CS544 Module 4 Assignment

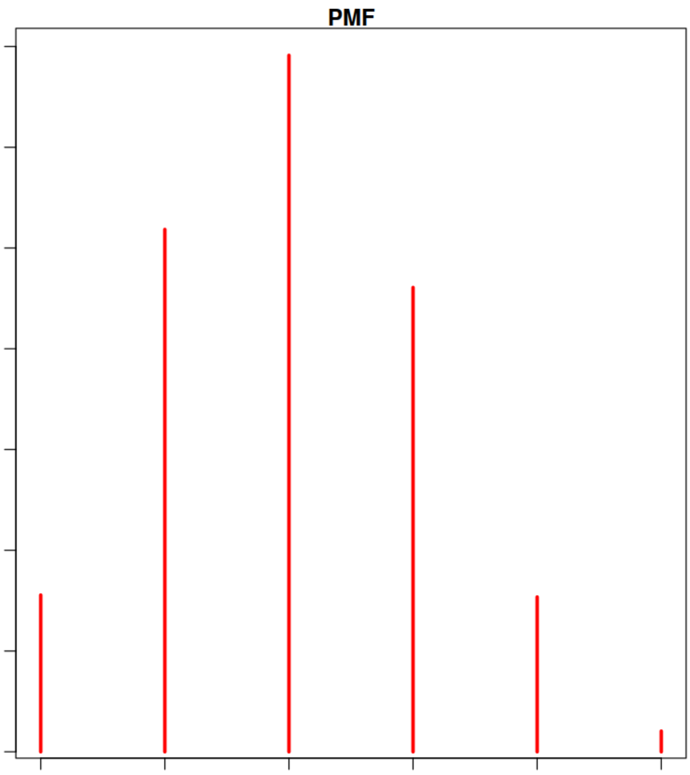
Part1) Binomial distribution (20 points) Suppose a student has 40% chance of scoring a perfect score in an exam with randomly selected questions. Each student will be provided 5 attempts.

```
> #part 1
> p1 <- 0.4
> n1 <- 5
```

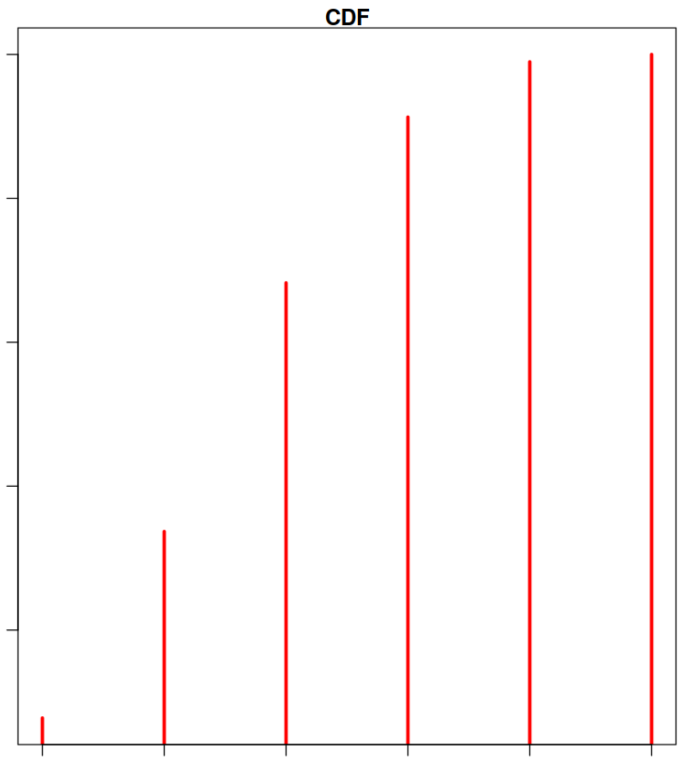
a) Compute and plot the probability distribution for the number of perfect scores over the 5 attempts (both the PMF and CDF)

```
> #a
> success <- 0:n1
> cdf_1 <- pbinom(success, size=n1, prob=p1)
> print(cdf_1)
[1] 0.07776 0.33696 0.68256 0.91296 0.98976 1.00000
> pmf_1 <- dbinom(success, size=n1, prob=p1)
> print(pmf_1)
[1] 0.07776 0.25920 0.34560 0.23040 0.07680 0.01024
> plot(success, cdf_1 ,type='h',xlab="Sucesses",ylab="Probabilities",col="red",lwd=3,main="CDF")
> plot(success, pmf_1,type='h',xlab="Sucesses",ylab="Probabilities",col="red",lwd=3,main="PMF")
```

PMF



CDF



b) What is the probability that a student will score a perfect score in exactly 2 out of the 5 attempts?

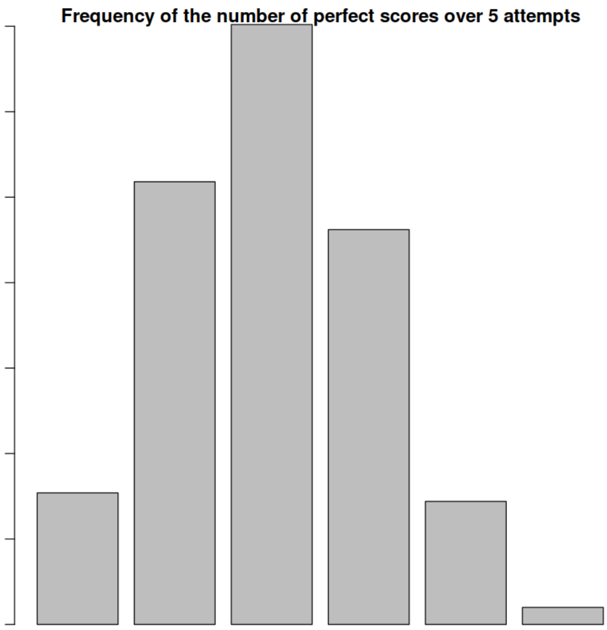
```
> #b
> p_b <- dbinom(2,n1,p1)
> print(p_b)
[1] 0.3456
```

c) What is the probability that a student will score a perfect score in at least 2 out of the 5 attempts?

```
> #c
> p_c <- 1-pbinom(1,n1,p1)
> print(p_c)
[1] 0.66304
```

d) Simulate the number of perfect scores over 5 attempts for 1000 students. Show the barplot of the frequencies of successes.

```
> #d
> set.seed(123)
> num_perfect_scores <- rbinom(1000, size = n1, prob = p1)
> barplot(table(num_perfect_scores), main = "Frequency of the number of perfect scores over 5 attempts",
+         xlab = "Number of perfect scores", ylab = "Frequency")
>
```

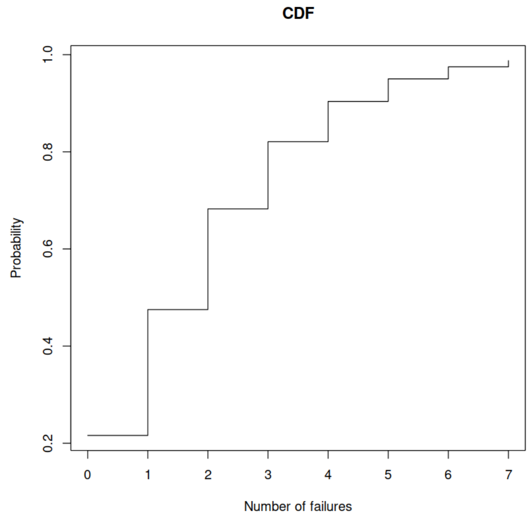


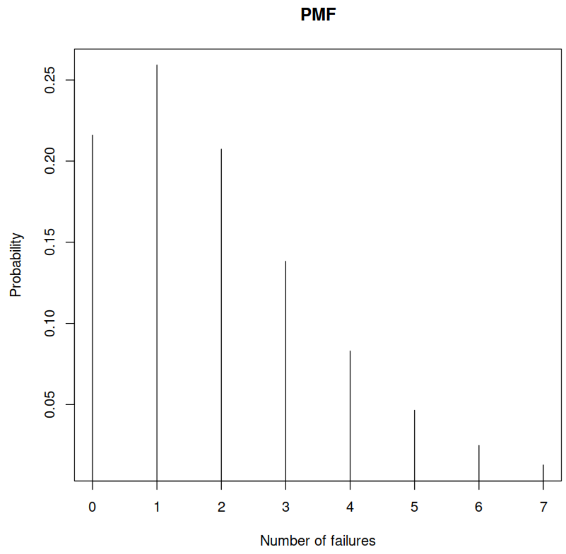
Part2) Negative Binomial distribution (20 points) Suppose a student has 60% chance of scoring a perfect score in an exam with randomly selected questions. The student has to repeatedly take the exam until they achieve three perfect scores.

```
> #part 2
> p2 <- 0.6
> r <- 3
> n2 <- 10
```

a) Compute and plot the probability distribution for scoring the three perfect scores (both the PMF and CDF). The student will only go for a maximum of 10 failures before giving up.

```
> #a
> x <- 0:(n2 - r)
> pmf2 <- dnbinom(x, size = r, prob = p2)
> print(pmf2)
[1] 0.21600000 0.25920000 0.20736000 0.13824000 0.08294400 0.04644864 0.02477261
[8] 0.01274020
> plot(x, pmf2, type = "h", xlab = "Number of failures", ylab = "Probability",
+      main = "PMF")
> cdf2 <- pnbinom(x, size = r, prob = p2)
> print(cdf2)
[1] 0.2160000 0.4752000 0.6825600 0.8208000 0.9037440 0.9501926 0.9749652 0.9877054
> plot(x, cdf2, type = "s", xlab = "Number of failures", ylab = "Probability",
+      main = "CDF")
>
```





b) What is the probability that the student will have the three perfect scores with exactly 4 failures?

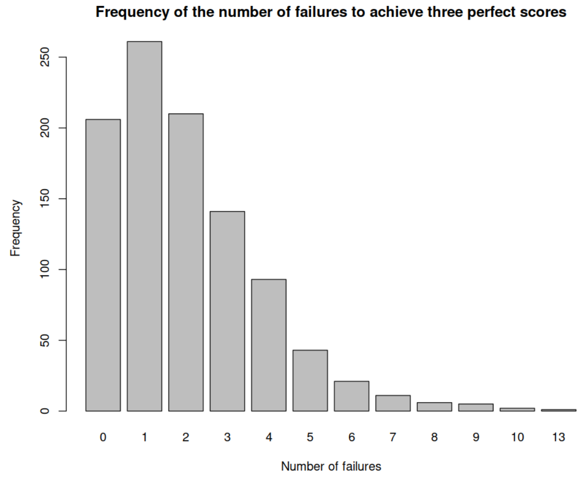
```
> #b
> prob_4 <- dnbinom(4, size = r, prob = p2)
> print(prob_4)
[1] 0.082944
```

c) What is the probability that the student will have the three perfect scores with at most 4 failures?

```
> #c
> prob_atmost_4 <- pnbinom(4, size = r, prob = p2)
> print(prob_atmost_4)
[1] 0.903744
```

d) Simulate the number of failures to get the three perfect scores for 1000 students. Show the barplot of the frequencies of the failures.

```
> #d
> set.seed(123)
> num_failures <- rnbinom(1000, size = r, prob = p2)
>
> barplot(table(num_failures), main = "Frequency of the number of failures to achieve three perfect scores",
+         xlab = "Number of failures", ylab = "Frequency")
>
```

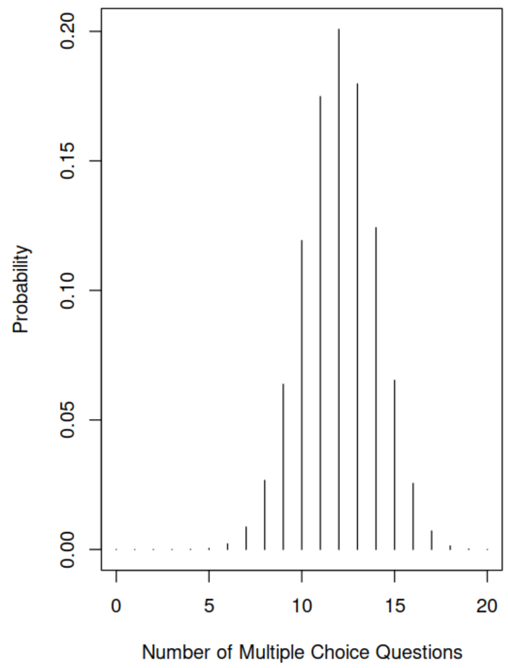


Part3) Hypergeometric distribution (20 points) Suppose that your professor created a pool of 60 multiple choice questions and 40 programming questions for the final exam. For each student, a random set of 20 distinct questions from the pool will be presented during the exam. The student has the opinion that the multiple-choice questions are easy to handle than the programming questions.

```
> #part 3
> n3 <- 20
> p3 <- 0.6
```

a) Compute and plot the probability distribution for the number of multiple choice questions out of the 20 questions that the student will be given?

```
#a
x <- 0:20
pmf3a <- dhyper(x, m = p3*100, n = 40, k = n3)
plot(x, pmf3a, type = "h", xlab = "Number of Multiple Choice Questions", ylab = "Probability")
```



b) What is the probability that the student will have exactly 10 multiple choice questions out of the 20 questions in the exam?

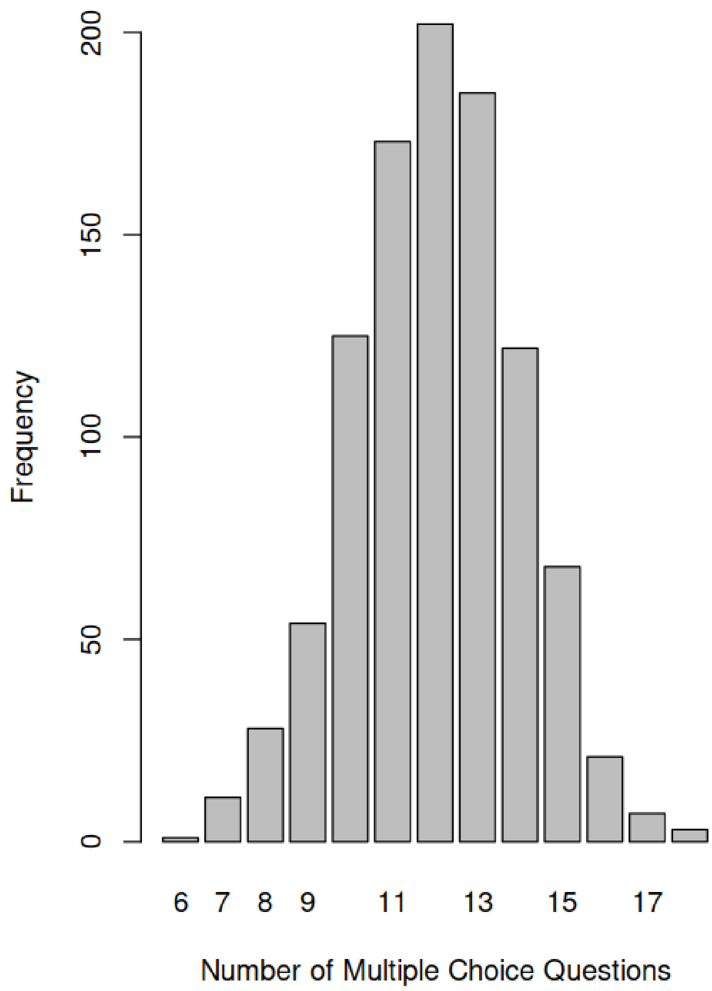
```
> #b
> p_exact10 <- dhyper(10, m = p3*100, n = 40, k = n3)
> print(p_exact10)
[1] 0.1192361
> (choose(p3*100, 10)*choose(40,10))/choose(p3*100 + 40,20)
[1] 0.1192361
```

c) What is the probability that the student will have at least 10 multiple choice questions out of the 20 questions in the exam?

```
> #c
> p_least10 <- 1 - phyper(9, m = p3*100, n = 40, k = n3)
> print(p_least10)
[1] 0.8982561
```

d) Simulate the number of multiple choice questions for 1000 students. Show the barplot of the frequencies of the multiple-choice questions.

```
> #d
> set.seed(123)
> sim_data <- rhyper(1000, m = p3*100, n = 40, k = n3)
> barplot(table(sim_data), xlab = "Number of Multiple Choice Questions", ylab = "Frequency")
> 
```



Part4) Poisson distribution (20 points) Suppose that, on an average, students email 10 questions per day to the professor.

```
> #part 4
> numQ <- 10
>
```

a) What is the probability that the professor will have to answer exactly 8 questions per day?

```
> #a
> print(dpois(8, numQ))
[1] 0.112599
```

b) What is the probability that the professor will have to answer at most 8 questions per day?

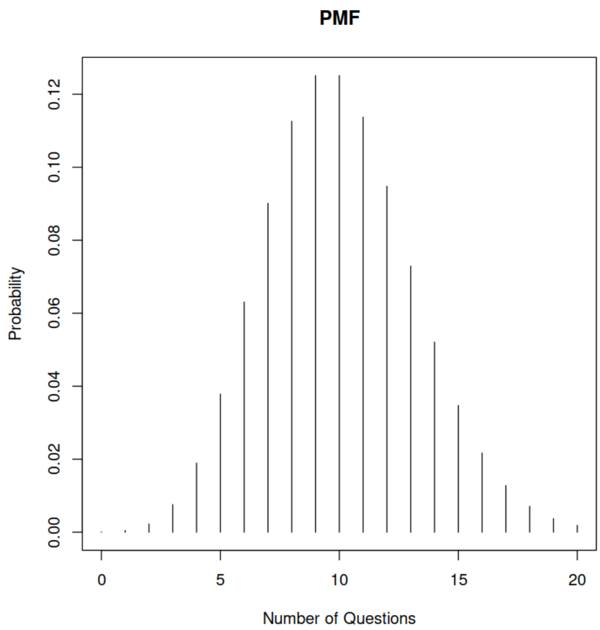
```
> #b
> print(ppois(8, numQ))
[1] 0.3328197
```

c) What is the probability that the professor will have to answer between 6 and 12 questions (inclusive)?

```
> #c
> print(ppois(12, numQ) - ppois(5, numQ))
[1] 0.7244705
```

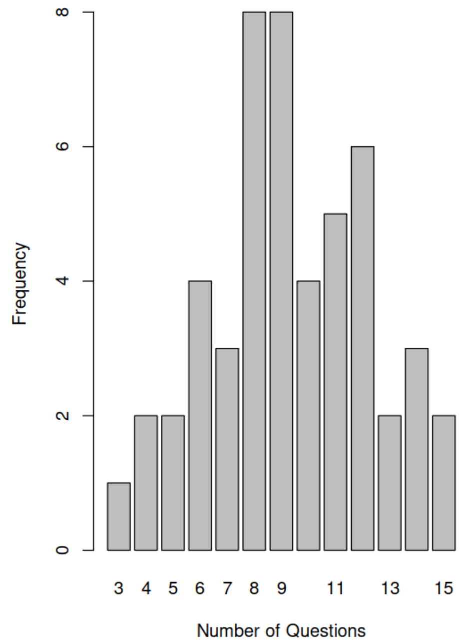
d) Calculate and plot the PMF for the first 20 questions.

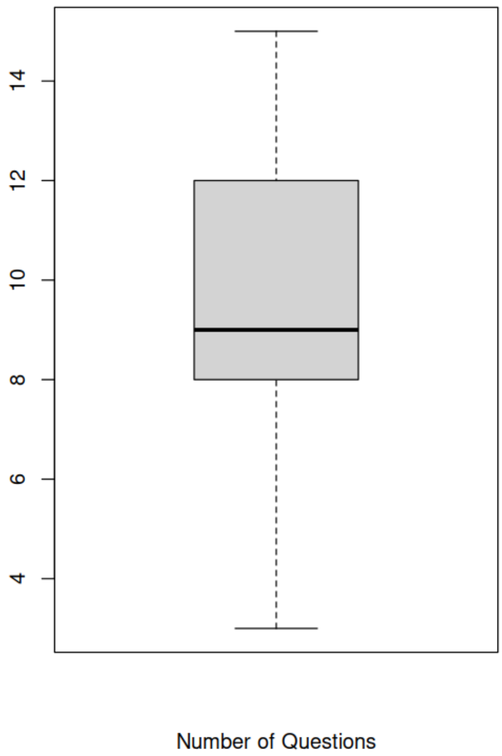
```
> #d
> x <- 0:20
> prob <- dpois(x, numQ)
> plot(x, prob, type = "h", xlab = "Number of Questions", ylab = "Probability", main = "PMF")
>
```



e) Suppose the course runs for 50 days. Simulate the number of questions the professor gets per day over the course run. Show a barplot of the frequencies of the number of questions. Show a boxplot of the number of questions. What do you infer from the plots?

```
> #e
> set.seed(123)
> sim_data <- rpois(50, numQ)
> barplot(table(sim_data), xlab = "Number of Questions", ylab = "Frequency")
> boxplot(sim_data, xlab = "Number of Questions")
>
```



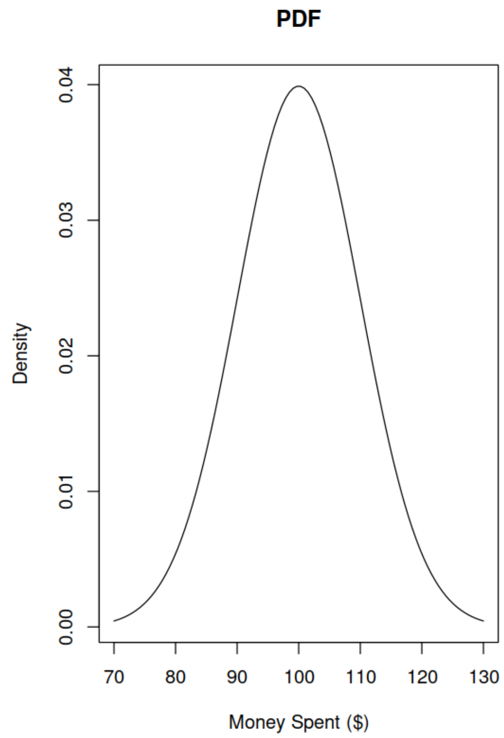


Part5) Normal distribution (20 points) Suppose that visitors at a theme park spend an average of $100 on souvenirs. Assume that the money spent is normally distributed with a standard deviation of $10.

```
> #part 5
> Avg <- 100
> stand_d <- 10
>
```

a) Plot the PDF of this distribution covering the three standard deviations on either side of the mean.

```
> #a
> x <- seq(Avg - 3*stand_d, Avg + 3*stand_d, length.out = 100)
> y <- dnorm(x, mean = Avg, sd = stand_d)
> plot(x, y, type = "l", main = "PDF",
+       xlab = "Money Spent ($)", ylab = "Density")
>
```



b) What are the chances that a randomly selected visitor will spend more than $120?

```
> #b
> print(1 - pnorm(120, mean = Avg, sd = stand_d))
[1] 0.02275013
```

c) What is the chance that a randomly selected visitor will spend between $80 and $90 (inclusive)?

```
> #c
> print(pnorm(90, mean = Avg, sd = stand_d) - pnorm(80, mean = Avg, sd = stand_d))
[1] 0.1359051
```

d) What are the chances of spending within one standard deviation, two standard deviations, and three standard deviations, respectively?

```
> #d
> print(1 - pnorm(Avg + stand_d, mean = Avg, sd = stand_d) + pnorm(Avg - stand_d, mean = Avg, sd = stand_d))
[1] 0.3173105
> print(1 - pnorm(Avg + 2*stand_d, mean = Avg, sd = stand_d) + pnorm(Avg - 2*stand_d, mean = Avg, sd = stand_d))
[1] 0.04550026
> print(1 - pnorm(Avg + 3*stand_d, mean = Avg, sd = stand_d) + pnorm(Avg - 3*stand_d, mean = Avg, sd = stand_d))
[1] 0.002699796
>
```

e) Between what two values will the middle 80% of the money spent will fall?

```
> #e
> print(qnorm(0.1, mean = Avg, sd = stand_d))
[1] 87.18448
> print(qnorm(0.9, mean = Avg, sd = stand_d))
[1] 112.8155
```

f) If the theme park gives a free T-shirt for the top 2% of the spenders, what will be the minimum amount you have to spend to get the free T-shirt?

```
> #f
> print(qnorm(0.98, mean = Avg, sd = stand_d))
[1] 120.5375
>
```

g) Show a plot for 10,000 visitors using the above distribution.

```
> #g
> set.seed(123)
> visitors <- rnorm(10000, mean = Avg, sd = stand_d)
> hist(visitors, main = "Money Spent by Visitors",
+      xlab = "Money Spent ($)", col = "red", breaks = 20)
>
```