FORMATIVE ASSESSMENT 6

Lindsy Rossel C. Masicat

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
```

- 2. Generate 1000 random variables from the geometric distribution.

```
rand_var <- 1000

x <- rgeom(rand_var, p) + 1
x
```

```
##    [1]  4 16  1  4 14  1 10  4  7  2  3  2  5 21  1 21  1  1  3  8  1  2  9  3
##   [25]  1  7  1  5  2  4  3  5  2  1  5  9  5 12  9 12  1  4  2  6  6  8  5 14
##   [49] 13 16  4  3  1  1  6  2  2  1  2  1  9  3  4  1 12 11  6  6  3  3  2  3
##   [73]  1  8  4  3 11  1  2  2  4 16  1  4  4  2  4  6  1  4  7  4  1  7  1 12
##   [97]  5  3  3  1 10  2  1  1  6  3  3  1  6 11  2  1  5  1  3  1  7  2  5 11
##  [121]  8  4  9 10  6  6 12  1  4  5 10  7  9  5  2  5  1  2  3  2  9  2  4  2
##  [145]  7  1  3  1  1  6  8  1  7  7  3  3  6 12  3  1 28  2  1  2  4 11  4 14
##  [169]  2  2  2  1 10  3 14  7  9 10  9 16  2 19  4  5  5  5 25  5  4  1  3  3
##  [193] 17 16  8  4  7  2  1  1 19  8  2  4  1  2  2  8  2  4 13  4  2  6 14 10
##  [217]  1  4  2  1  3  7  8  2  3 11 13  3  3 17  3  4  7  5  2  2  2  4  4  3
##  [241]  1  2  2  2  3  8  3  1  3  5  5  3  6  5  2  8  3  5  1 17  2 10  2  4
##  [265]  2  1  2  2  1  1  2  6  9  2  3  7  4  5  2  2  5  7 11  9  3  1  5  2
##  [289]  3  4  3  2  1  5  2  2 11  3  5  4  2 10  4 15  2  2  2  5  3  4  2  2
##  [313]  4  9  6  3  5  5  5  2  1  6  8  3  3  1  4  1  5  7  9  1  3  1  4 11
##  [337]  8  4 13 11  1  5  7 23  2  1 14 10  6 16 11  2  2  6  3  4  6  1  8  3
##  [361]  1  1  8  7  5  7  2  2  5  2  2  2 25  4  8  3  5  2  2  1  8  2  1  8
##  [385]  2  5  4  6  2  2  6  3  3  5 16  5  4  1  5 15  4  1  7  2  1  5  1  1
##  [409]  6  4  9 11  1  2  1 15  7  1  3 14  3  7  3  5  2  1  1  6  5  5  1  1
##  [433]  4  4  1  3  7  3  1  2  1 34  3  1  4 14  3 11  2  7 13  6  2  8 12 10
##  [457] 10  1  1  5  6  1  6  5  1  4  3  4  1  4  1  2 11  6  9  1 22  6  2  3
##  [481]  2  1  1  1  3  2  2  5  4  4  1  8  3  1  8  4  4  6  3  2  3  6  2  1
##  [505]  3 26  1 14  4  4  6  5  4  1  4  9  3  2  2  2  5  3  9  7  2  5  2  1
##  [529]  5  7  2 19  1  3 11  6 32  2  5  3  3 10  3  2 12  2 10  1  5  7  6  7
##  [553]  5  6  5  3 24  6 18 10  8  1  8  3 18  2  6 11  1  3  1  1  1  8  3  3
##  [577]  4 15  1  1  1  2  3  2  3  3  7  1  9  2 12  4  8  3  4 13 15  7  5  2
##  [601]  2  4  2  9  1  4  7  2  6  8  7  8  1  3 12  1 10  1  2 13  8  1  3  2
##  [625]  6  2  6 16  1 13  1  3  6  8  5  5  8  1  2  3  5  1  1  3  5  5  2  1
##  [649]  2  8  1  4  3  5 15  8  1  1 12  4  9  5  8  6  5  1  8  1  2  7  2  4
##  [673]  4  7 14  1  7  4  6  1  5 11  2  2  2 11  7  1  7  3  3 10  4  4  1 14
##  [697]  2  5  3  4  3  1 12 10  3  2  1  6  6  9  6  2 11  2  5  5  8  2  4  1
##  [721]  5  9  5  7  1  3  2  2  1 28  4  5 10  5  2  6  7  6  4  4  2  3  7  8
##  [745] 10  8  6  2  1  2  3  2 13  9 20 13  1  4  6  3  2  3  5  3  1  3  1  2
##  [769]  1  5  3  1  1  6  2  4 10 13  1  6 29  5  4  1  4  7  4  3  1 10  6  9
##  [793]  3  7  8  4 11  5  2  3  3 17  3  3  2  4  6 10  2  1  2  8  7  2  1  4
##  [817]  4  1  7  4  5  2  6  4 26  6  8 10 12  3  2  1  1  5  2  9  1  5  2  2
##  [841] 11  3  3  6  8 13  1  3  4  5  5  7  8  6  3 15  2  2  3  3  1  2  4  9
##  [865] 11  1  1  7  2  3  5  4  4  1  5  3  8 19  3  6  5  3  7  3  3  5  3  1
##  [889]  2  3  1  1  4  7 10  3  1  1  5  3  1  3  2  5  3  2  7  5  1  1  4  1
##  [913]  2  6 13  3  5  3  1  2  3  4  5  3  3  2  7  4  1 18  1  4  2  9  5  7
##  [937]  1  1  4 12  2 11  2  5  4  1  7  1  4  1  3  1  7  5 11  9  6  2  3  5
##  [961] 14  2  5  2  2  1  1  2  5 15  2  1  6  2  1  2  2  4 17  3  6  2  1  1
##  [985]  3  2  1  6  9  6 10  3  2  4  2  7  2  2  7  2
```

- 3. Calculate some basic statistics:
  mean_x = mean
  var_x = variance
  sd_x = standard deviation

```
mean_x <- mean(x)
var_x <- var(x)
sd_x <- sd(x)
```

- 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):

```
cat("Number of trials required to achieve first success: \n")
```

```
## Number of trials required to achieve first success:
```

```
cat("Mean (in 2 decimal places): ", round(mean_x, 2), "\n")
```

```
## Mean (in 2 decimal places):  5.04
```

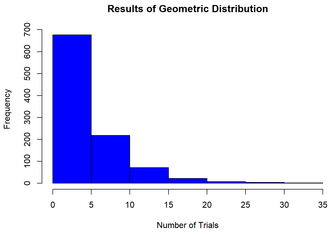
```
cat("Variance (in 2 decimal places): ", round(var_x, 2), "\n" )
```

```
## Variance (in 2 decimal places):  20.23
```

```
cat("Standard deviation (in 2 decimal places): ", round(sd_x, 2), "\n")
```

```
## Standard deviation (in 2 decimal places):  4.5
```

- Plot the histogram of the results.



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:

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

```
x <- 2:10
sample_drawn <- 10
sample_size <- 40
def_percent <- 0.1
num_def <- sample_size*def_percent

prob_1 <- dhyper(x, num_def, sample_size-num_def, sample_drawn)
prob_1
```

```
## [1] 0.214191925 0.039391618 0.002297844 0.000000000 0.000000000 0.000000000
## [7] 0.000000000 0.000000000 0.000000000 0.000000000
```

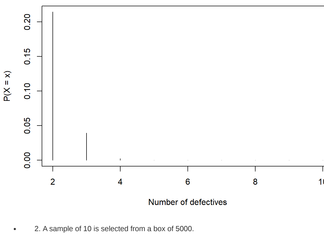
```
totalprob_1 <- sum(prob_1)

cat("The probability that the sample contains more than 10% defectives when 10 samples are drawn from a box of 40 is ", totalprob_1, "\n")
```

```
## The probability that the sample contains more than 10% defectives when 10 samples are drawn from a box of 40 i
## s  0.2558814
```

- Plot for sample of 10 selected from a box of 40.

```
plot(x, prob_1,
xlab = "Number of defectives", type = "h",
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.

```
x <- 2:10
sample_drawn <- 10
sample_size <- 5000
def_percent <- 0.1
num_def <- sample_size*def_percent

prob_2 <- dhyper(x, num_def, sample_size-num_def, sample_drawn)
prob_2
```

```
## [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
```

```
totalprob_2 <- sum(prob_2)

cat("The probability that the sample contains more than 10% defectives when 10 samples are drawn from a box of 50
00 is ", totalprob_2 , "\n")
```

```
## The probability that the sample contains more than 10% defectives when 10 samples are drawn from a box of 5000
## is  0.2638622
```

- Plot for sample of 10 selected from a box of 5000.

```
plot(x, prob_2,
xlab = "Number of defectives", type = "h",
ylab = "P(X = x)",
main = "N = 5000, M = 500, n = 10", font.main = 1)
```