Regina Wong
AMS 394
Homework 1: Due 11:59 p.m. June 1

Key
Blue is code
Green is output/solution

1. Consider the following heights 1.55, 1.92, 1.60, 1.75, 1.58, 1.67, 1.63, 1.82, 1.76, 1.77, 1.72, 1.85.
   (1) Assign all these heights as vector “height”.
   ```
   > height= c(1.55, 1.92, 1.60, 1.75, 1.58, 1.67, 1.63, 1.82, 1.76, 1.77, 1.72, 1.85)
   > height
        [1] 1.55 1.92 1.60 1.75 1.58 1.67 1.63 1.82 1.76 1.77 1.72 1.85
   ```
   (2) Compute the mean and standard deviation of “height”?
   ```
   > mean(height)
     [1] 1.718333
   > sd(height)
     [1] 0.1149572
   ```
   (3) What is the length of “height”?
   ```
   > length(height)
     [1] 12
   ```
   (4) How many heights are less than 1.65?
   ```
   > length(height[which(height>1.65)])
     [1] 8
   ```
   (5) Show if each height is larger than 1.60 and smaller than 1.75.
   ```
   > subset(height, 1.60<height & height<1.75)
     [1] 1.67 1.63 1.72
   ```

2. Use the following script, we can generate a 3x4 matrix: tmp <- matrix(rnorm(12), 3, 4)
   ```
   > tmp <- matrix(rnorm(12), 3, 4)
   > tmp
   ```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.62215317	-1.1369437	0.8994971	-2.3855232
[2,]	0.41709484	-0.1602884	0.2469736	-0.1514158
[3,]	-0.09940534	0.4233244	0.1886511	-1.1970627

   (1) Compute the sum of the second and third row, respectively.
   ```
   > sum(tmp[2,])
     [1] 0.3523642
   > sum(tmp[3,])
     [1] -0.6844926
   ```
   (2) Compute the product of second and fourth column, respectively.
   ```
   > sum(tmp[,2])
     [1] -0.8739078
   > sum(tmp[,4])
     [1] -3.734002
   ```

(3) Show the dimension of the matrix.

```
> dim(tmp)
[1] 3 4
```

(4) Use "cat" function to output elements in the second row that are less than 0.2.

```
> cat(subset(tmp[2,],tmp[2,]< 0.2), sep = " ", fill =FALSE)
-0.1602884 -0.1514158
```

3.

(1) Use "sample" function to generate a random vector that follows a multinomial distribution with probability (0.2, 0.3, 0.5).

```
> sample(c('1','2','3'),100,replace = TRUE, prob = c(0.2,0.3,0.5))
 [1] "3" "2" "1" "2" "3" "3" "3" "3" "3" "3" "3" "3" "3" "2" "2" "2" "3" "3" "3" "3" "2" "1"
[23] "2" "3" "3" "1" "1" "3" "3" "2" "1" "1" "3" "3" "3" "3" "3" "3" "1" "3" "1" "3" "3" "3"
[45] "2" "3" "1" "3" "3" "1" "3" "1" "2" "3" "2" "3" "2" "1" "2" "1" "3" "1" "3" "1" "3" "2"
[67] "2" "3" "1" "3" "2" "1" "3" "2" "3" "2" "3" "3" "1" "1" "2" "1" "2" "3" "1" "3" "2" "3"
[89] "2" "3" "2" "3" "3" "3" "2" "2" "3" "2" "3" "3"
```

(2) Without use the sample function, generate a random vector that follows a multinomial distribution with probability (0.2, 0.3, 0.5).

```
> x = runif(100)
> rv = c()
> for( i in x) {
+   if(i < 0.2) {
+     rv = c(rv, '1')
+   }
+   else if (i < 0.5) {
+     rv = c(rv, '2')
+   }
+   else{
+     rv = c(rv, '3')
+   }
+ }
> rv
 [1] "3" "3" "2" "2" "1" "1" "1" "2" "3" "3" "2" "3" "3" "1" "3" "1" "2" "3" "2" "1" "3" "3"
[23] "3" "3" "3" "1" "1" "1" "3" "2" "3" "3" "3" "3" "3" "2" "1" "2" "2" "1" "3" "2" "1" "3"
[45] "2" "2" "2" "1" "2" "3" "3" "3" "3" "2" "2" "2" "1" "1" "3" "2" "1" "2" "3" "3" "3" "2"
[67] "3" "3" "3" "3" "3" "2" "3" "2" "1" "3" "1" "3" "3" "1" "2" "3" "3" "1" "1" "3" "2" "2"
[89] "3" "2" "3" "3" "3" "1" "3" "2" "1" "2" "3" "1"
```

4. Calculate the probability for each of the following events:

(1) A normally distributed variable with mean 15 and standard deviation 3 is less than 16.

```
> pnorm(16, mean=15, sd =3)
[1] 0.6305587
```

(2) X < 8 in a chi-square distribution with 10 degrees of freedom.

```
> pchisq(8,10)
[1] 0.3711631
```

(3) Getting 5 out of 10 successes in a binomial distribution with probability 0.4.

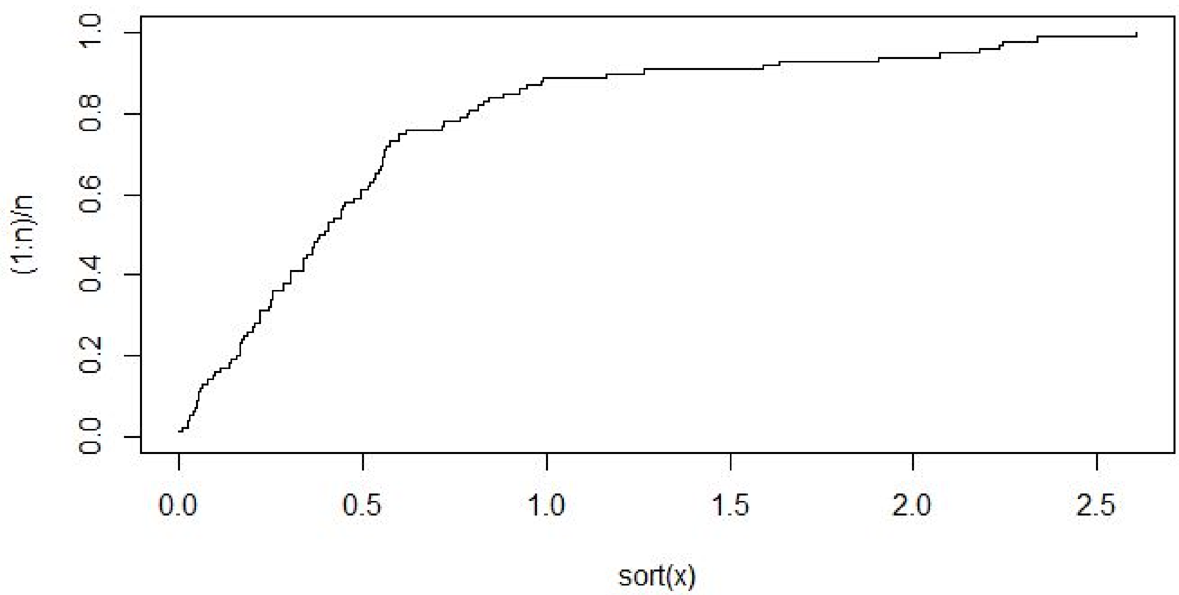
```
>   dbinom(5, 10, 0.4)
    [1] 0.2006581
```

(4) X = 5 in a Poisson distribution with λ is 3.

```
>   dpois(5,3)
    [1] 0.1008188
```

5. Generate 100 exponentially distributed random variables with rate 2 and plot the empirical distribution function.

```
>   x=rexp(100, rate=2)
>   n= length(x)
>   plot(sort(x),(1:n)/n,type="s",ylim=c(0,1))
```



6. Over the past 5 years, the mean time for a warehouse to fill a buyer's order has been 25 minutes. Officials of the company believe that the length of time has increased recently, either due to a change in the workforce or due to a change in customer purchasing policies. The processing time (in minutes) was recorded for a random sample of 15 orders processed over the past month.

   28 25 27 31 10 26 30 15 55 12 24 32 28 42 38

```
>   time = c(28, 25, 27, 31, 10, 26, 30, 15, 55, 12, 24, 32, 28, 42, 38)
```

(1) Check the normality of the data

```
>   shapiro.test(time)
            Shapiro-Wilk normality test
    data:  time
    W = 0.94167, p-value = 0.4038
```

H_0: The population is normal

H_1: The population is not normal

Since the p-value of the Shapiro-Wilk normality test is 0.4038. Since the p-value is greater than 0.05 we fail to reject the null hypotheses(H_0) and conclude that normality is met for the t-test.

(2) Please test the research hypothesis at the significance level α = 0.05

```
>     t.test(time, mu=25, alternative="greater")

        One Sample t-test

data:  time
t = 1.0833, df = 14, p-value = 0.1485
alternative hypothesis: true mean is greater than 25
95 percent confidence interval:
 22.99721      Inf
sample estimates:
mean of x
    28.2
```

H_0: mu = 25

H_1: mu > 25

Since the p-value of the t-test is 0.1485 which is greater than 0.05, we fail to reject the null hypothesis (H_0) at the 0.05 significance level. In other words, there isn't enough evidence to conclude that the mean time for the warehouse to fill a buyer's order is greater than 25 minutes.