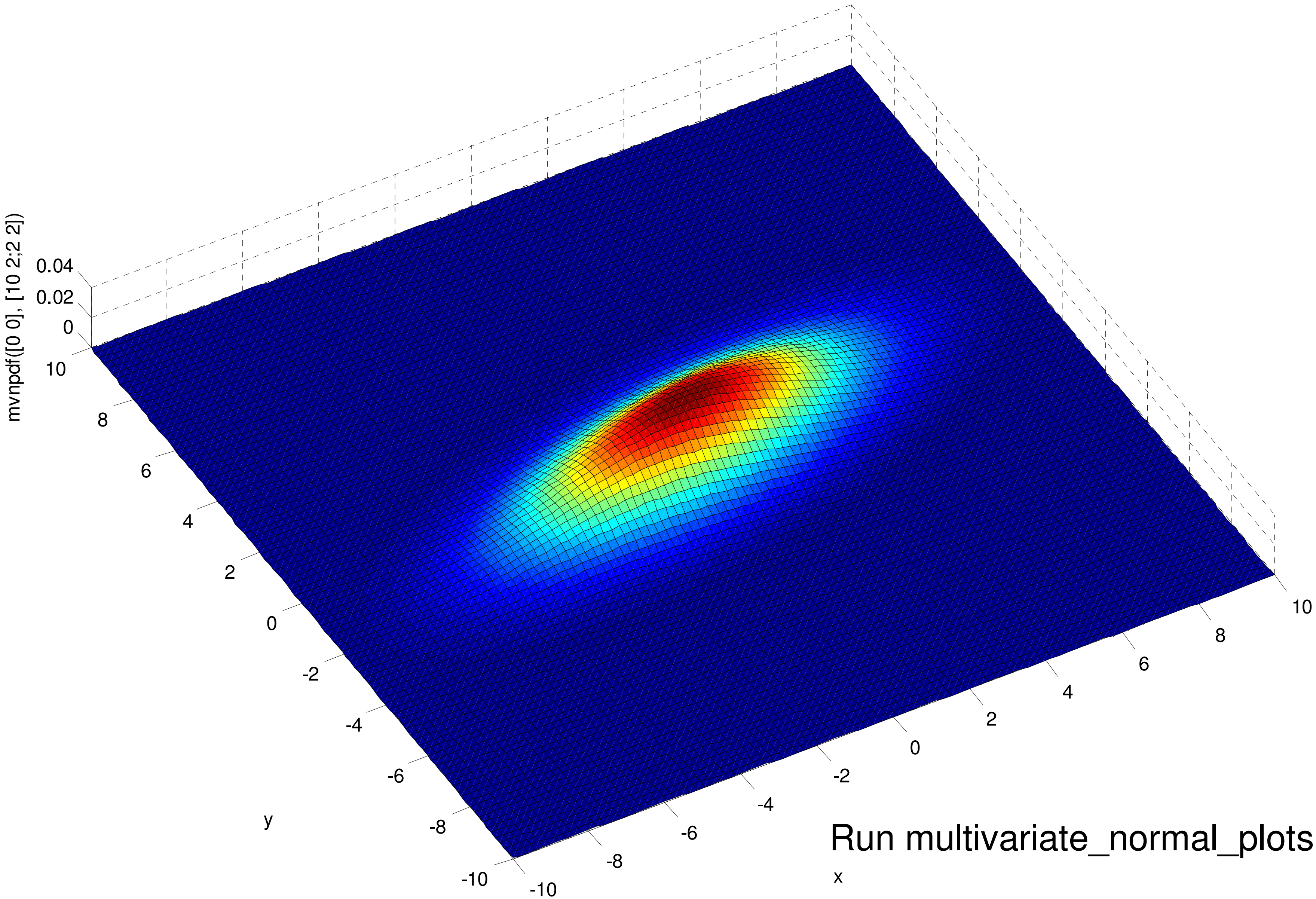
- Let § be the covariance matrix of Y

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

- Then the multivariate normal density is given by

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right]$$

Example 2d Multivariate Normal Distribution



Run multivariate_normal_plots.m

Matrix Simple Linear Regression

- Nothing new – only matrix formalism for previous results
- Remember the normal error regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \ldots, n$$

- This implies

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}$$

Regression Matrices

- If we identify the following matrices

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- We can write the linear regression equations in a compact form

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\,\underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

Regression Matrices

- Of course, in the normal regression model the expected value of each of the ϵ_i's is zero, we can write

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \underset{n\times 1}{\mathbf{X}\boldsymbol{\beta}}$$

- This is because

$$\underset{n\times 1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{n\times 1}{\mathbf{0}} \qquad \begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \vdots \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Error Covariance

- Because the error terms are independent and have constant variance σ^2

$$\underset{n \times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$\underset{n \times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \sigma^2 \underset{n \times n}{\mathbf{I}}$$

Matrix Normal Regression Model

- In matrix terms the normal regression model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{E}\{\boldsymbol{\varepsilon}\} = \mathbf{0}$$

and

$$\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}$$

Least Squares Estimation

- Starting from the normal equations you have derived

$$nb_0 + b_1 \sum X_i = \sum Y_i$$
$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i$$

we can see that these equations are equivalent to the following matrix operations

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}} \ \underset{2\times 1}{\mathbf{b}} = \underset{2\times 1}{\mathbf{X}'\mathbf{Y}}$$

with

$$\underset{2\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

demonstrate this on board

Estimation

- We can solve this equation

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}} = \underset{2\times 1}{\mathbf{X}'\mathbf{Y}}$$

(if the inverse of X'X exists) by the following

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

and since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$

we have $\underset{2\times 1}{\mathbf{b}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\ \underset{2\times 1}{\mathbf{X}'\mathbf{Y}}$

Least Squares Solution

- The matrix normal equations can be derived directly from the minimization of

$$Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

w.r.t. to β

Do this on board.

Fitted Values and Residuals

- Let the vector of the fitted values be

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

in matrix notation we then have

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}}$$

Hat Matrix – Puts hat on Y

- We can also directly express the fitted values in terms of only the X and Y matrices

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

and we can further define H, the “hat matrix”

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times n}{\mathbf{H}}\ \underset{n\times 1}{\mathbf{Y}} \qquad \underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

- The hat matrix plans an important role in diagnostics for regression analysis.

write H on board

Hat Matrix Properties

- The hat matrix is symmetric
- The hat matrix is idempotent, i.e.

$$\mathbf{HH} = \mathbf{H}$$

demonstrate on board

Residuals

- The residuals, like the fitted values of \hat{Y_i} can be expressed as linear combinations of the response variable observations Y_i

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\mathbf{Xb}}$$

$$\underset{n\times 1}{\mathbf{e}} = (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}}) \underset{n\times 1}{\mathbf{Y}}$$

Covariance of Residuals

- Starting with $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$

  we see that $\boldsymbol{\sigma}^2\{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\boldsymbol{\sigma}^2\{\mathbf{Y}\}(\mathbf{I} - \mathbf{H})'$

  but $\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}$

  which means that

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

$$= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})$$

and since I-H is idempotent (check) we have

$$\underset{n \times n}{\boldsymbol{\sigma}^2\{\mathbf{e}\}} = \sigma^2(\mathbf{I} - \mathbf{H})$$

we can plug in MSE for σ^2 as an estimate

ANOVA

- We can express the ANOVA results in matrix form as well, starting with

$$SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

where

$$\mathbf{Y}'\mathbf{Y} = \sum Y_i^2 \qquad \frac{(\sum Y_i)^2}{n} = \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$$

J is matrix of all ones, do 3x3 example

leaving

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

SSE

- Remember

$$SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

- We have

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

derive this on board

$$SSE = \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}}_{\mathbf{b}}$$

and this

$$= \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{I}\mathbf{X}'\mathbf{Y}$$

- Simplified

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$$

SSR

- It can be shown that
  - for instance, remember SSR = SSTO-SSE

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \qquad SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$$

write these on board

Tests and Inference

- The ANOVA tests and inferences we can perform are the same as before
- Only the algebraic method of getting the quantities changes
- Matrix notation is a writing short-cut, not a computational shortcut

Quadratic Forms

- The ANOVA sums of squares can be shown to be quadratic forms. An example of a quadratic form is given by

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2$$

- Note that this can be expressed in matrix notation as (where A is a symmetric matrix)

$$[Y_1 \quad Y_2] \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$$

do on board

Quadratic Forms

- In general, a quadratic form is defined by

$$\underset{1\times 1}{\mathbf{Y}'\mathbf{A}\mathbf{Y}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} Y_i Y_j \qquad \text{where } a_{ij} = a_{ji}$$

  A is the matrix of the quadratic form.
- The ANOVA sums SSTO, SSE, and SSR are all quadratic forms.

ANOVA quadratic forms

- Consider the following rexpression of b'X'

$$\mathbf{b}'\mathbf{X}' = (\mathbf{X}\mathbf{b})' = \hat{\mathbf{Y}}' \qquad \mathbf{b}'\mathbf{X}' = (\mathbf{H}\mathbf{Y})'$$

$$\mathbf{b}'\mathbf{X}' = \mathbf{Y}'\mathbf{H}$$

- With this it is easy to see that

$$SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$$

$$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}$$

Inference

- We can derive the sampling variance of the β vector estimator by remembering that

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

where A is a constant matrix

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad \mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

which yields

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{A}'$$

Variance of b

- Since $(\mathbf{X}'\mathbf{X})^{-1}$ is symmetric we can write

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

and thus

$$\begin{aligned}\boldsymbol{\sigma}^2\{\mathbf{b}\} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{I} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Variance of b

- Of course this assumes that we know σ^2. If we don't, we, as usual, replace it with the MSE.

$$\underset{2\times 2}{\sigma^2\{\mathbf{b}\}} = \begin{bmatrix} \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum(X_i - \bar{X})^2} & \frac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} \\ \frac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} & \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \end{bmatrix}$$

$$\underset{2\times 2}{\mathbf{s}^2\{\mathbf{b}\}} = MSE(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{MSE}{n} + \frac{\bar{X}^2 MSE}{\sum(X_i - \bar{X})^2} & \frac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} \\ \frac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} & \frac{MSE}{\sum(X_i - \bar{X})^2} \end{bmatrix}$$

Mean Response

- To estimate the mean response we can create the following matrix

$$\underset{2\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \qquad \text{or} \qquad \underset{1\times 2}{\mathbf{X}'_h} = [1 \quad X_h]$$

- The fit (or prediction) is then

$$\hat{Y}_h = \mathbf{X}'_h \mathbf{b}$$

since

$$\mathbf{X}'_h \mathbf{b} = [1 \quad X_h] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h] = [\hat{Y}_h] = \hat{Y}_h$$

Variance of Mean Response

- Is given by

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

  and is arrived at in the same way as for the variance of \beta
- Similarly the estimated variance in matrix notation is given by

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

Wrap-Up

- Expectation and variance of random vector and matrices
- Simple linear regression in matrix form

- Next: multiple regression