Question 3:

```r
#3d)i)
set.seed(20870815)
N  = 10000
all_samples= rbinom(N, 8, 1/8)
head(all_samples)

#3d)ii)
mean(all_samples)
```

```
> head(all_samples)
[1] 1 3 2 2 2 0
> #3d)ii)
> mean(all_samples)
[1] 0.986
```

```r
#3d)i)
set.seed(20870815)
N <- 10000
all_samples <- t(replicate(N, sample(c("Correct", "Incorrect"), 8,
                                     replace = TRUE, prob = c(1/8, 7/8))))
a <- 1:N
for (row in 1:N) {
  a[row] <- (sum(all_samples[row,] == "Correct"))
}
head(a)
```

4 2 0 1 0 0

```r
#3d)ii)
mean(a)
```

1.0125

```r
#3e)i)
# I'd expect Y to have a larger mean as it is a distribution without replacement
# thus the probability of the letter reaching the correct person gets larger
# with each trial due to the number of people to send it to decreasing each time
# (as p = 1/N). Thus we can expect that on average the number of people who
# received the correct letters based off of Y's distribution will be higher.
```

```r
#3e)ii)
set.seed(20870815)
N <- 10000
all_sample2 <- t(replicate(N, sample(1:8, 8, replace = FALSE, prob = rep(1/8, 8))))
a2 <- 1:N
for (row in 1:N) {
  count <- 0
  for (col in 1:8) {
    if (col == all_sample2[row,col]) {
      count <- count + 1
    }
  }
  a2[row] <- count
}
head(a2)
```

0 0 0 2 1 0

```r
#3e)iii)
mean(a2)
```

1.0155

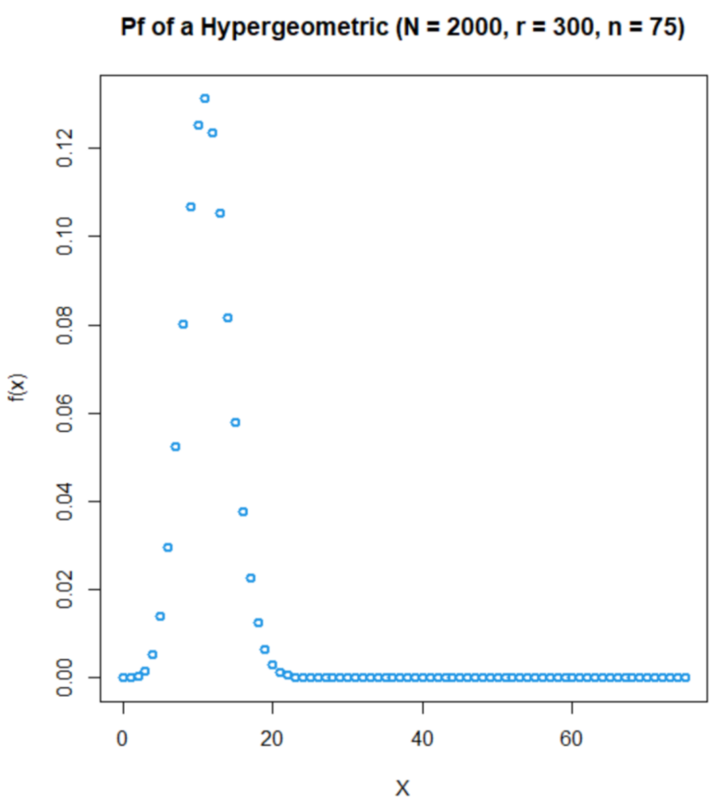
```r
#3f)
# The mean of the random variable Y is larger than the mean of variable X
# by 0.0295. This isn't a very large difference, though it is a notable one.
# Note that we can say that X has a binomial distribution while Y has a
# hypergeometric one as it is without replacement and since the size of the
# sample taken is not very large, X can not be used to approximate Y, thus a
# a notable discrepancy is expected.
```

Question 1:

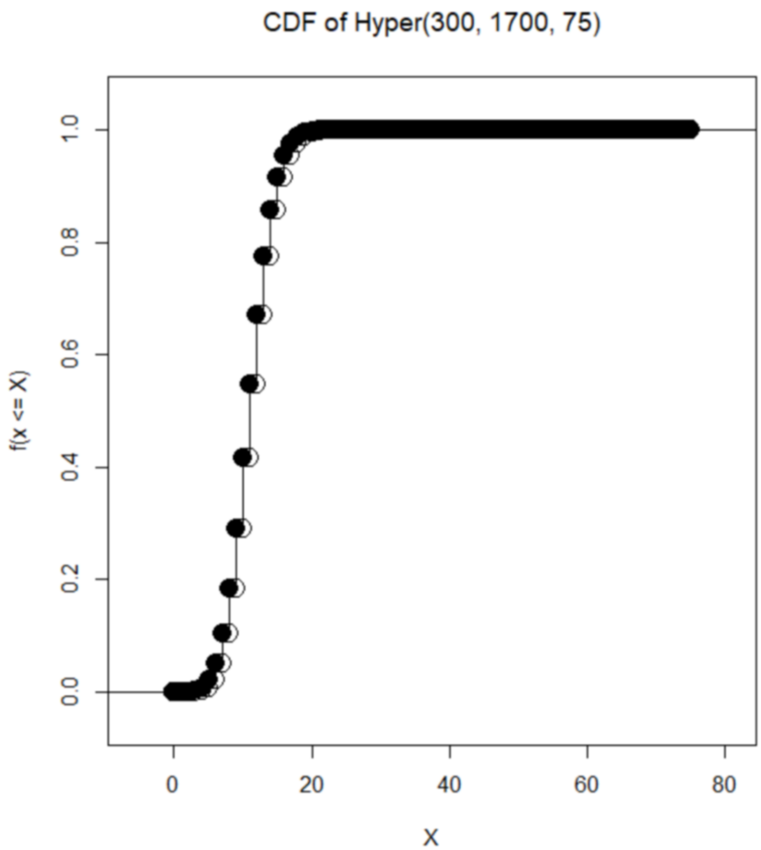
```
# Q1)
# a) i)
# This is a Hypergeometric distribution, with the population,
# N = 2000, number of successes, r = 300 and number of objects in
# the sample is 75

# a)ii)
lessthanfourtyh <- phyper(40, 300, 1700, 75)
lessthanfifteenh <- phyper(15, 300, 1700, 75)
fifteenh <- dhyper(15, 300, 1700, 75)
fifteentofourtyh <- lessthanfourtyh - lessthanfifteenh + fifteenh
fifteentofourtyh
```

P(15 <= X <= 40): 0.1426305

```
# a) iii)
x <- seq(0, 75, by = 1)
y <- dhyper(x, 300, 1700, 75)
plot(x, y, type="p", xlab="X", ylab="f(x)", col=4,
     main="Pf of a Hypergeometric (N = 2000, r = 300, n = 75)",
     lwd=2)
```



```r
# a) iv)
library(distr)
X <- Hyper(m=300, n=1700, k=75)
plot(X, to.draw.arg = "p")
```

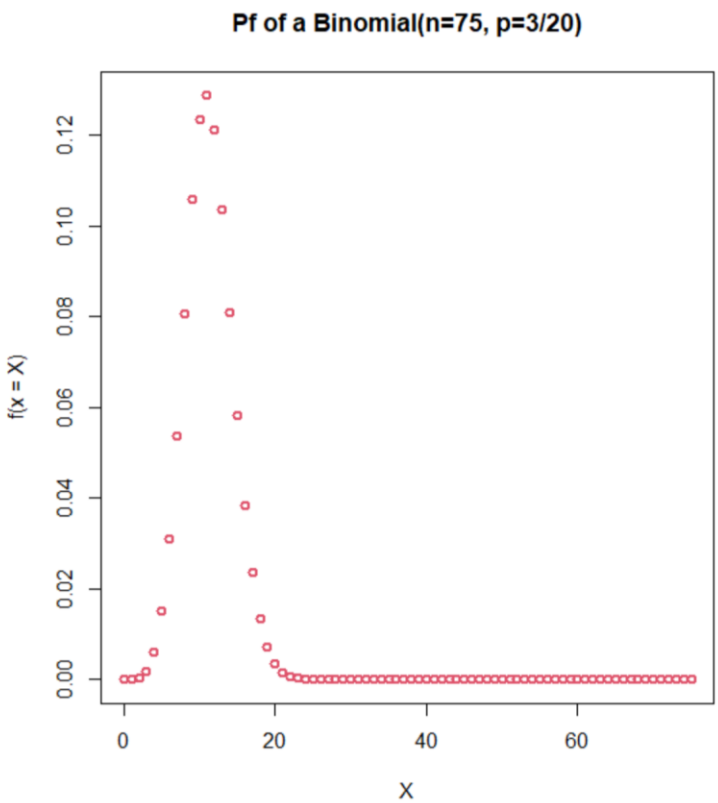



```r
#a) v)
twentyfive <- qhyper(0.25, 300, 1700, 75)
fifty <- qhyper(0.50, 300, 1700, 75)
seventyfive <- qhyper(0.75, 300, 1700, 75)
```

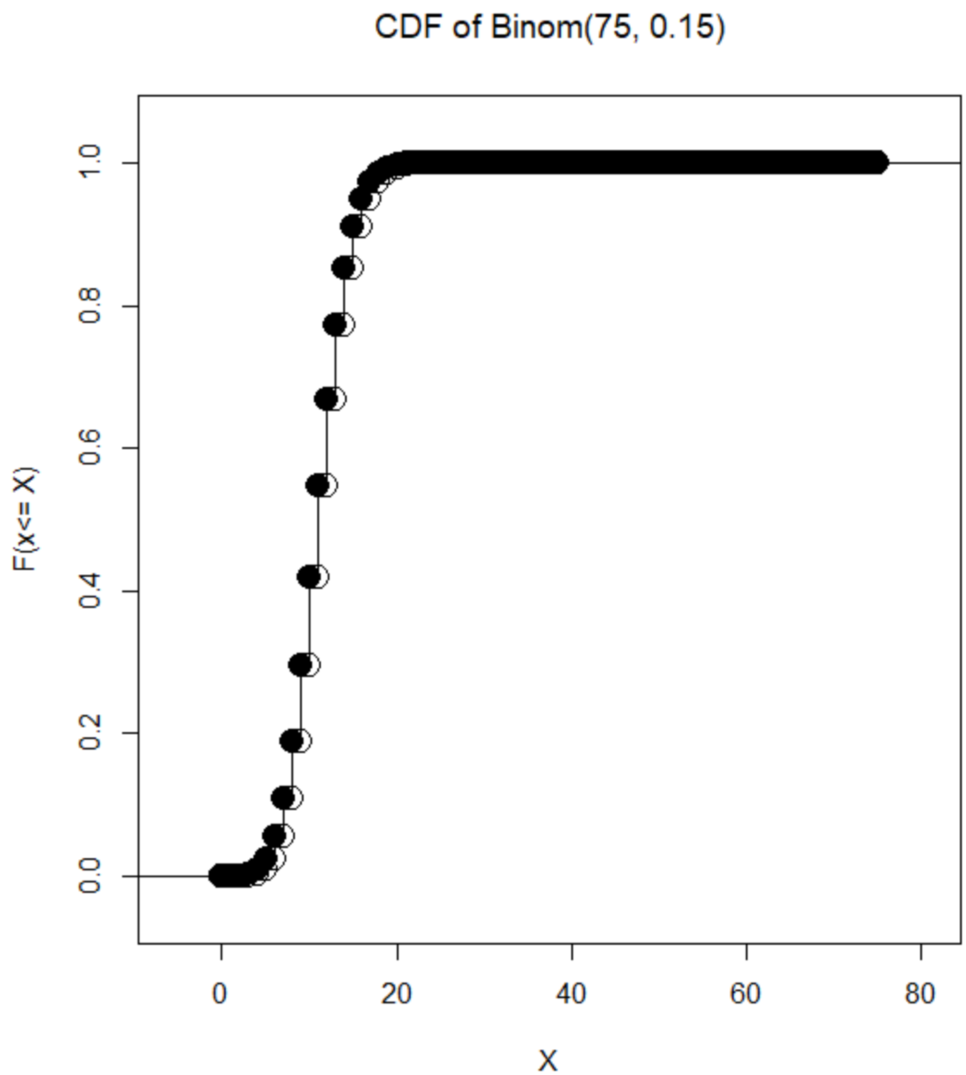
```
> print(twentyfive)
[1] 9
> print(fifty)
[1] 11
> print(seventyfive)
[1] 13
```

```r
#b) i)
# This is a binomial distribution with size of the sample 75 and
# probability 0.15 of success

#b) ii)
lessthanfourtyb <- pbinom(40, 75, 3/20)
lessthanfourteenb <- pbinom(14, 75, 3/20)
fifteentofourteenb <- lessthanfourtyb - lessthanfourteenb
fifteentofourteenb
```

P(15 <= X <= 40): 0.1469411



```r
#b) iv)
library(distr)
X <- Binom(size=75, p=3/20)
plot(X, to.draw.arg = "p", xlab="X", ylab="F(x<= X)")
```



```r
#b) v)
twentyfiveq <- qbinom(0.25, 75, 3/20)
fiftyq <- qbinom(0.5, 75, 3/20)
seventyfiveq <- qbinom(0.75, 75, 3/20)
print(twentyfiveq)
print(fiftyq)
print(seventyfiveq)
```

```
> print(twentyfiveq)
[1] 9
> print(fiftyq)
[1] 11
> print(seventyfiveq)
[1] 13
```

```r
#c)
# The probabilities and the percentiles are similar and we can expect
# this as the total number of objects is large and the number being drawn is
# relatively small thus the binomial distribution of this case is a good
# approximation for the hypergeometric distribution
```

Question 2:

```
#2a)
# I think the distribution of Xa is a binomial distribution with N = 3 and
# probability of success p = 1 - pa
# I think the distribution of Xb is a binomial distribution with N = 6 and
# probability of success p = 1 - pb
```

```
#2b)i)
set.seed(20870815)
N <- 10000
all_sampleAb <- rbinom(N, 3, 0.9)
head(all_sampleAb)
```

```
3 1 2 2 2 3
```

```
all_sampleBb <- rbinom(N, 6, 0.8)
head(all_sampleBb)
```

```
4 5 5 5 4 5
```

```
#2b)ii)
my_count_A <- 0
for (i in all_sampleAb) {
  if (i >= 2) {
    my_count_A <- my_count_A + 1
  }
}
```

```
> my_count_A
[1] 9708
```

```
my_count_B <- 0
for (i in all_sampleBb) {
  if (i >= 3) {
    my_count_B <- my_count_B + 1
  }
}
```

```
> my_count_B
[1] 9809
```

```
# Since the number of days out of 10000 that Plant B passes the manufacturing
# requirements (9809) is more than that of Plant A (9708), Pavlov should choose
# Plant B
```

```
#2c)i)
set.seed(20870815)
N <- 10000
all_sampleAc <- rbinom(N, 3, 0.1)
head(all_sampleAc)
```

0 2 1 1 1 0

```
all_sampleBc <- rbinom(N, 6, 0.2)
head(all_sampleBc)
```

2 1 1 1 2 1

```
#2c)ii)
my_count_Ac <- 0
for (i in all_sampleAc) {
  if (i >= 2) {
    my_count_Ac <- my_count_Ac + 1
  }
}
```

```
> my_count_Ac
[1] 292
```

```
my_count_Bc <- 0
for (i in all_sampleBc) {
  if (i >= 3) {
    my_count_Bc <- my_count_Bc + 1
  }
}
```

```
> my_count_Bc
[1] 1005
```

```
# Since the number of days out of 10000 that Plant B passes the manufacturing
# requirements (1005) is more than that of Plant A (292), Pavlov should choose
# Plant B
```

Q2d) Suppose now that $p = p_A = p_B$. For what values of $p \in (0,1)$ is plant B more likely to meet manufacturing requirements than plant A on a randomly selected day?

Find p such that $P(X_B \geqslant 3) > P(X_A \geqslant 2)$

In this case let $\bar{p}$ denote the probability a machine is operation

$P(X_B \geqslant 3) = P(X_B = 3) + P(X_B = 4) + P(X_B = 5) + P(X_B = 6)$

$= 6C3 \times \bar{p}^3 \times (1-\bar{p})^3 + 6C4 \times \bar{p}^4 \times (1-\bar{p})^2 + 6C5 \times \bar{p}^5 \times (1-\bar{p}) + 6C6 \times \bar{p}^6 \times (1-\bar{p})$

$= 20\bar{p}^3(1-\bar{p})^3 + 15\bar{p}^4(1-\bar{p})^2 + 6\bar{p}^5(1-\bar{p}) + \bar{p}^6$

$P(X_A \geqslant 2) = P(X_A = 2) + P(X_A = 3)$

$= 3C2 \times \bar{p}^2 \times (1-\bar{p})^1 + 3C3 \times \bar{p}^3 \times (1-\bar{p})^0$

$20\bar{p}^3(1-\bar{p})^3 + 15\bar{p}^4(1-\bar{p})^2 + 6\bar{p}^5(1-\bar{p}) + \bar{p}^6 > 3\bar{p}^2(1-\bar{p}) + \bar{p}^3$

$\Rightarrow 20\bar{p}(1-\bar{p})^3 + 15\bar{p}^2(1-\bar{p})^2 + 6\bar{p}^3(1-\bar{p}) + \bar{p}^4 > 3(1-\bar{p}) + \bar{p}$

$\Rightarrow 20\bar{p}(1-3\bar{p}+3\bar{p}^2-\bar{p}^3) + 15\bar{p}^2(1-2\bar{p}+\bar{p}^2) + 6\bar{p}^3(1-\bar{p}) + \bar{p}^4 > 3 - 3\bar{p} + \bar{p}$

$\Rightarrow 20\bar{p} - 60\bar{p}^2 + 60\bar{p}^3 - 20\bar{p}^4 + 15\bar{p}^2 - 30\bar{p}^3 + 15\bar{p}^4 + 6\bar{p}^3 - 6\bar{p}^4 + \bar{p}^4 > 3 - 2\bar{p}$

$\Rightarrow -10\bar{p}^4 + 36\bar{p}^3 - 45\bar{p}^2 + 22\bar{p} - 3 > 0$

Note that for this polynomial, $\bar{p} = 1$ gives

$-10(1) + 36(1) - 45(1) + 22(1) - 3 = 0$

Thus $\bar{p} = 1$ is a root and so $(\bar{p} - 1)$ is a factor.

So we will proceed with polynomial division.

$$
\begin{array}{r|l}
 & -10\bar{p}^3+26\bar{p}^2-19\bar{p}+3 \\
\hline
\bar{p}-1 & -10\cancel{\bar{p}^4}+36\bar{p}^3-45\bar{p}^2+22\bar{p}-3 \\
 & \cancel{10}\bar{p}^4-10\bar{p}^3 \\
\hline
 & 26\cancel{\bar{p}^3}-45\bar{p}^2+22\bar{p}-3 \\
 & -\cancel{26}\bar{p}^3+26\bar{p}^2 \\
\hline
 & -19\bar{p}^2+22\bar{p}-3 \\
 & 19\bar{p}^2+19\bar{p} \\
\hline
 & 3\bar{p}-3 \\
 & -3\bar{p}+3 \\
\hline
 & 0
\end{array}
$$

Thus our polynomial is of the form:

$$(\bar{p}-1)(-10\bar{p}^3+26\bar{p}^2-19\bar{p}+3)$$

Once again we have that for $\bar{p}=1$:

$$-10(1)+26(1)-19(1)+3=0$$

Thus $p=1$ is root of the above qutient.

$\therefore$ $(\bar{p}-1)$ is a root.

Once more we proceed with polynomial division.

$$
\begin{array}{r|l}
 & -10\bar{p}^2+16\bar{p}-3 \\
\hline
\bar{p}-1 & -10\bar{p}^3+26\bar{p}^2-19\bar{p}+3 \\
 & 10\bar{p}^3-10\bar{p}^2 \\
\hline
 & 16\bar{p}^2-19\bar{p}+3 \\
 & -16\bar{p}^2+16\bar{p} \\
\hline
 & -3\bar{p}+3 \\
 & 3\bar{p}-3 \\
\hline
 & 0
\end{array}
$$

Thus our polynomial of degree 4 is of the form

$$(\bar{p}-1)^2(-10\bar{p}^2+16\bar{p}-3)$$

The two other roots of our polynomial are –

$$\frac{-16+\sqrt{16^2-4(-10)(-3)}}{2(-10)} = \frac{8-\sqrt{34}}{10} = 0.2169$$

$$\frac{-16-\sqrt{16^2-4(-10)(-3)}}{2(-10)} = \frac{8+\sqrt{34}}{10} = 1.3831$$

Thus for $\bar{p}$ (probability that a machine is operational) such that $\bar{p} \in \left(\frac{8-\sqrt{34}}{10}, 1\right)$, plant B is more likely to meet requirements.

Thus for $p_a = p_B = p$ where p is the probability that a machine will breakdown,

$$p \in \left(0,\ 1 - \frac{8-\sqrt{34}}{10}\right)$$

will make plant B more likely to meet requirements

$$\lim_{N\to\infty} P(X=1)$$

As N is getting infinitely larger and thus $p = \frac{1}{N}$ is getting infinitely smaller, we can use the Poisson Distribution to approximate it.

$$P(x) = \frac{\mu^{x} \times e^{-\mu}}{x!}$$

* $\mu = np = N \times \frac{1}{N} = 1$
* $x! = 1! = 1$

so we have:

$$\frac{1^1 e^{-1}}{1} = \frac{1}{e} \Rightarrow \lim_{N\to\infty} P(X=1) = 1/e$$