Percia_FA6

2024-02-29

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.
3. Calculate some basic statistics:

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

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):

5. Plot the histogram of the results.

```
## I. Geometric Distribution
# Set the probability of success
p <- 0.2

# Generate 1000 random variables from the geometric distribution
x <- rgeom(1000, p)

# Calculate basic statistics
mean_x <- mean(x)
var_x <- var(x)
sd_x <- sd(x)

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

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

```
cat("Mean (in 2 decimal places): ", sprintf("%.2f", mean_x), "\n", sep="")
```

```
## Mean (in 2 decimal places): 3.83
```

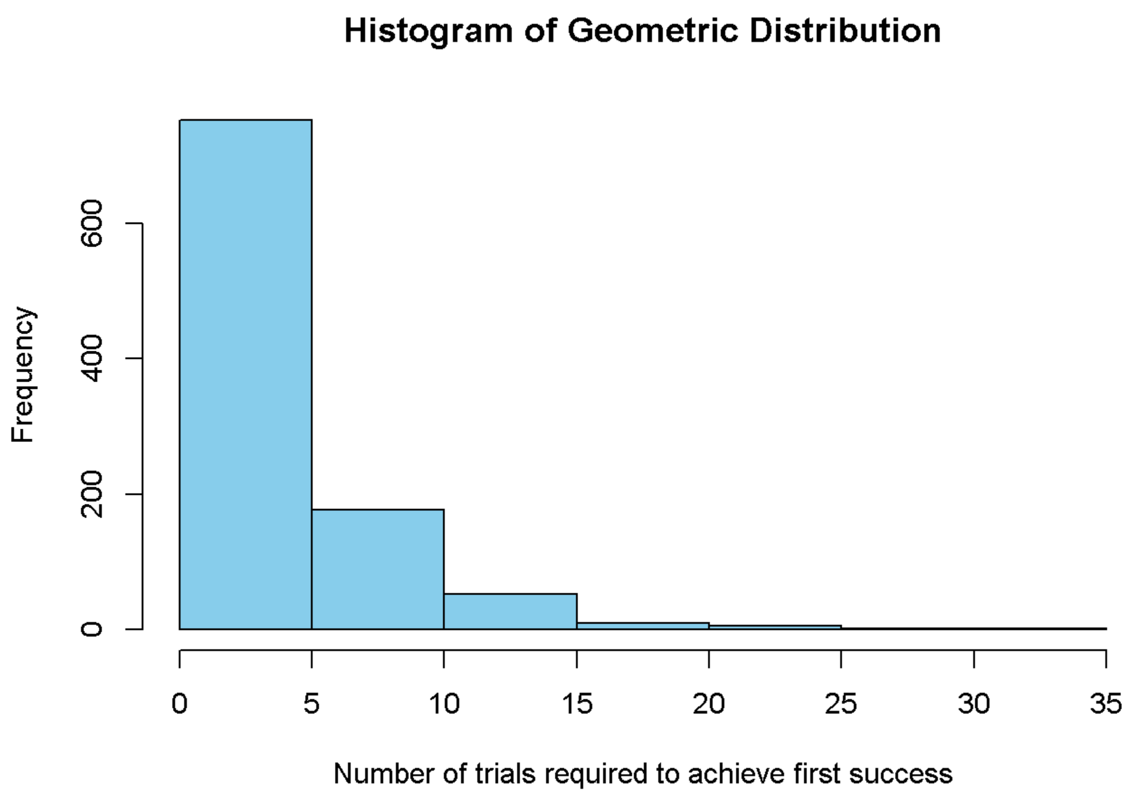
```
cat("Variance (in 2 decimal places): ", sprintf("%.2f", var_x), "\n", sep="")
```

```
## Variance (in 2 decimal places): 17.43
```

```
cat("Standard deviation (in 2 decimal places): ", sprintf("%.2f", sd_x), "\n", sep="")
```

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

```
# Plot the histogram
hist(x, main="Histogram of Geometric Distribution", xlab="Number of trials required to achieve first success", yl
ab="Frequency", col="skyblue")
```



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

```
## II. Hypergeometric Distribution

# (1)

# Parameters
sample_size <- 10
box_size <- 40
defectives_in_population <- 0.1 * box_size  # 10% of 40

# Calculate probability
probability_more_than_10_percent <- 1 - phyper(sample_size * 0.1 - 1, defectives_in_population, box_size - defect
ives_in_population, sample_size)
print(probability_more_than_10_percent)
```

```
## [1] 0.7001313
```

```
#(2)

# Parameters
sample_size <- 10
box_size <- 5000
defectives_in_population <- 0.1 * box_size  # 10% of 5000

# Calculate probability
probability_more_than_10_percent <- 1 - phyper(sample_size * 0.1 - 1, defectives_in_population, box_size - defect
ives_in_population, sample_size)
print(probability_more_than_10_percent)
```

```
## [1] 0.6516705
```

```
##PLOTTING

# Required libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

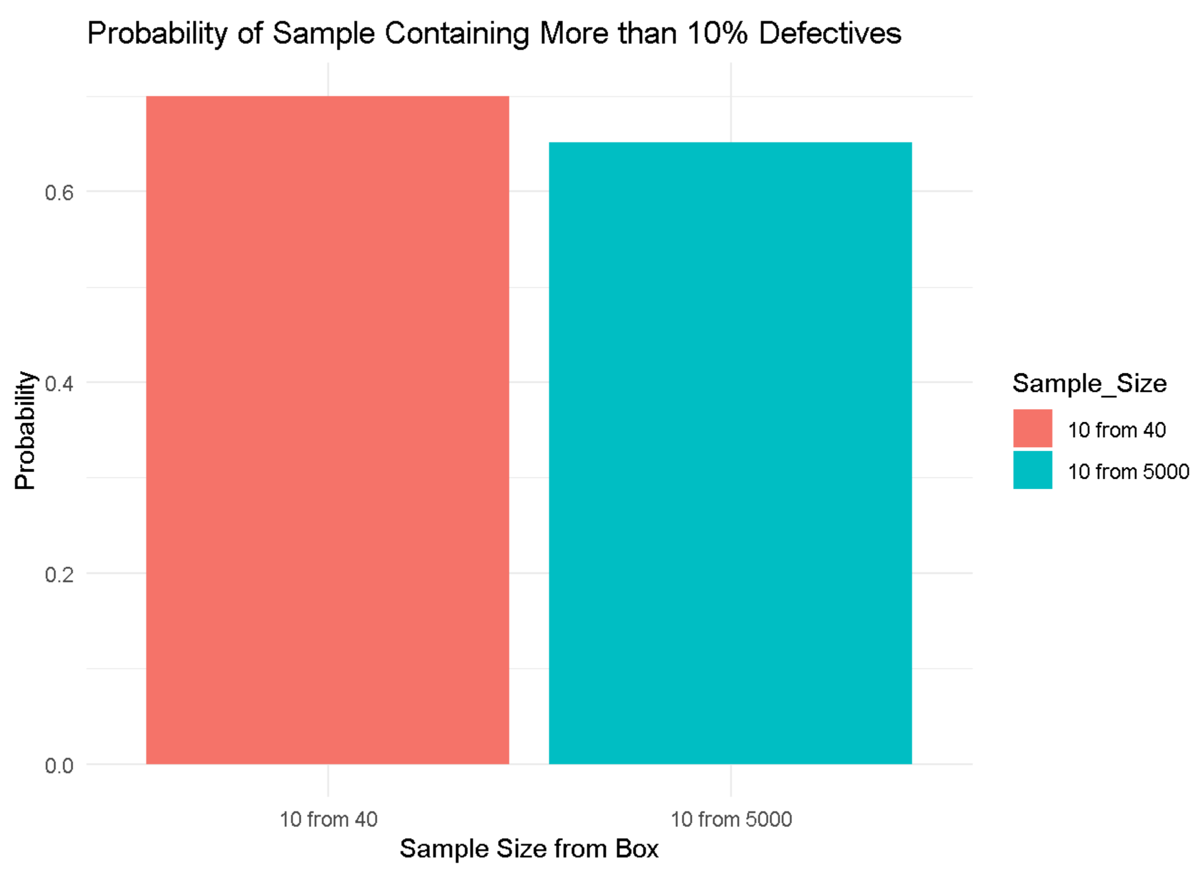
```
# Function to calculate probability
calculate_probability <- function(sample_size, box_size, defectives_percentage) {
  defectives_in_population <- defectives_percentage * box_size
  probability_more_than_10_percent <- 1 - phyper(sample_size * 0.1 - 1, defectives_in_population, box_size - defe
ctives_in_population, sample_size)
  return(probability_more_than_10_percent)
}

# Parameters
sample_sizes <- c(10, 10)
box_sizes <- c(40, 5000)
defectives_percentages <- c(0.1, 0.1)

# Calculate probabilities
probabilities <- sapply(1:length(sample_sizes), function(i) {
  calculate_probability(sample_sizes[i], box_sizes[i], defectives_percentages[i])
})

# Data frame for plotting
data <- data.frame(Sample_Size = rep(c("10 from 40", "10 from 5000"), each = 1),
                   Probability = probabilities)

# Plot
ggplot(data, aes(x = Sample_Size, y = Probability, fill = Sample_Size)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Probability of Sample Containing More than 10% Defectives",
       x = "Sample Size from Box",
       y = "Probability") +
  theme_minimal()
```



Interpretation

1. Hypergeometric Distribution (Probability of Sample Containing More than 10% Defectives):

- The hypergeometric distribution models the probability of obtaining a certain number of successes (defectives) in a sample drawn without replacement from a finite population (the box of IC chips).
- In both scenarios, the hypergeometric distribution indicates that there is a considerable probability of encountering a sample containing more than 10% defectives.
- Scenario I, where a sample of 10 chips is drawn from a box of 40, demonstrates a relatively high probability of finding a defective chip within a small sample due to the smaller population size and high defect rate.
- In Scenario II, despite the larger population size (5000 chips), the relatively high defect rate still results in a notable probability of finding a sample containing more than 10% defectives, albeit slightly lower compared to Scenario I.

2. Geometric Distribution (Number of Trials to Achieve First Success):

- The geometric distribution models the number of trials (selections) required to achieve the first success (finding a defective chip) in a sequence of Bernoulli trials (sampling chips from the box).
- Both scenarios suggest that it's likely to find a defective chip within a few trials, as indicated by the geometric distribution.
- Even with a larger population size in Scenario II, the geometric distribution implies that the probability of finding a defective chip within a few trials remains considerable due to the relatively high defect rate.