# Formative Assessment 6

Vera Aguila

2024-02-28

I. **Geometric Distribution.** Provide an R code for the geometric distribution. The geometric distribution is a probability distribution that models the number of trials required to achieve the first success in a sequence of Bernoulli trials, where each trial has a constant probability of success.

1. Set the probability of success: p <- 0.2

```
p <- 0.2
```

2. Generate 1000 random variables from the geometric distribution.

```
n <- 1000
k <- rgeom(n, p)
```

3. Calculate some basic statistics:

mean_x <- mean(x)

var_x <- var(x)

sd_x <- sd(x)

```
tofirstsuccess <- k+1
mean_x <- mean(tofirstsuccess)
var_x <- var(tofirstsuccess)
sd_x <- sd(tofirstsuccess)
```

4. Print the results in item 3 with the following output (string):

Number of trials required to achieve first success:

Mean (in 2 decimal places):

Variance (in 2 decimal places):

Standard deviation ( in 2 decimal places):

```
## Number of trials required to achieve first success: 4 1 4 0 0 4 0 0 3 0 0 2 4 8 12 3 4 0 0 1 5 11 0 5 1 2 3 1
1 3 1 4 1 0 1 5 7 2 6 2 6 20 0 0 5 6 4 3 3 2 6 3 0 12 4 2 0 4 1 2 0 5 8 0 1 0 7 6 1 5 17 11 2 0 1 8 1 5 6 0 0 0 2
2 2 2 5 1 0 1 0 3 2 0 0 1 4 11 7 12 1 3 1 0 0 3 1 1 1 1 3 0 0 6 2 2 2 0 3 0 3 0 0 4 1 10 4 5 1 1 9 1 1 0 4 5 1
15 0 1 2 5 2 0 1 0 1 13 6 2 13 1 4 3 4 0 1 0 4 6 0 8 3 5 0 2 1 19 1 6 14 30 3 9 0 4 7 1 0 5 1 3 4 3 3 0 2 0 4 3 1
7 2 2 1 3 5 0 0 6 6 2 1 4 0 4 12 4 5 0 1 0 6 5 5 8 1 9 3 6 2 9 19 4 3 7 11 5 4 16 2 5 1 6 0 0 7 4 0 0 1 7 9 5 0
1 6 5 5 1 3 0 2 12 3 9 1 11 0 5 5 3 1 1 1 0 1 4 6 9 9 2 0 1 9 4 1 4 1 6 1 1 1 19 1 0 2 3 4 2 10 5 0 2 9 5 17 0 1
3 2 1 12 8 9 0 1 6 0 8 6 5 1 1 9 0 4 0 1 8 17 0 2 3 8 4 10 0 4 4 1 1 0 0 0 2 1 0 3 0 4 0 8 1 1 0 3 2 12 2 13
12 0 3 0 6 10 1 0 0 3 0 0 0 18 0 4 0 0 3 5 2 0 4 9 7 2 6 4 5 2 3 1 7 14 2 0 1 0 12 17 1 9 0 9 3 0 0 3 0 4 1 6 0 3
13 3 0 2 1 2 4 20 2 3 10 8 0 3 3 0 2 3 0 3 0 1 1 7 3 4 0 1 3 2 5 4 4 12 3 2 2 2 4 6 2 5 4 4 1 1 3 2 0 14 2 30 0 1
6 4 1 13 13 0 0 2 0 8 2 1 10 2 0 10 7 1 9 9 1 1 4 9 0 6 8 0 2 4 4 0 6 5 4 1 6 1 8 3 0 13 1 5 0 1 7 5 6 6 8 6
1 6 0 9 4 0 1 4 0 3 1 14 8 0 4 9 2 0 6 3 6 0 10 1 1 1 17 19 21 4 2 4 1 2 10 6 2 3 1 2 5 0 2 2 1 9 1 10 1 1 11 2 4
10 0 1 6 7 6 1 9 6 3 0 0 1 7 1 6 0 5 3 9 2 2 0 0 1 9 0 2 5 0 9 0 2 5 8 1 1 0 6 4 5 10 6 2 3 2 9 8 0 11 1 0 3 0 2
9 14 10 5 1 0 9 3 1 9 0 0 13 13 4 2 1 1 2 1 3 8 2 6 0 0 3 2 2 0 0 5 7 2 2 10 3 6 6 1 0 0 3 0 6 2 9 8 1 0 1 2 3 0 9
5 0 0 1 10 1 7 3 10 4 0 4 4 1 2 7 3 2 3 0 2 8 1 0 2 3 9 7 6 3 5 1 3 0 4 1 0 2 1 1 8 4 8 1 3 2 11 7 6 5 4 3 16 1 6
10 3 1 2 13 2 5 5 1 5 1 0 8 0 3 0 3 3 1 7 0 1 2 7 5 1 4 1 2 4 2 4 8 5 6 2 0 1 6 0 2 3 1 9 2 5 3 3 2 8 5 6 0 2 1 2
0 1 0 2 4 9 3 8 8 3 4 13 2 3 13 5 4 2 8 3 0 4 1 0 0 2 3 4 5 9 0 11 5 5 21 2 3 2 4 6 13 0 6 1 10 1 9 1 16 6 1 1 1
2 0 6 0 0 19 15 3 8 0 1 0 0 2 5 3 0 0 0 2 6 0 8 10 0 3 0 8 9 13 0 5 2 9 3 7 10 0 4 14 5 4 6 8 0 4 2 2 3 0 3 5 1 5
1 3 2 0 0 4 5 1 1 1 3 11 6 3 1 0 1 0 1 1 16 0 1 2 4 10 2 2 6 1 0 0 0 12 1 6 0 11 1 0 2 1 1 9 5 0 2 1 1 4 3 12 2 3
1 7 2 5 10 3 1 2 1 4 4 6 1 0 1 3 4 5 0 3 0 0 1 16 0 1 0 13 4 9 1 2 2 2 5 0 5 1 3 8 9 1 5 4 11 0 8 1 1 8 0
```

```
## Mean (in 2 decimal places): 4.84
```

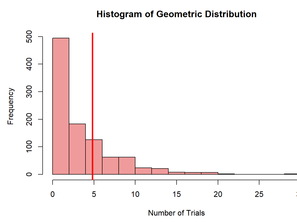
```
## Variance (in 2 decimal places): 17.37
```

```
## Standard deviation ( in 2 decimal places): 4.17
```

5. Plot the histogram of the results.

The red line represents the mean.

```
hist(k, breaks = 20, col = "#f29b9b", main = "Histogram of Geometric Distribution", xlab = "Number of Trials", yl
ab = "Frequency")
abline(v = mean_x, col='red', lwd = 3)
```




II. **Hypergeometric Distribution.** Consider a plant manufacturing IC chips of which 10% are expected to be defective. The chips are packed in boxes for export. Before transportation, a sample is drawn from each box. Estimate the probability that the sample contains more than 10% defectives, when:

proportion of successes = p = 0.10

1. A sample of 10 is selected from a box of 40;

total number = N = 40
number chosen = n = 10

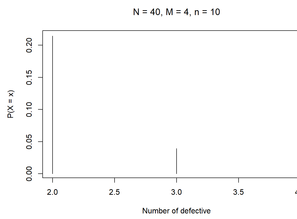
```
x <- 2:4
k <- dhyper(x, 4, 36, 10)
k
```

```
## [1] 0.214191925 0.039391618 0.002297844
```

```
sum(k)
```

```
## [1] 0.2558814
```

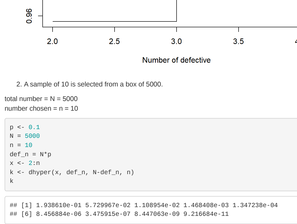
```
plot(x, dhyper(x, 4, 36, 10),
xlab = "Number of defective", type = "h",
ylab = "P(X = x)",
main = "N = 40, M = 4, n = 10", font.main = 1)
```




```
phyper(4, 4, 36, 10)
```

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

```
plot(x, phyper(x, 4, 36, 10),
xlab = "Number of defective", type = "s",
ylab = "P(X <= x)",
main = "N = 40, M = 4, n = 10", font.main =1)
```




2. A sample of 10 is selected from a box of 5000.

total number = N = 5000
number chosen = n = 10

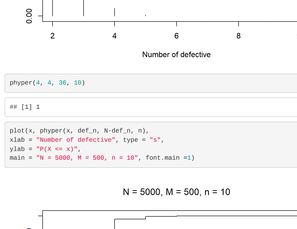
```
p <- 0.1
N = 5000
n = 10
def_n = N*p
x <- 2:n
k <- dhyper(x, def_n, N-def_n, n)
k
```

```
## [1] 1.938610e-01 5.729967e-02 1.108954e-02 1.468408e-03 1.347238e-04
## [6] 8.456884e-06 3.475915e-07 8.447063e-09 9.216684e-11
```

```
sum(k)
```

```
## [1] 0.2638622
```

```
plot(x, k,
xlab = "Number of defective", type = "h",
ylab = "P(X = x)",
main = "N = 5000, M = 4500, n = 10", font.main = 1)
```




```
phyper(4, 4, 36, 10)
```

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

```
plot(x, phyper(x, def_n, N-def_n, n),
xlab = "Number of defective", type = "s",
ylab = "P(X <= x)",
main = "N = 5000, M = 500, n = 10", font.main =1)
```