Stats 401 Lab 3

401 GSI team

1/18/2018 and 1/19/2018

Announcements

- Homework 2 is due today
- Please remember to include "Sources" and "Please explain"
- First quiz is coming up! (Feb. 1st or 2nd)
- Homework 1 and 2 solutions will be posted Friday afternoon.

Quick Review: Matrix Properties

- Addition

$$\mathbb{A} + \mathbb{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

- Scalar multiplication

$$s\mathbb{A} = \begin{bmatrix} sa_{11} & sa_{12} \\ sa_{21} & sa_{22} \end{bmatrix}$$

- Transpose

$$\mathbb{A}^\top = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

- Matrix multiplication

$$\mathbb{A}\mathbb{B} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Quick Review: Matrix Properties (cont.)

- Inverse

$$\mathbb{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Quick Review: How to Input Matrices into R

- Take the following matrix:

$$\mathbb{A} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 0 & 2 \end{bmatrix}$$

```r
# R takes a column vector and
# splits it into 3 different rows
matrix_by_col <- matrix(c(1,1,3,1,2,0,3,1,2),nrow=3)
matrix_by_row <- matrix(c(1,1,3,1,2,1,3,0,2),
                        byrow = TRUE, nrow=3)
```

Quick Review: How to Input Matrices into R

- Don't forget to check that your output is correct!

```
matrix_by_col
```

```
##      [,1] [,2] [,3]
## [1,]    1    1    3
## [2,]    1    2    1
## [3,]    3    0    2
```

```
matrix_by_row
```

```
##      [,1] [,2] [,3]
## [1,]    1    1    3
## [2,]    1    2    1
## [3,]    3    0    2
```

Matrices in Action

- Solving a system of linear equations
- Recall the US wages dataset that we saw in Lab 01

```
##          wage educ exper race smsa ne mw so we pt
## 6085  771.60   18    18    0    1  1  0  0  0  0
## 23701 617.28   15    20    0    1  0  0  0  1  0
## 16208 957.83   16     9    0    1  0  0  1  0  0
## 2720  617.28   12    24    0    1  1  0  0  0  0
## 9723  902.18   14    12    0    1  0  1  0  0  0
## 22239 299.15   12    33    0    1  0  0  0  1  0
```

Matrices in Action (cont.)

- Write the sample version of the linear model for wages with all other variables as explanatory variables using vector notation.

Step 1) Notation:

Let $x_{i1}, x_{i2}, \dots, x_{ip}$ and y_i be the values of predictor variable j and the wage of worker i, respectively.

Step 2) Recognize the hidden matrix:

Linear model for each individual is

$$y_1 = b_0 + b_1 x_{11} + b_2 x_{12} + \cdots + b_p x_{1p} + e_1$$

$$y_2 = b_0 + b_1 x_{21} + b_2 x_{22} + \cdots + b_p x_{2p} + e_2$$

$$\vdots$$

$$y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + \cdots + b_p x_{np} + e_n$$

- This looks like a system of linear equations that we can put into matrix form!

Matrices in Action (cont.)

Step 3) Define the matrices and vectors:

Let

$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$

Let $\mathbf{y} = (y_1, y_2, ..., y_n)$, $\mathbf{b} = (b_1, b_2)$, and $\mathbf{e} = (e_1, e_2, ..., e_n)$ be the vector of wages, predictor variables, and error terms. Then

$$\mathbf{y} = \mathbb{X}\mathbf{b} + \mathbf{e}$$

Solving for the least squares estimate

- Recall the least squares estimate for the predictor variable coefficients is
$$\mathbf{b} = (\mathbb{X}^{\mathbb{T}}\mathbb{X})^{-1}\mathbb{X}\mathbf{y}$$
- Let's solve for $\mathbf{b}$ together.

Solving for **b**

- Construct $\mathbb{X}$ matrix corresponding to the linear equation from before.

```
# Note that I've left out one of the
# regional dummy varibles
# Question for students: Why did I need to do this?
attach(uswages)
X <- cbind(intercept=rep(1, length(educ)),
           educ, exper, race,
           smsa, ne, mw,
           so, pt)

head(X, n=3)
```

```
##      intercept educ exper race smsa ne mw so pt
## [1,]         1   18    18    0    1  1  0  0  0
## [2,]         1   15    20    0    1  0  0  0  0
## [3,]         1   16     9    0    1  0  0  1  0
```

Solving for **b**

```r
solve(t(X) %*% X) %*% t(X) %*% uswages$wage
```

```
##                  [,1]
## intercept -203.918425
## educ        48.803359
## exper        9.135332
## race      -119.158469
## smsa       115.678257
## ne         -53.926540
## mw         -60.199034
## so         -50.433257
## pt        -336.215572
```

Checking our result **b**

```r
# Use the linear model function in R for
# including all the variables
wage_lm <- lm(wage ~ ., data = uswages)
coef(wage_lm)
```

```
## (Intercept)        educ       exper        race
## -203.918425   48.803359    9.135332 -119.158469
##        smsa          ne          mw          so
##  115.678257  -53.926540  -60.199034  -50.433257
##          we          pt
##          NA -336.215572
```

In Lab Activity

- Using the library(faraway) and data("infmort"):

1. Construct the linear equation using vector and matrix notation.
2. Estimate the least squares estimate of **b** using the design matrix $\mathbb{X}$.
3. Check your estimate by using the lm function in R.

Lab ticket

- Write your least squares estimate of the fitted values for the infant mortality data.