week1quiz

Introduction

R Refresher

QUESTION 1.1.1: R REFRESHER (1 point possible)

If you haven't done so already, install the library UsingR

```
install.packages("UsingR")
```

Then once you load it you have access to Galton's father and son heights:

```
## Loading required package: MASS
## Loading required package: HistData
## Loading required package: Hmisc
## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
##
##
## Attaching package: 'UsingR'
##
## The following object is masked from 'package:ggplot2':
##
##     movies
##
## The following object is masked from 'package:survival':
##
##     cancer
```

What is the average height of the sons (don't round off)?

```r
library(UsingR)
data(father.son)
mean(father.son$sheight)
```

```
## [1] 68.68407
```

QUESTION 1.1.2 R REFRESHER

One of the defining features of regression is that we stratify one variable based on others. In Statistics we use the verb "condition". For example, the linear model for son and father heights answers the question how tall do I expect a son to be if I condition on his father being x inches. The regression line answers this question for any x.

Using the father.son dataset described above, we want to know the expected height of sons if we condition on the father being 71 inches. Create a list of son height's for sons that have fathers with height of 71 inches (round to the nearest inch).

```r
x=father.son$fheight
y=father.son$sheight

l<-split(y,round(x))
```

What is the mean of the son heights for fathers that have a height of 71 inches (don't round off your answer)? (Hint: use the function round() on the fathers' heights)

```r
mean(y[round(x)==71])
```

```
## [1] 70.54082
```

QUESTION 1.1.3

We say a statistical model is a linear model when we can write it as a linear combination of parameters and known covariates plus random error terms. In the choices below, Y represents our observations, time t is our only covariate, unknown parameters are represented with letters a,b,c,d and measurment error is represented by the letter e. Note that if t is known, then any transformation of t is also known. So, for example, both Y=a+bt +e and Y=a+b f(t) +e are linear models. Which of the following can't be written as a linear model?

- Y = a + bt + e
- Y = a + b cos(t) + e
- Y = a + b^t + e
- Y = a + b t + c t^2 + d t^3 + e

Answer is Y = a + b^t + e

EXPLANATION In every other case we can write the model as linear combination of parameters and known covariates. b^t is not a linear combination of b and t.

QUESTION 1.1.4 (1/1 point)

Supposed you model the relationship between weight and height across individuals with a linear model. You assume that the height of individuals for a fixed weight x follows a liner model Y = a + b x + e. Which of the following do you feel best describes what e represents?

- Measurement error: scales are not perfect
- Within individual random fluctuations: you don't weigh the same in the morning as the afternoon
- Round off error introduced by the computer we use to analyze the data
- Between individual variability: people of the same height vary in their weight

Answer is Between individual variability: people of the same height vary in their weight

EXPLANATION Remember the model is across individuals and we fix x. People of the same height can greatly in other aspects of their physiology: for example different bone density or differing amounts of muscle and fat.

Matrix Notations

QUESTION 1.2.1

In R we have vectors and matrices. You can create your own vectors with the function c.

`c(1,5,3,4)`

They are also the output of many functions such as

```
rnorm(10)
```

```
##  [1]  1.0013195 -0.1651693 -0.7890390 -0.6383341 -1.5839791 -0.6363041
##  [7] -1.8208719  0.8967536 -0.6178332 -1.4296602
```

You can turn vectors into matrices using functions such as rbind, cbind or matrix.

Create the matrix from the vector 1:1000 like this:

```
X = matrix(1:1000,100,10)
```

What is the entry in row 25, column 3 ?

```
X[25,3]
```

```
## [1] 225
```

QUESTION 1.2.2

Using the function cbind, create a 10 x 5 matrix with first column

```
x=1:10
```

Then columns 2*x, 3*x, 4*x and 5*x in columns 2 through 5.

```
m<-cbind(x,2*x,3*x,4*x,5*x)
```

What is the sum of the elements of the 7th row?

```
sum(m[7,])
```

```
## [1] 105
```

QUESTION 1.2.3

Which of the following creates a matrix with multiples of 3 in the third column?

- matrix(1:60,20,3)
- matrix(1:60,20,3,byrow=TRUE)
- x=11:20;rbind(x,2*x,3*x)
- x=1:40;matrix(3*x,20,2)

Answer is

```
matrix(1:60,20,3,byrow=TRUE)
```

```
##       [,1] [,2] [,3]
##  [1,]    1    2    3
##  [2,]    4    5    6
##  [3,]    7    8    9
##  [4,]   10   11   12
##  [5,]   13   14   15
##  [6,]   16   17   18
##  [7,]   19   20   21
##  [8,]   22   23   24
##  [9,]   25   26   27
## [10,]   28   29   30
## [11,]   31   32   33
## [12,]   34   35   36
## [13,]   37   38   39
## [14,]   40   41   42
## [15,]   43   44   45
## [16,]   46   47   48
## [17,]   49   50   51
## [18,]   52   53   54
## [19,]   55   56   57
## [20,]   58   59   60
```

Matrix Operations

QUESTION 1.3.1

Suppose X is a matrix in R. Which of the following is not equivalent to X?

- t(t(X))
- X %*% matrix(1,ncol(X))
- X*1
- X%*%diag(ncol(X))

```
#suppose X is a 3x3 matrix with nine element
X<-matrix(1:9,3,3,byrow = TRUE)
#then
X %*% matrix(1,ncol(X) )
```

```
##      [,1]
## [1,]    6
## [2,]   15
## [3,]   24
```

QUESTION 1.3.2

Solve the following system of equations using R:

3a + 4b - 5c + d = 10

2a + 2b + 2c - d = 5

a -b + 5c - 5d = 7

5a + d = 4

```
a<-rbind(c(3 , 4 ,-5 , 1),
         c(2 , 2 , 2 ,-1),
         c(1 ,-1 , 5 ,-5),
         c(5 , 0 , 0 , 1))

b<-c(10,5,7,4)
sol<-solve(a,b)
```

What is the solution for c:

```
sol[3]
```

```
## [1] -0.8849558
```

Matrix Multiplication

Load the following two matrices into R:

```
a <- matrix(1:12, nrow=4)
b <- matrix(1:15, nrow=3)
dim(a)
```

```
## [1] 4 3
```

```
dim(b)
```

```
## [1] 3 5
```

Note the dimension of 'a' and the dimension of 'b'.

In the question below, we will use the matrix multiplication operator in R, %*%, to multiply these two matrices.

QUESTION 1.4.1

What is the value in the 3rd row and the 2nd column of the matrix product of 'a' and 'b'

```
j<-a %*% b
j[3,2]
```

```
## [1] 113
```

QUESTION 1.4.2

Multiply the 3rd row of 'a' with the 2nd column of 'b', using the element-wise vector multiplication with *.

What is the sum of the elements in the resulting vector?

```
sum(a[3,] * b[,2])
```

```
## [1] 113
```

Matrix Algebra

Linear Model As Matrix Mult. Assessment

Suppose we are analyzing a set of 4 samples. The first two samples are from a treatment group A and the second two samples are from a treatment group B. This design can be represented with a model matrix like so:

```
X <- matrix(c(1,1,1,1,0,0,1,1),nrow=4)
rownames(X) <- c("a","a","b","b")
X
```

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

Suppose that the fitted parameters for a linear model give us:

```
beta <- c(5, 2)
```

QUESTION 1.5.1

What is the fitted value for the A samples? (The fitted Y values.)

```
fitted<-X %*% beta
fitted[1:2,]
```

```
## a a
## 5 5
```

QUESTION 1.5.2

What is the fitted value for the B samples? (The fitted Y values.)

```
fitted<-X %*% beta
fitted[3:4,]
```

```
## b b
## 7 7
```

Suppose now we are comparing two treatments B and C to a control group A, each with two samples. This design can be represented with a model matrix like so:

```
X <- matrix(c(1,1,1,1,1,1,0,0,1,1,0,0,0,0,0,0,1,1),nrow=6)
rownames(X) <- c("a","a","b","b","c","c")
X
```

```
##   [,1] [,2] [,3]
## a    1    0    0
## a    1    0    0
## b    1    1    0
## b    1    1    0
## c    1    0    1
## c    1    0    1
```

Suppose that the fitted values for the linear model are given by:

```
beta <- c(10,3,-3)
```

QUESTION 1.5.3

What is the fitted value for the B samples?

```
fitted<-X %*% beta
fitted[3:4,]
```

```
##  b  b
## 13 13
```

QUESTION 1.5.4

What is the fitted value for the C samples?

```
fitted<-X %*% beta
fitted[5:6,]
```

```
## c c
## 7 7
```