HW03

Victor Majestic

11/11/22

Question 1

1a.i

```
p0 = (4.5^0*exp(-4.5))/factorial(0)
p1 = (4.5^1*exp(-4.5))/factorial(1)
p2 = (4.5^2*exp(-4.5))/factorial(2)
p3 = (4.5^3*exp(-4.5))/factorial(3)
p0 + p1 + p2 + p3
```

```
## [1] 0.342296
```

1a.ii

```
ppois(3, 4.5, lower.tail = TRUE)
```

```
## [1] 0.342296
```

1b.i

```
p4 = (4.5^4*exp(-4.5))/factorial(4)
p5 = (4.5^5*exp(-4.5))/factorial(5)
p6 = (4.5^6*exp(-4.5))/factorial(6)
p3 + p4 + p5 + p6
```

```
## [1] 0.6574725
```

1b.ii

```
ppois(6, 4.5, lower.tail = TRUE) - ppois(2, 4.5, lower.tail = TRUE)
```

```
## [1] 0.6574725
```

Question 2

2a

```
pnorm(6.3, 5, 3, lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 0.6676137
```

2b

```
pnorm(7.8, 5, 3, lower.tail = FALSE, log.p = FALSE)
```

```
## [1] 0.1753239
```

2c

```
pnorm(7.8, 5, 3, lower.tail = TRUE, log.p = FALSE) - pnorm(1.6, 5, 3,
lower.tail = TRUE, log.p = FALSE)

## [1] 0.6961389
```

2d

```
qnorm(0.95, 5, 3, lower.tail = TRUE)

## [1] 9.934561
```

2e

```
lowerbound = qnorm(0.025, 5, 3, lower.tail = TRUE)
-lowerbound + 5

## [1] 5.879892
```

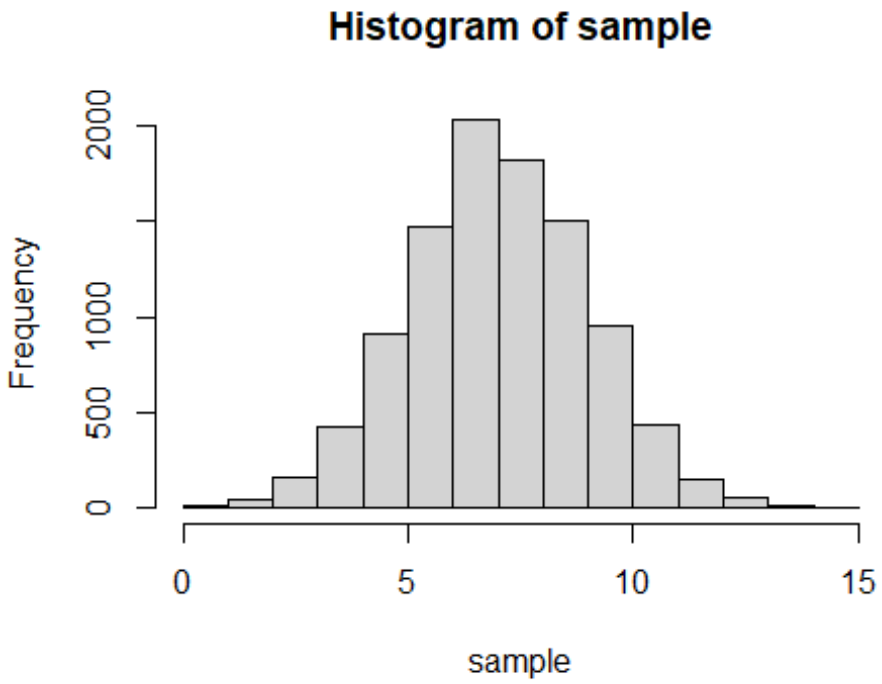
Question 3

3a

```
set.seed(101)
sample = rnorm(10000, 7, 2)
mean(sample)

## [1] 7.010556

sd(sample)

## [1] 1.986348
```

3b

```
hist(sample)
```



Question 6.1.12

6.1.12a

```
Rainfall <- read.csv("Ex6_1_12.csv", header=T)
mean(Rainfall$rainfall)

## [1] 303.2865

sd(Rainfall$rainfall)

## [1] 514.9998

range(Rainfall$rainfall)

## [1]    1.0 2745.6

max(Rainfall$rainfall) - min(Rainfall$rainfall)

## [1] 2744.6
```

6.1.12b

```
Unseeded <- Rainfall[1:26, 1]
mean(Unseeded)

## [1] 164.5885

sd(Unseeded)

## [1] 278.4264
```

```
range(Unseeded)

## [1]    1.0 1202.6

max(Unseeded) - min(Unseeded)

## [1] 1201.6
```

6.1.12c

```
Seeded <- Rainfall[27:52, 1]
mean(Seeded)

## [1] 441.9846

sd(Seeded)

## [1] 650.7872

range(Seeded)

## [1]    4.1 2745.6

max(Seeded) - min(Seeded)

## [1] 2741.5
```

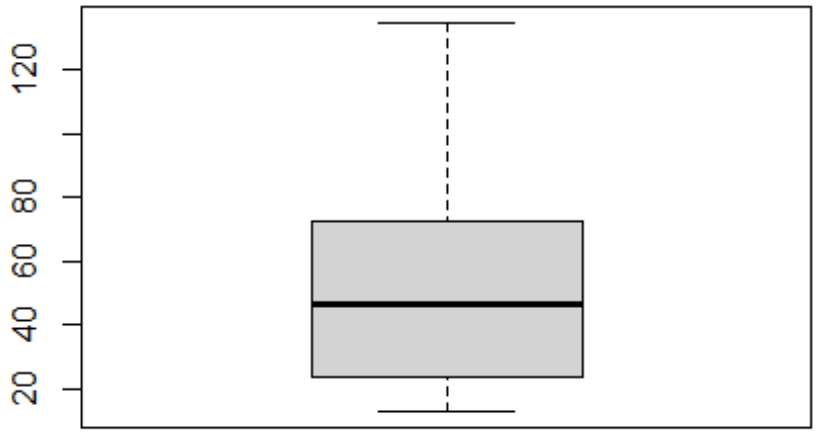
Question 6.2.4

```
Gasoline <- read.csv("Ex6_2_4.csv", header=T)
stem(Gasoline$OctRating)

##
##   The decimal point is at the |
##
##    82 | 4
##    84 | 333
##    86 | 777456789
##    88 | 233345566790233678899
##    90 | 011134445678900011122566 88
##    92 | 22236777023347
##    94 | 2247
##    96 | 15
##    98 | 8
##   100 | 3

quantile(Gasoline$OctRating, probs = c(0, 0.25, 0.5, 0.75, 1))

##    0%   25%   50%   75%  100%
##  83.4  88.6  90.4  92.2 100.3
```

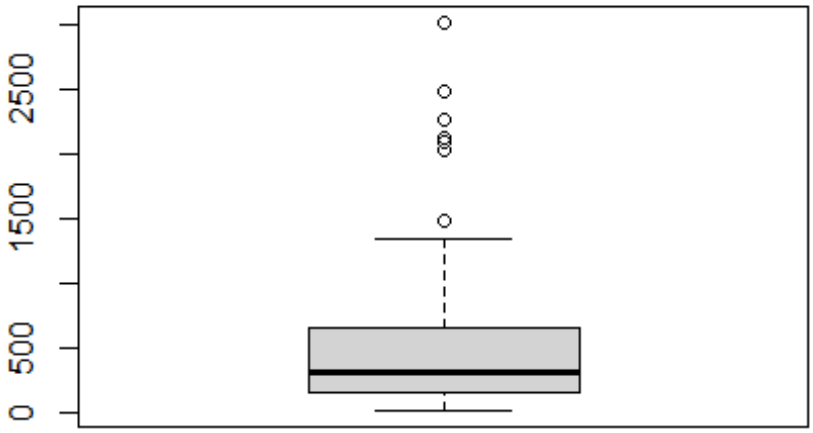
Question 6.4.9

```
Drug <- read.csv("Ex6_4_9.csv", header = T)
HighDose <- Drug[1:21, 1]
```

```
Control <- Drug[22:82, 1]
boxplot(HighDose)
```




```
boxplot(Control)
```



Neither distribution follows a normal curve. Both distributions are right-skewed, with the control group being an almost extreme case. The high-dose distribution has a median between 40 and 60. 25% of gene activity values fall between roughly 75 and the maximum 134.9; meanwhile the lower 75% fall within a similar range of the minimum 12.9 and roughly 75. The control group is even more right-skewed; 50% of the data values are greater than roughly 330. The maximum is around 3000. It appears as if this distribution has 7 extreme outliers. It is also evident that the median (and basically all of the distribution) of the high-dose group is much lower than that of the control group, indicating that the high-dose is effective in lowering gene activity.