Lab 1 - R Intro

Spring 2018 - Multivariate Data Analysis

R is

- A language and environment for statistical computing and graphics ("S" language)
- Free from www.r-project.org
- Fine graphics
- Easy and efficient handling of data
- Rich modern statistical routines
- Keep in mind !
- R is case sensitive !
- x and X are different !
- Rstudio : download from www.rstudio.com

Assignment

```r
X<-sqrt(2*3)
X
```

```
## [1] 2.44949
```

```r
class(X)
```

```
## [1] "numeric"
```

manual entry : vector

```
x<-c(-1,2,5)
x
```

```
## [1] -1  2  5
```

```
x<-c("A", "B", "C")
x
```

```
## [1] "A" "B" "C"
```

equispaced sequences of numbers

```
x<-3:10
x
```

```
## [1]  3  4  5  6  7  8  9 10
```

```
seq(3,10)
```

```
## [1]  3  4  5  6  7  8  9 10
```

```
seq(3,5,by=1/2)
```

```
## [1] 3.0 3.5 4.0 4.5 5.0
```

```
seq(3,10,length=8)
```

```
## [1]  3  4  5  6  7  8  9 10
```

logical values T and F

```r
x<-c(T,T,T,F,F,T,F)
x
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE
```

```r
x<-c("T", "F")
x
```

```
## [1] "T" "F"
```

```r
x<-((1:5)>3)
x
```

```
## [1] FALSE FALSE FALSE  TRUE  TRUE
```

matrices

```
x<-matrix(c(1,2,3,4,5,6),ncol=2)
x
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

```
x<-matrix(c(1,2,3,4,5,6),ncol=2,byrow=T)
x
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    3    4
## [3,]    5    6
```

data frames

```
x1<-rnorm(100)
x2<-rt(df=3,100)
x3<-sample(1:10,100,replace=T)
x.dat<-data.frame(X1=x1,X2=x2,X3=x3)
head(x.dat)
```

```
##           X1          X2 X3
## 1  1.3201508  0.05827621  3
## 2 -0.4735321 -0.68844877  2
## 3  0.8408570 -0.03072463 10
## 4 -0.5586731 -0.28889224  9
## 5  0.8963130 -1.74912818  7
## 6  1.3398131  2.61117629  7
```

list (1)

```
x<-list(name="Cox",wife="Mary",husband="Fred",
        no.child=3, child.ages=c(4,8,9))
x
```

```
## $name
## [1] "Cox"
##
## $wife
## [1] "Mary"
##
## $husband
## [1] "Fred"
##
## $no.child
## [1] 3
##
## $child.ages
## [1] 4 8 9
```

list (2)

```r
x<-list(name=c("Cox", "Wang"),
          wife=c("Mary", "Pearl"),
          husband=c("Fred", "Val"), no.child=c(3,2),
          child.ages=c(c(4,8,9),c(1,5)))
x$name
```

```
## [1] "Cox"  "Wang"
```

```r
x$no.child
```

```
## [1] 3 2
```

sub-setting/selecting (1)

```r
x<-11:20
x[c(1,5,10)]
```

```
## [1] 11 15 20
```

```r
x[-c(1:5)]
```

```
## [1] 16 17 18 19 20
```

```r
x[x>15]
```

```
## [1] 16 17 18 19 20
```

```r
id<-which(x>15)
id
```

```
## [1]  6  7  8  9 10
```

```r
x[id]
```

```
## [1] 16 17 18 19 20
```

sub-setting/selecting (2)

```r
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
tail(iris[iris$Species=="setosa" ,])
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 45          5.1         3.8          1.9         0.4  setosa
## 46          4.8         3.0          1.4         0.3  setosa
## 47          5.1         3.8          1.6         0.2  setosa
## 48          4.6         3.2          1.4         0.2  setosa
## 49          5.3         3.7          1.5         0.2  setosa
## 50          5.0         3.3          1.4         0.2  setosa
```

sub-setting/selecting (3)

```
iris[1:10,1]
```

```
##  [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9
```

```
iris$Sepal.Length[1:10]
```

```
##  [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9
```

```
iris[1:10, "Sepal.Length"]
```

```
##  [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9
```

simple statistics (1)

```
x<-rnorm(1000)
length(x)
```

```
## [1] 1000
```

```
sum(x)
```

```
## [1] 41.52639
```

```
mean(x)
```

```
## [1] 0.04152639
```

```
var(x)
```

```
## [1] 1.053724
```

simple statistics (2)

```r
sqrt(var(x));sd(x)
```

```
## [1] 1.026511

## [1] 1.026511
```

```r
min(x) ;max(x)
```

```
## [1] -3.235745

## [1] 3.362943
```

```r
range(x) ; median(x)
```

```
## [1] -3.235745  3.362943

## [1] 0.02442099
```

```r
summary(x)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.23575 -0.63132  0.02442  0.04153  0.73801  3.36294
```

simple function (3)

```
x<-runif(5,-10,10)
round(x)
```

```
## [1]  -9  -2  -7  -2 -10
```

```
abs(x)
```

```
## [1] 8.702106 2.099357 7.455881 1.677736 9.559698
```

```
x^2
```

```
## [1] 75.726655  4.407301 55.590155  2.814797 91.387824
```

```
log(x^2)
```

```
## [1] 4.327130 1.483263 4.018006 1.034890 4.515112
```

```
x>0
```

```
## [1] FALSE FALSE FALSE FALSE FALSE
```

simple function (4)

```r
tapply(iris[, "Sepal.Length"],iris[, "Species"],
          function(x){mean(x,na.rm=T)})
```

```
##     setosa versicolor  virginica
##      5.006      5.936      6.588
```

```r
mean.A<-function(x)
{  mean(x,na.rm=T) }
tapply(iris[, "Sepal.Length"],iris[, "Species"], mean.A)
```

```
##     setosa versicolor  virginica
##      5.006      5.936      6.588
```

ranking, sorting

```
x<-sample(11:20)
x
```

```
##  [1] 15 13 16 20 17 14 12 11 18 19
```

```
rank(x)
```

```
##  [1]  5  3  6 10  7  4  2  1  8  9
```

```
order(x)
```

```
##  [1]  8  7  2  6  1  3  5  9 10  4
```

```
sort(x)
```

```
##  [1] 11 12 13 14 15 16 17 18 19 20
```

```
x[order(x)]
```

```
##  [1] 11 12 13 14 15 16 17 18 19 20
```

Matrix Algebra (1)

```
A<-matrix(1:10,ncol=5)
B<-matrix(11:20,ncol=5)
A+B
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   12   16   20   24   28
## [2,]   14   18   22   26   30
```

```
A-B
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  -10  -10  -10  -10  -10
## [2,]  -10  -10  -10  -10  -10
```

Matrix Algebra (2)

```r
A*B
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   11   39   75  119  171
## [2,]   24   56   96  144  200
```

```r
#A%*%B
A%*%t(B)
```

```
##      [,1] [,2]
## [1,]  415  440
## [2,]  490  520
```

Matrix Algebra (3)

```
A<-matrix(c(1,0,1,2,3,5,1,4,6),ncol=3)
x<-1:3
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    1
## [2,]    0    3    4
## [3,]    1    5    6
```

```
x
```

```
## [1] 1 2 3
```

```
x%*%A%*%x
```

```
##      [,1]
## [1,]  131
```

```
x%*%A
```

```
##      [,1] [,2] [,3]
## [1,]    4   23   27
```

Linear system

```
solve(A,x)
```

```
## [1] -0.3333333  0.6666667  0.0000000
```

```
solve(A)
```

```
##            [,1]      [,2]      [,3]
## [1,] -0.6666667 -2.333333  1.666667
## [2,]  1.3333333  1.666667 -1.333333
## [3,] -1.0000000 -1.000000  1.000000
```

```
t(x)%*%solve(A)%*%x # quadratic form
```

```
##      [,1]
## [1,]    1
```

```
t(x)%*%solve(A,x) # quadratic form
```

```
##      [,1]
## [1,]    1
```

diagonal matrix

```r
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    1
## [2,]    0    3    4
## [3,]    1    5    6
```

```r
diag(A)
```

```
## [1] 1 3 6
```

```r
x
```

```
## [1] 1 2 3
```

```r
diag(x)
```

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

matrix composition

```r
x1<-c(1,2,3)
x2<-c(4,5,6)
cbind(x1,x2)
```

```
##      x1 x2
## [1,]  1  4
## [2,]  2  5
## [3,]  3  6
```

```r
rbind(x1,x2)
```

```
##    [,1] [,2] [,3]
## x1    1    2    3
## x2    4    5    6
```

Matrix calculation (1)

```
A<-matrix(c(1,-5,-5,1),nrow=2)
A
```

```
##      [,1] [,2]
## [1,]    1   -5
## [2,]   -5    1
```

```
eigen(A)
```

```
## eigen() decomposition
## $values
## [1]  6 -4
##
## $vectors
##            [,1]       [,2]
## [1,] -0.7071068 -0.7071068
## [2,]  0.7071068 -0.7071068
```

Matrix calculation (2)

```
P<-eigen(A)$vectors
E<-diag(eigen(A)$values)
P
```

```
##            [,1]       [,2]
## [1,] -0.7071068 -0.7071068
## [2,]  0.7071068 -0.7071068
```

```
E
```

```
##      [,1] [,2]
## [1,]    6    0
## [2,]    0   -4
```

```
P%*%E%*%t(E)
```

```
##           [,1]      [,2]
## [1,] -25.45584 -11.31371
## [2,]  25.45584 -11.31371
```

Matrix calculation (3)

```r
P%*%t(P)
```

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

```r
t(P)%*%P
```

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

Matrix calculation (4)

```
solve(A)
```

```
##              [,1]        [,2]
## [1,] -0.04166667 -0.20833333
## [2,] -0.20833333 -0.04166667
```

```
E.inv<-diag(1/eigen(A)$values)
P%*%E.inv%*%t(P)
```

```
##              [,1]        [,2]
## [1,] -0.04166667 -0.20833333
## [2,] -0.20833333 -0.04166667
```

Matrix calculation (5)

```r
A<-matrix(c(13,-4,2,-4,13,-2,2,-2,10),ncol=3)
eigen(A)
```

```
## eigen() decomposition
## $values
## [1] 18  9  9
##
## $vectors
##            [,1]       [,2]      [,3]
## [1,]  0.6666667 -0.7453560 0.0000000
## [2,] -0.6666667 -0.5962848 0.4472136
## [3,]  0.3333333  0.2981424 0.8944272
```

Standardization (1)

```r
sample1<-iris[,1:4]
X.mean<-colMeans(sample1)
S<-var(sample1)
P<-eigen(S)$vectors
E.inv.5<-diag(1/sqrt(eigen(S)$values))
one<-matrix(rep(1,nrow(sample1)),ncol=1)
head(one%*%X.mean)
```

```
##          [,1]     [,2]  [,3]     [,4]
## [1,] 5.843333 3.057333 3.758 1.199333
## [2,] 5.843333 3.057333 3.758 1.199333
## [3,] 5.843333 3.057333 3.758 1.199333
## [4,] 5.843333 3.057333 3.758 1.199333
## [5,] 5.843333 3.057333 3.758 1.199333
## [6,] 5.843333 3.057333 3.758 1.199333
```

Standardization (2)

```
Z1<-apply(sample1,2,function(x) (x-mean(x))/sd(x))
apply(Z1,2,mean)
```

```
##  Sepal.Length   Sepal.Width  Petal.Length   Petal.Width
## -4.484318e-16  2.034094e-16 -2.895326e-17 -2.989362e-17
```

```
var(Z1)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
## Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
## Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
## Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

```
cor(sample1)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
## Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
## Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
## Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

Standardization (3)

```
S.inv.5<-P%*%E.inv.5%*%t(P)
Z2<-as.matrix(sample1-one%*%X.mean)%*%S.inv.5
apply(Z2,2,mean)
```

```
## [1] -1.101363e-15  6.016108e-16  5.421314e-16 -2.859174e-16
```

```
var(Z2)
```

```
##               [,1]          [,2]          [,3]          [,4]
## [1,]  1.000000e+00  2.782462e-15  1.192112e-15 -7.337280e-16
## [2,]  2.782462e-15  1.000000e+00 -1.673888e-16 -3.318026e-15
## [3,]  1.192112e-15 -1.673888e-16  1.000000e+00 -8.502763e-16
## [4,] -7.337280e-16 -3.318026e-15 -8.502763e-16  1.000000e+00
```