

Assignment 1

Rachel Cardarelli

9/19/2019

```
#install.packages('tinytex')  
#tinytex::install_tinytex()
```

Question 1

```
x = c(1:2019)  
S1 = sum(x)  
S2 = sum(x^2)  
S3 = sum(x^x)  
y <- c(1,-1)  
S4 = sum((x^x)*y)
```

```
## Warning in (x^x) * y: longer object length is not a multiple of shorter  
## object length
```

```
S5 = sum(x^-2)  
S6 = sum(x^-1)  
S7 = sum(x^-3)  
S8 = sum((x^-1)*y)
```

```
## Warning in (x^-1) * y: longer object length is not a multiple of shorter  
## object length
```

Question 2

```
x <- rnorm(1000,mean=10,sd=1)  
mean(x)
```

```
## [1] 10.04155
```

```
sd(x)
```

```
## [1] 1.005077
```

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

```
## [1] 498
```

```
#hist(x)
x <- rnorm(1000,mean=2,sd=1)
sum(x>1)/1000
```

```
## [1] 0.855
```

Question 3

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

```
## [1] 3.533
```

```
sd(x)
```

```
## [1] 1.697584
```

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

```
## [1] 163
```

```
table(x)
```

```
## x
##  1  2  3  4  5  6
## 164 155 170 169 179 163
```

```
#prop.table(table(x))
#hist(x)
```

Question 4a

```
x1 = sample(c(1:6), 1000, replace=TRUE)
x2 = sample(c(1:6), 1000, replace=TRUE)
x3 = sample(c(1:6), 1000, replace=TRUE)
sum(x1 > x2 +x3)/1000
```

```
## [1] 0.091
```

Question 4b

```
sum(x1^2 > x2^2 +x3^2)/ 1000
```

```
## [1] 0.222
```

Question 5

```
m <- matrix(sample(c(0,1),3000, replace = TRUE), ncol = 3)
sum(rowSums(m)==3)/3000
```

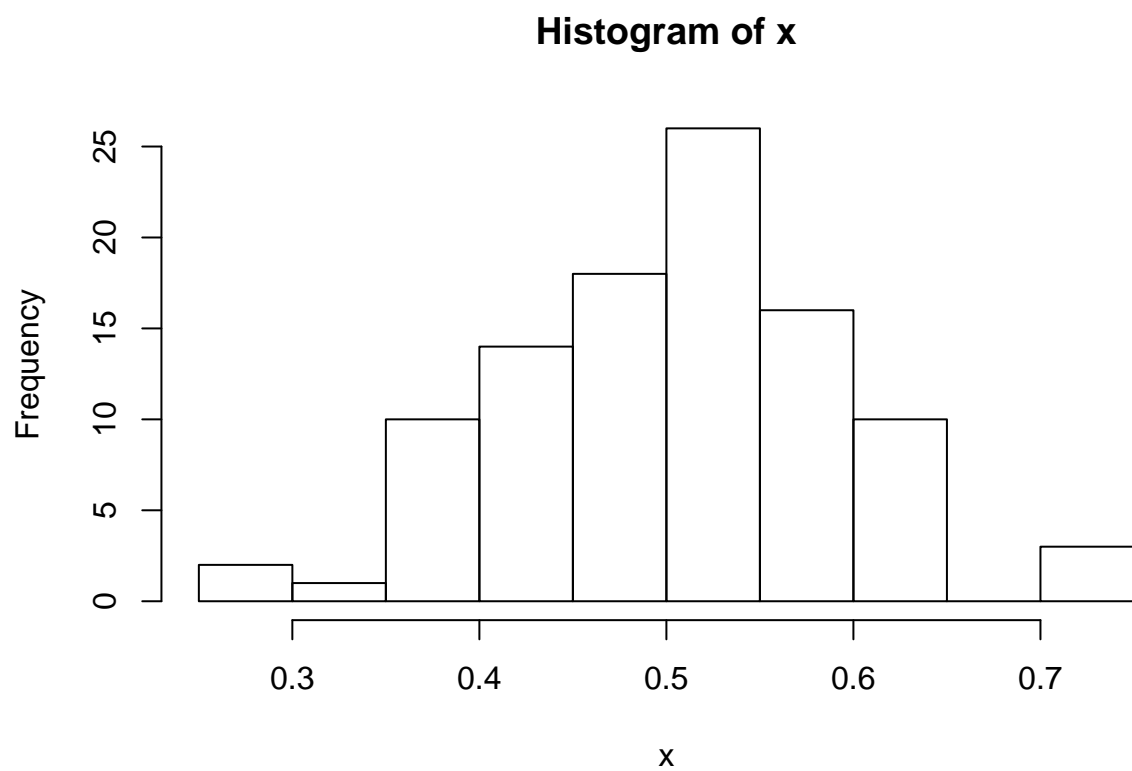
```
## [1] 0.041
```

Question 6

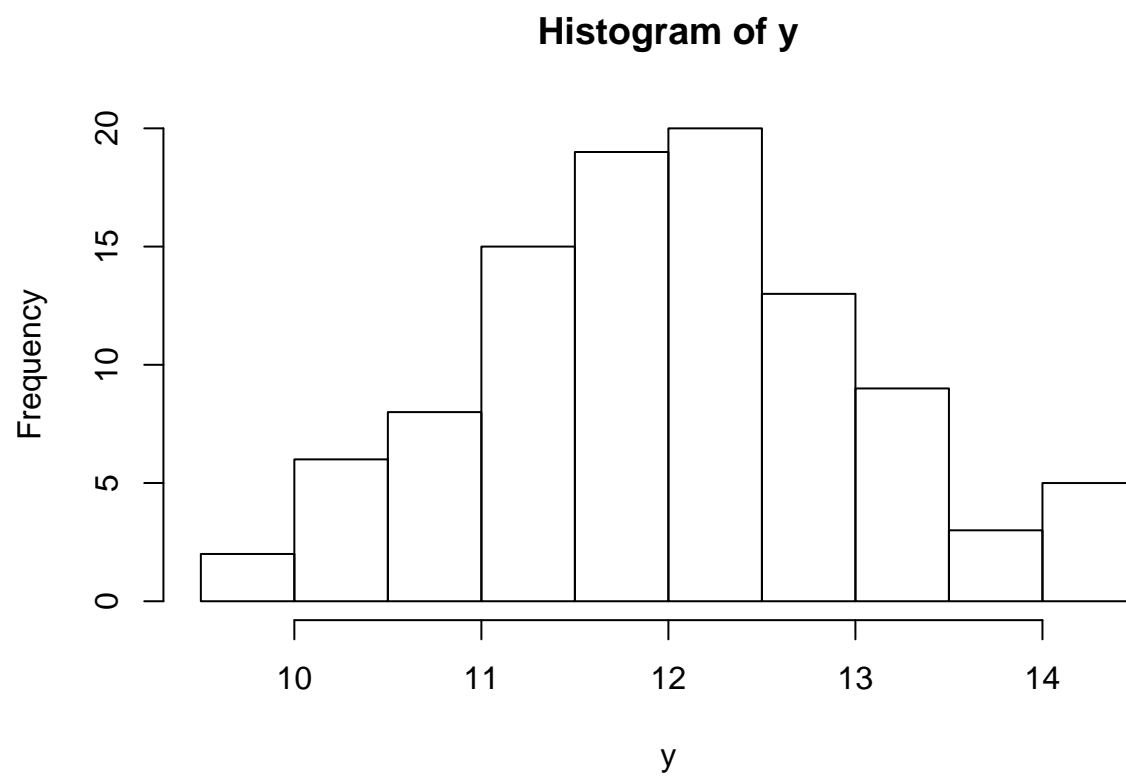
```
m <- matrix(sample(c(0,1),10000, replace = TRUE), ncol = 10)
for (n in 1000){
  for (i in 8){
    if(sum(m[n,i:(i+2)])==0){
      t = t + 1
      i = i + 1 #trying to make sure it's exactly three in a row
    }}
#t/1000
```

Question 7

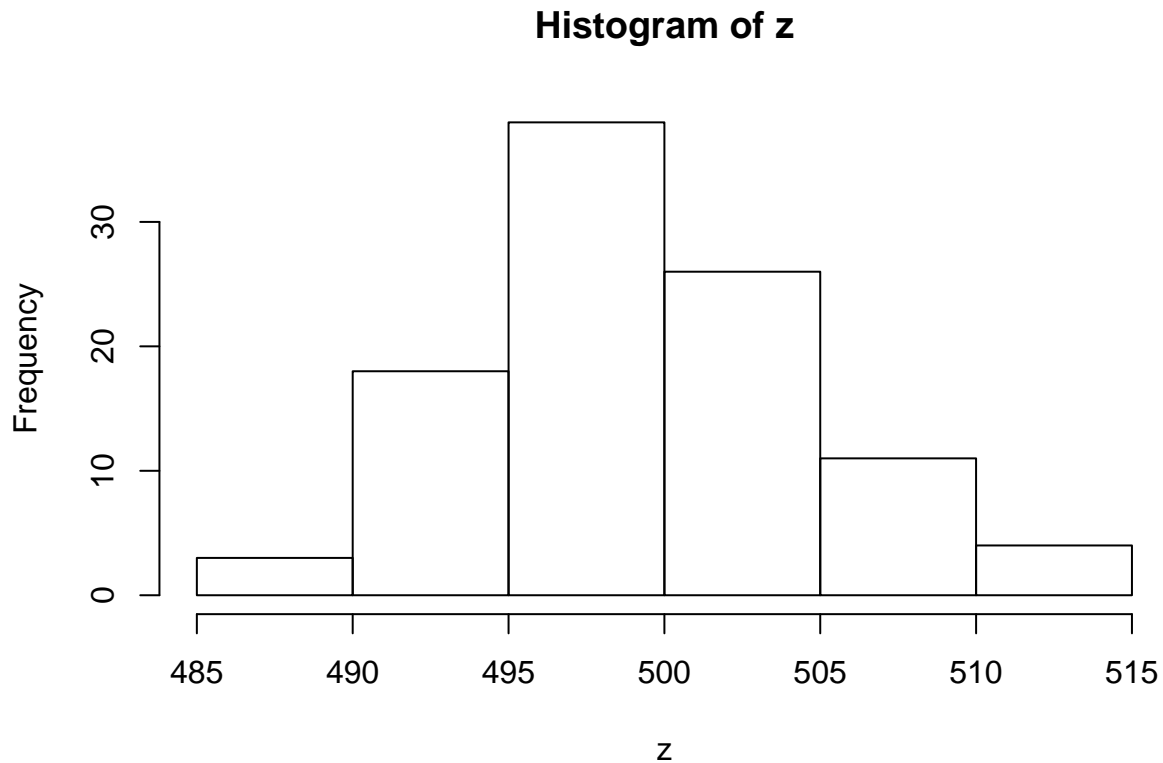
```
m <- matrix(rep(runif(1000, min = 0, max = 1),100), nrow = 100)
x <- rowMeans(m,na.rm = FALSE, dims = 1)
h <- hist(x)
```



```
n <- matrix(rep(rpois(1000,lambda = 12),100),nrow = 100)
y <- rowMeans(n,na.rm = FALSE, dims = 1)
i <- hist(y)
```



```
o <- matrix(rep(rbinom(1000,1000,.5),100),nrow = 100)
z <- rowMeans(o,na.rm = FALSE, dims = 1)
s <- hist(z)
```



Question 7

```
library(readr)
library(stringr)
titanic <- read.csv("C:/Users/student/Documents/Honors Thesis/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 ...
```

Question 8

```
knitr::kable(titanic[1:10,], format = 'markdown')
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	S	
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	S	
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	S	
6	0	3	Moran, Mr. James	male	NA	0	0	330877	8.4583	Q	
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.0750	S	
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333	S	
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708	C	

Question 9

```
sum(is.na(titanic))
```

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

```
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
```

Question 10

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

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

Question 11

```
mean_age <- mean(titanic$Age, na.rm = TRUE)
missing_Age = which(is.na(titanic$Age))
titanic$Age[missing_Age] = mean_age
```

Question 12

```
titanic <- subset(titanic, select = -c(Name, PassengerId, Ticket, Cabin))
```

Question 13

```
mean(titanic$Age[titanic$Sex == "female"], na.rm = TRUE)
```

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

Question 14

```
median(titanic$Fare[titanic$Pclass==1], na.rm = TRUE)
```

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

```
median(subset(titanic, (Pclass == 1))$Fare, na.rm = TRUE)
```

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

```
median(titanic[titanic$Pclass==1,"Fare"], na.rm = TRUE)
```

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

Question 15

```
median(titanic$Fare[titanic$Sex == "female" & titanic$Pclass != 1], na.rm = TRUE)
```

```
## [1] 14.45625
```

Question 16

```
median(subset(titanic, (Survived == 1) & (Sex == "female") & (Pclass != 3))$Age, na.rm=TRUE)
```

```
## [1] 30
```


Question 17

```
teenager = c(13:19)
mean(titanic$Fare[titanic$Sex == 'female' & titanic$Survived == 1 & match(titanic$Age, teenager)], na.rm=TRUE)

## [1] 49.17966
```

Question 18

```
mean(titanic$Fare[titanic$Sex == 'female' & titanic$Survived == 1 & match(titanic$Age, teenager)& titanic$Survived == 0])

## [1] 107.5407

mean(titanic$Fare[titanic$Sex == 'female' & titanic$Survived == 1 & match(titanic$Age, teenager)& titanic$Survived == 0] & titanic$Age < 18)

## [1] 20.00885

mean(titanic$Fare[titanic$Sex == 'female' & titanic$Survived == 1 & match(titanic$Age, teenager)& titanic$Survived == 0] & titanic$Age < 18)

## [1] 8.769885
```

Question 19

```
subset1 <- subset(titanic, titanic$Fare > mean(titanic$Fare))
sum(subset1$Survived == 1)/sum(subset1$Survived==0)

## [1] 1.482353
```

Question 20

```
sfare <- (titanic$Fare - mean(titanic$Fare))/sd(titanic$Fare)
titanic <- cbind(titanic, sfare)
```

Question 21

```
cfare <- ifelse(titanic$Fare < mean(titanic$Fare), c("cheap"), c("expensive"))
titanic <- cbind(titanic, cfare)
```

Question 22

```
attach(titanic)
```

```
## The following objects are masked _by_ .GlobalEnv:
```

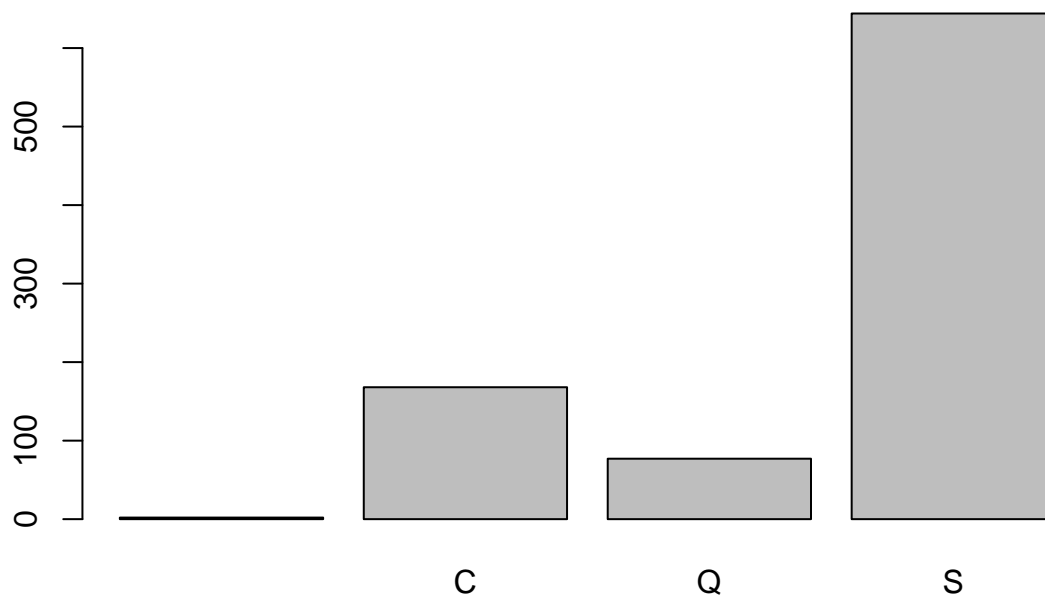
```
##
```

```
##      cfare, sfare
```

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

Question 23

```
plot(titanic$Embarked)
```



```
levels(titanic$Embarked)[1] <- "S"
```