Lecture 1

Brennan Becerra

2024-01-09

Items of Interest:

- Basics of R
- Basics of Test of Hypothesis
- Textbook Topics

When you open R :

```
> 2+3 # Enter
[1] 5 # Output
> 2 * 3 ; 2/3 ; 2^3 ; log(3) ...
```

Codeblock 1 (language-r)

Assignment of Variables:

```
>a <- 2 # Assignment of a to 2 (use this one)
>a =2 # Assignment of a to 2
>a +3
[1] 5
>a^3
>log(a)
```

Codeblock 2 (language-r)

```
>a <- c(1,2,3,4,5)
>a
[1] 1 2 3 4 5
>mean(a) #Average values
>sum(a) #Add up all data points
>sd(a) #Standard Deviation
>var(a) #Variance
```

Codeblock 3 (language-r)

Scan function:

```
>a <- scan("datafile name")
>a <- scan() #Will also work, just specifies nothing
1: 1
2: 2
3: 3
4:
Read 3 items
>a
[1] 1 2 3
>sum(a)
[1] 6
>a+3
[1] 4 5 6 #adds 3 to each data value
```

Codeblock 4 (language-r)

(Note) the scan() function read an external text or ASCII-formatted data file

Matrix and matrix computation

Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Ex.

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}_{2\times 3(\text{Dimension of A})}$$

```
>matrix(data, nrow=n, ncol=m, byrow= #true or false) # byrow = T
```

Codeblock 5 (language-r)

```
>seq(1:6)
[1] 1 2 3 4 5 6
>DF = matrix()
```

Codeblock 6 (language-r)

---

byrow=F

$$DD = DF = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

byrow=T

$$DF = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

```
>DD[row, column] # this gives you the value at that point
>DD[2, ]
[1] 2 4 6
>DD[ , 3]
[1] 5 6
```

Codeblock 7 (language-r)

```
>sum(DF)
>sum(DF[2, ]) # only interested in summing the values of row 2
```

Codeblock 8 (language-r)

$$t(DF) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Say

$$A = \begin{pmatrix} 1 & 10 \\ 2 & 5 \end{pmatrix}$$
$$A^T = \begin{pmatrix} 1 & 2 \\ 10 & 5 \end{pmatrix}$$

Observe the following:

```
>Y= matrix(x(1,2,3,4,5,6),nrow=3,ncol=2)
```

Codeblock 9 (language-r)

We get

$$Y = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \qquad <- \qquad \text{cbind(Y,Z)}$$
$$X = \begin{pmatrix} 7 & 9 & 11 \\ 8 & 10 & 12 \end{pmatrix}$$
$$Z = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$$

```
>betahat=solve(t(X)%*%X)%*%t(X)%*%y # Calculating the following
```

Codeblock 10 (language-r)

$$(X'X)^{-1}X'y$$

```
>A'%*%A
```

Codeblock 11 (language-r)

$$\begin{pmatrix} 1 & 2 \\ 10 & 5 \end{pmatrix} \begin{pmatrix} 1 & 10 \\ 2 & 5 \end{pmatrix}$$

Learning

- Supervised Learning: $\underbrace{Y}_{\text{Output}} = \underbrace{f(x)}_{\text{Input}} + \text{Error}.$
  - Ch 1 ~ Ch 9.
- Unsupervised Learning: We do not know y.
  - Neural Network $\rightarrow$ AI

$$\underbrace{Y}_{\text{Response/Dependent}} = \underbrace{f(x)}_{\text{Independent}} + \text{ERROR}$$

- Linear; Linear Regression Analysis (Ch.3)
- Non-linear; Non-parametric Analysis

Ex.

$$Y = \beta_0 + \beta_1 x + \text{ERROR}$$
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \text{ERROR}$$