Assignment1

Huiyi Huang

9/9/2019

Part 1 #1

```
S<-c(1:2019)
sum(S)
```

```
## [1] 2039190
```

```
sum(S^3)
```

```
## [1] 4.158296e+12
```

```
sum(S^S)
```

```
## [1] Inf
```

```
sum(((-1)^S-1)*S^S)
```

```
## [1] NaN
```

```
sum(1/(S^2))
```

```
## [1] 1.644439
```

```
sum(1/S)
```

```
## [1] 8.187821
```

```
sum(1/(S^3))
```

```
## [1] 1.202057
```

```
sum(((-1)^S-1)*1/S)
```

```
## [1] -8.881216
```

#2

```
x<-rnorm(1000,mean=10,sd=1)
mean(x)
```
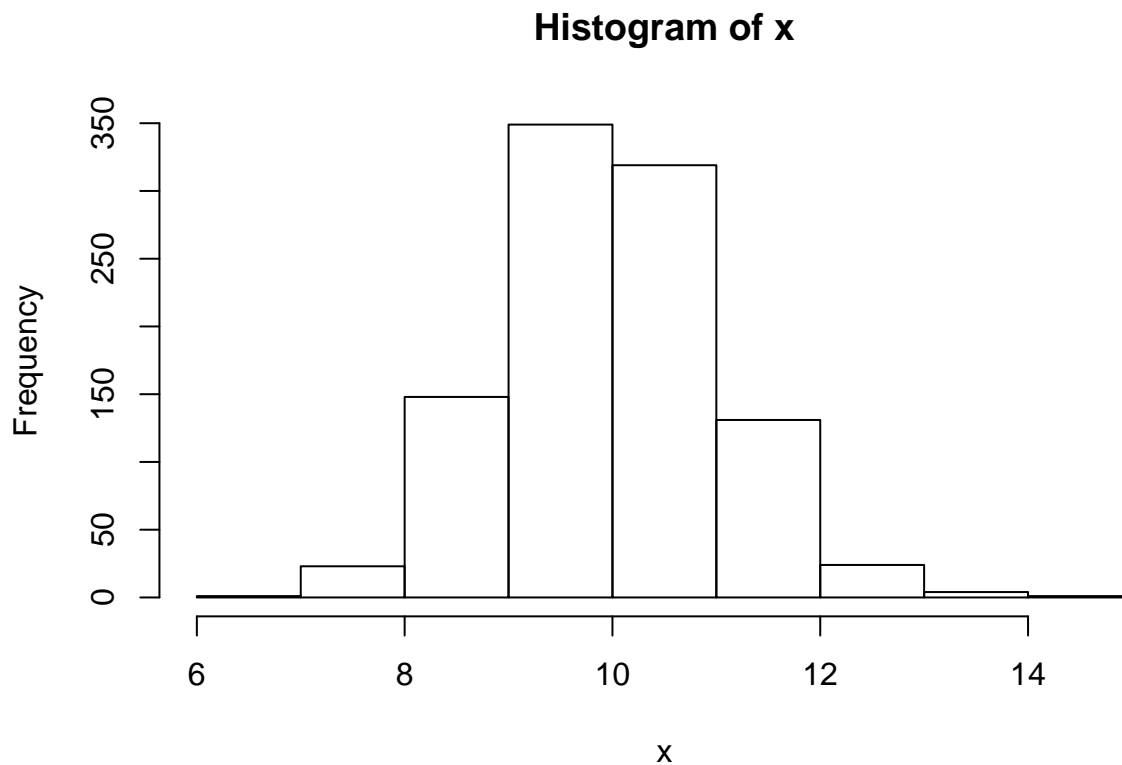
```
## [1] 9.969499
```

```r
sd(x)
```

```
## [1] 1.046466
```

```r
sum(x>10)
```

```
## [1] 479
```

```r
hist(x)
```

Histogram of x


```r
X<-rnorm(1000,mean=2,sd=1)
prob=sum(X>1)/1000
prob
```

```
## [1] 0.84
```

#3

```r
x = sample(c(1:6) ,1000, replace = TRUE)
mean(x)
```

```
## [1] 3.515
```

```
sd(x)
```

```
## [1] 1.673492
```

```
sum(x==6)
```

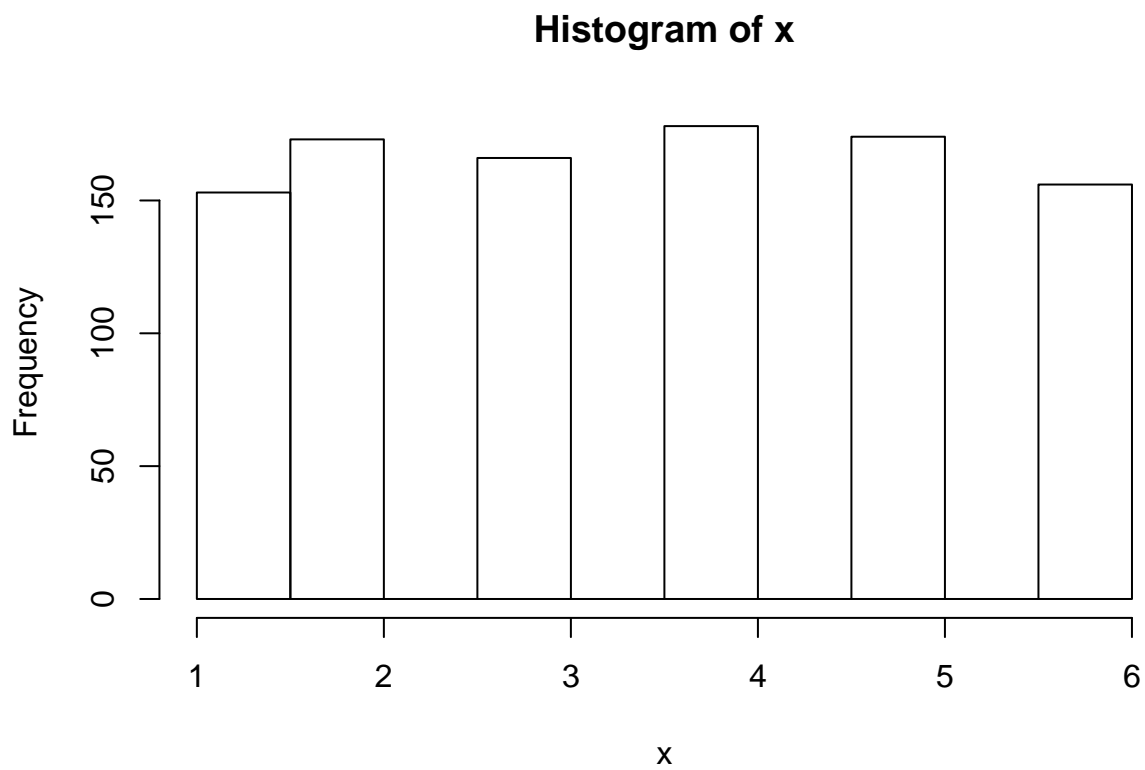```
[1] 156
```

```
table(x)
```

```
## x
##   1   2   3   4   5   6
## 153 173 166 178 174 156
```

```
prop.table(table(x))
```

```
## x
##     1     2     3     4     5     6
## 0.153 0.173 0.166 0.178 0.174 0.156
```

```
hist(x)
```

Histogram of x

#4

```
X1<-sample(1:6,1000,replace=TRUE)
X2<-sample(1:6,1000,replace=TRUE)
X3<-sample(1:6,1000,replace=TRUE)
z=sum(X1>X2+X3)
z/1000
```

```
## [1] 0.089
```

```
y=sum(X1^2>X2^2+X3^2)
y/1000
```

```
## [1] 0.227
```

#5

```
x=sample(c(0:1),3000,replace=TRUE)
d=matrix(x,ncol=3)
sum(rowSums(d)==0)/1000
```
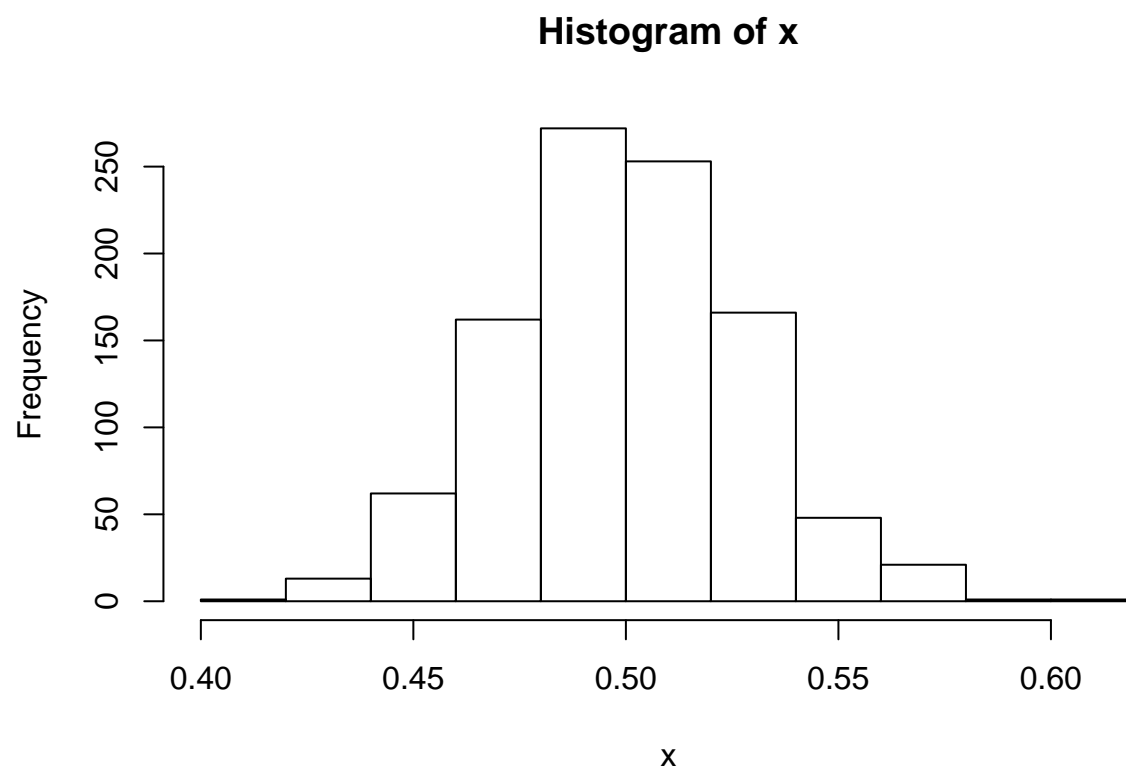
```
## [1] 0.124
```

#6

```
x=sample(c(0,1),10000,replace=TRUE)
d=matrix(x,ncol=10)
sum(rowSums(d)==3)/1000
```
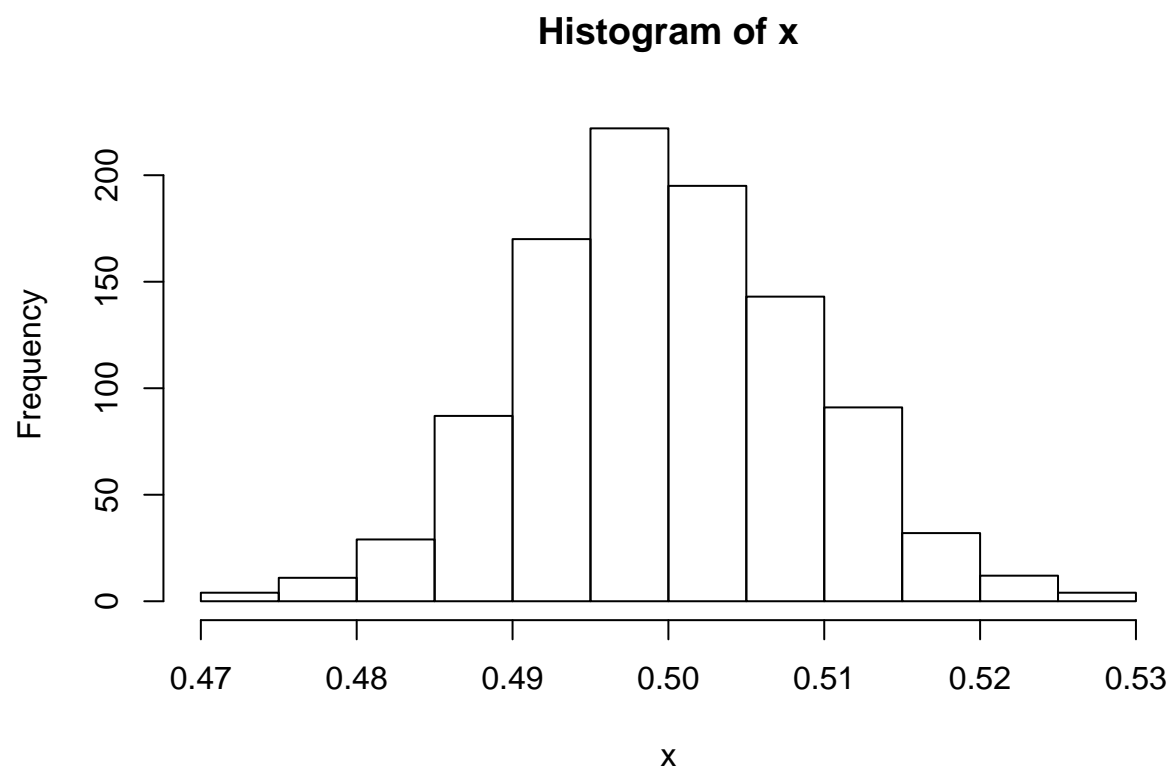
```
## [1] 0.108
```

#7

```
y=runif(100000,0,1)

d=matrix(y,ncol =100)

x=rowMeans(d)

hist(x)
```
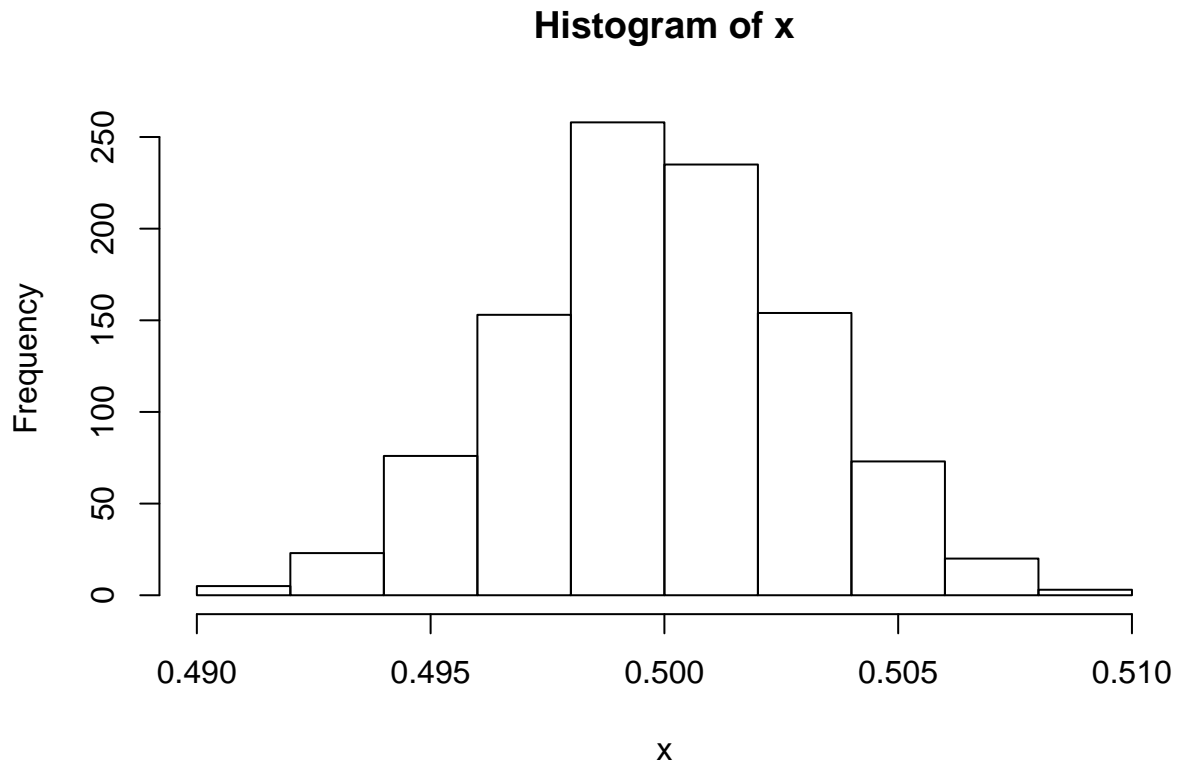
Histogram of x


```r
y=runif(1000000,0,1)

d=matrix(y,ncol =1000)

x=rowMeans(d)

hist(x)
```

Histogram of x


```
y=runif(10000000,0,1)

d=matrix(y,ncol =10000)

x=rowMeans(d)

hist(x)
```

Histogram of x

Part 2 #7

```r
setwd("C:/Users/student/Documents/Fall2019/")

Titanic<-read.csv(file= 'titanic.csv')
str(Titanic)
```

```
## 'data.frame':    891 obs. of  12 variables:
##  $ PassengerId: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Survived   : int  0 1 1 1 0 0 0 0 1 1 ...
##  $ Pclass     : int  3 1 3 1 3 3 1 3 3 2 ...
##  $ Name       : Factor w/ 891 levels "Abbing, Mr. Anthony",..: 109 191 358 277 16 559 520 629 417 58:
##  $ Sex        : Factor w/ 2 levels "female","male": 2 1 1 1 2 2 2 2 1 1 ...
##  $ Age        : num  22 38 26 35 35 NA 54 2 27 14 ...
##  $ SibSp      : int  1 1 0 1 0 0 0 3 0 1 ...
##  $ Parch      : int  0 0 0 0 0 0 0 1 2 0 ...
##  $ Ticket     : Factor w/ 681 levels "110152","110413",..: 524 597 670 50 473 276 86 396 345 133 ...
##  $ Fare       : num  7.25 71.28 7.92 53.1 8.05 ...
##  $ Cabin      : Factor w/ 148 levels "","A10","A14",..: 1 83 1 57 1 1 131 1 1 1 ...
##  $ Embarked   : Factor w/ 4 levels "","C","Q","S": 4 2 4 4 4 3 4 4 4 2 ...
```

#8

```r
#knitr::kable(Titanic)
```

#9

```r
sum(is.na(Titanic))
```

```
## [1] 177
```

```r
colSums(is.na(Titanic))
```

```
## PassengerId     Survived       Pclass         Name          Sex          Age
##           0            0            0            0            0          177
##        SibSp        Parch       Ticket         Fare        Cabin     Embarked
##           0            0            0            0            0            0
```

#10

```r
mean(Titanic$Age,na.rm=TRUE)
```

```
## [1] 29.69912
```

#11

```r
Titanic$Age[is.na(Titanic$Age)]<-mean(Titanic$Age,na.rm=TRUE)
sum(is.na(Titanic$Age))
```

```
## [1] 0
```

#12

```r
Titanic$Name = NULL
Titanic$PassingerId = NULL
Titanic$Ticket = NULL
Titanic$Cabin = NULL
```

#13

```r
mean(Titanic$Age[Titanic$Sex =='female'])
```

```
## [1] 28.21673
```

#14

```r
median(Titanic$Fare[Titanic$Pclass=='1'])
```

```
## [1] 60.2875
```

#15

```r
median(Titanic$Fare[Titanic$Sex=="female" & Titanic$Pclass=='2' | Titanic$Pclass=='3'])
```

```
## [1] 8.6625
```

#16

```r
median(Titanic$Age[Titanic$Survived=="1" & Titanic$Sex == "female" & Titanic$Pclass=='2' | Titanic$Pcla
```

```
## [1] 32.25
```

#17

```r
mean(Titanic$Fare[Titanic$Sex == "female" & Titanic$Survived == "1" & Titanic$Age < 18])
```

```
## [1] 33.17226
```

#18

```r
mean(Titanic$Fare[Titanic$Sex == "female" & Titanic$Survived == "1" & Titanic$Age < 18 & Titanic$Pclass=
```

```
## [1] 97.30239
```

```r
mean(Titanic$Fare[Titanic$Sex == "female" & Titanic$Survived == "1" & Titanic$Age < 18 & Titanic$Pclass=
```

```
## [1] 26.24167
```

```r
mean(Titanic$Fare[Titanic$Sex == "female" & Titanic$Survived == "1" & Titanic$Age < 18 & Titanic$Pclass=
```

```
## [1] 13.92259
```

#19

```r
y=mean(Titanic$Fare)
sum(Titanic$Survived == '1'& Titanic$Fare > y)/ sum(Titanic$Survived == '0'& Titanic$Fare> y)
```

```
## [1] 1.482353
```

#20

```r
Titanic$sfare <- (Titanic$Fare-mean(Titanic$Fare))/sd(Titanic$Fare)
```

#21

```r
Titanic$cfare <- ifelse(Titanic$Fare > mean(Titanic$Fare),'expensive','cheap')
names(Titanic)
```

```
##  [1] "PassengerId" "Survived"    "Pclass"      "Sex"         "Age"
##  [6] "SibSp"       "Parch"       "Fare"        "Embarked"    "sfare"
## [11] "cfare"
```

#22

```
Titanic$cage <- 0
Titanic$cage[Titanic$Age>10 & Titanic$Age<20]=1
Titanic$cage[20<Titanic$Age & Titanic$Age<30]=2
Titanic$cage[30<Titanic$Age & Titanic$Age<40]=3
Titanic$cage[40<Titanic$Age & Titanic$Age<50]=4
Titanic$cage[50<Titanic$Age & Titanic$Age<60]=5
Titanic$cage[60<Titanic$Age & Titanic$Age<70]=6
Titanic$cage[70<Titanic$Age & Titanic$Age<80]=7
```

#23

```
table(Titanic$Embarked)
```

```
##
##       C   Q   S
##   2 168  77 644
```

```
levels(Titanic$Embarked)[ ]=c("S","C","Q","S")
levels(Titanic$Embarked)
```

```
## [1] "S" "C" "Q"
```