36-401 Class Notes and Exercise for 9/14/00

Howard Seltman

Simple linear regression in matrix form (Appendix 3)

Column and Row Vectors ("Prime" or superscript T means transpose)

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = [Y_1\ Y_2\ \cdots\ Y_n]'$$

The covariates are in the "design matrix". For simple regression this has n rows and 2 columns so it is of "order" or "dimension" $n \times 2$.

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

The coefficients are in a $K \times 1$ column vector:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = [\beta_0\ \beta_1]'$$

Errors are in an $n \times 1$ column matrix:

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = [\epsilon_1\ \epsilon_2\ \cdots\ \epsilon_n]'$$

Model:

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\epsilon}_{n\times 1} \quad \boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2\mathbf{I})$$

Matrix addition requires same dimensions. Matrix multiplication is non-commutative and requires an inner dimension match. E.g. $\mathbf{T}_{axb}\mathbf{U}_{cxd}$ is possible only if $b = c$, and the product has order $a \times d$.

$$\mathbf{X}\boldsymbol{\beta} = [\qquad\qquad\qquad \cdots \qquad\qquad]'$$

No change if pre- or post-multiplied by identity matrix. Inverse of a square matrix is defined such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. Instead of division, multiply by inverse.

Splus matrix example

```
> X_matrix(c(rep(1,6),4,2,7,1,8,5), 6,2)
> Y_c(16,11,30,7,33,22)
> X
     [,1] [,2]
[1,]    1    4
[2,]    1    2
[3,]    1    7
[4,]    1    1
[5,]    1    8
[6,]    1    5
> XtX_t(X) %*% X
> XtX
     [,1] [,2]
[1,]    6   27
[2,]   27  159

> XtXi_solve(XtX)
> XtXi
           [,1]        [,2]
[1,]  0.7066667 -0.12000000
[2,] -0.1200000  0.02666667
> XtXi %*% XtX
             [,1]         [,2]
[1,] 1.000000e+00 7.105427e-15
[2,] 7.771561e-16 1.000000e+00
> t(X) %*% Y
     [,1]
[1,]  119
[2,]  677
> lm(Y~X-1)
```

Parameter Estimation

Choose the parameter estimate $\mathbf{B} = [b_0\ b_1]'$ to minimize RSS.

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{e} \Leftrightarrow \mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e} \qquad \mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{B} \Leftrightarrow \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

$$\text{RSS} = \mathbf{e}'\mathbf{e} = e_1^2 + e_2^2 + \cdots + e_n^2$$

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

$$\Rightarrow -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{0} \Leftrightarrow \mathbf{B} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$$

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

Variance covariance matrix of the parameter estimates is

$$\mathbf{S} = s_e^2(\mathbf{X}'\mathbf{X})^{-1}$$

Breakout Session for 9/14/00

1) Consider the following matrices and vectors and their dimensions: $\mathbf{Y}_{14\times1}$, $\mathbf{X}_{14\times4}$, $\boldsymbol{\beta}$, $\boldsymbol{\epsilon}$. If $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, what are the missing dimensions?

2) If the 5^{th} row of $\mathbf{X}$ is (1, 4, -3, 2.5), what is the 5^{th} row of $\mathbf{X}\boldsymbol{\beta}$?

3) Using the Splus code above, what is $\mathbf{B}$? Why does the lm() call use "$X - 1$"?

4) Using the Splus code above (and $s_e = 1.11$) calculate
a) the variance of b_0:
b) the standard error of b_1:
c) the covariance of b_0 and b_1:
d) the correlation of b_0 and b_1: