



Chapter 15: Normal, Binomial Distribution

Exercise 1: Buying T-shirt - Binominal Distribution

- Giả sử chúng ta biết xác suất khách mua áo thun từ cửa hàng là 0,3. Có 8 người trong cửa hàng đang xem hàng và họ không trao đổi gì với nhau.
- Vậy xác suất để hai người mua áo thun là bao nhiêu?
- Vậy xác suất để bảy người trong số họ mua áo thun là bao nhiêu?
- Xác suất để ít nhất 2 người mua áo thun là bao nhiêu?

Exercise 2: Chiều cao - Normal Distribution

- Chiều cao trung bình hiện nay của phụ nữ Việt Nam là 156cm, với độ lệch chuẩn là 4.6cm
- Hãy tạo ra các mẫu chiều cao lần lượt từ 140 cm -> 180cm, cách nhau 0.5 cm
- Hãy tính phân phối xác suất chiều cao của phụ nữ Việt Nam
- Vẽ histogram tương ứng
- Nếu chiều cao của phụ nữ là 1.6m thì xác suất là bao nhiêu?

Exercise 3: Tạo mẫu theo mean và sd - Normal Distribution

- Tạo ra 500 mẫu theo normal distribution với trung bình là 35 và độ lệch chuẩn là 0.1
- In 10 mẫu đầu tiên được tạo ra
- Vẽ histogram của bộ 500 mẫu này
- Tính lại mean và std của 500 mẫu này

Exercise 4: Binominal Distribution

Vấn đề 1: Trắc nghiệm

- Giả sử có mười hai câu hỏi trắc nghiệm trong một bài kiểm tra lớp tiếng Anh. Mỗi câu hỏi có năm câu trả lời, và chỉ một trong số đó là đúng.
- Tìm xác suất để có bốn câu trắc nghiệm trả lời đúng nếu một học sinh cố gắng trả lời ngẫu nhiên mọi câu hỏi.
- Tìm xác suất để có bốn câu trắc nghiệm trả lời đúng hoặc ít hơn nếu một học sinh cố gắng trả lời ngẫu nhiên mọi câu hỏi.

Vấn đề 2: Dùng thuốc

- Giả sử rằng 80% người trưởng thành bị dị ứng xác nhận có thể giảm triệu chứng bằng một loại thuốc cụ thể. Nếu thuốc được dùng cho 10 bệnh nhân mới bị dị ứng thì xác suất mà thuốc có hiệu quả cho chính xác bảy người là bao nhiêu?

Vấn đề 3: Bệnh đau tim

- Khả năng một bệnh nhân bị đau tim chết vì lên cơn đau tim là 0,04 (tức là cứ 100 người thì có 4 người chết).
- Giả sử chúng ta có 5 bệnh nhân bị đau tim, xác suất tất cả cùng sống sót là bao nhiêu?

Exercise 5: Thi cuối kỳ - Normal Distribution

- Giả sử rằng điểm kiểm tra cuối kỳ môn tiếng Anh phù hợp với Normal Distribution. Với điểm kiểm tra trung bình là 72 và độ lệch chuẩn là 15.2.
- Cho biết tỷ lệ học sinh đạt 84 điểm trở lên trong kỳ thi là bao nhiêu?
- Giả sử có 100 học sinh => hãy tạo ra các mẫu và vẽ histogram của các mẫu này.

Gợi ý:

Exercise 1: Buying T-shirt - Binominal Distribution

```
In [1]: # What is the probability of 2 of them buying a t-shirt?  $P(X = 2)$ 
p1 <- dbinom(2, size = 8, prob = 0.3)
print(p1)
# What is the probability of 7 of them buying a t-shirt?  $P(X = 7)$ 
p2 <- dbinom(7, size = 8, prob = 0.3)
print(p2)
# What is the probability that at least 2 of the customers buy a t-shirt?  $P(X \geq 2)$ 
p3 <- pbinom(1, size = 8, prob = 0.3, lower.tail = FALSE)
print(p3)
```

```
[1] 0.2964755
[1] 0.00122472
[1] 0.7447017
```

Exercise 2: Chiều cao - Normal Distribution

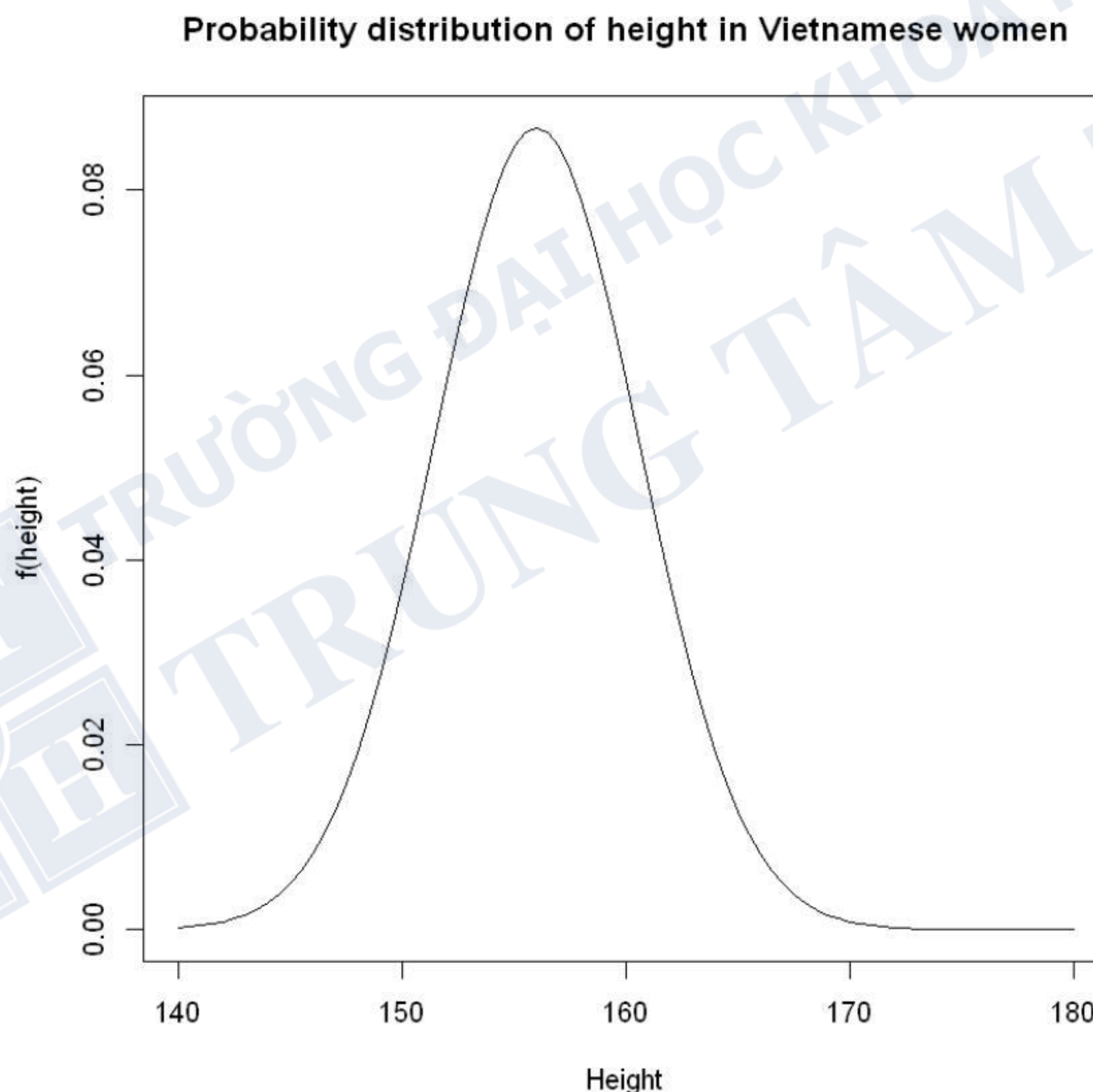




```
In [2]: height <- seq(140, 180, 0.5)
print("Samples: ")
print(height)
value <- dnorm(height, 156, 4.6)
```

```
[1] "Samples: "
[1] 140.0 140.5 141.0 141.5 142.0 142.5 143.0 143.5 144.0 144.5 145.0 145.5
[13] 146.0 146.5 147.0 147.5 148.0 148.5 149.0 149.5 150.0 150.5 151.0 151.5
[25] 152.0 152.5 153.0 153.5 154.0 154.5 155.0 155.5 156.0 156.5 157.0 157.5
[37] 158.0 158.5 159.0 159.5 160.0 160.5 161.0 161.5 162.0 162.5 163.0 163.5
[49] 164.0 164.5 165.0 165.5 166.0 166.5 167.0 167.5 168.0 168.5 169.0 169.5
[61] 170.0 170.5 171.0 171.5 172.0 172.5 173.0 173.5 174.0 174.5 175.0 175.5
[73] 176.0 176.5 177.0 177.5 178.0 178.5 179.0 179.5 180.0
```

```
In [3]: plot(height, value,
             type="l",
             ylab="f(height)",
             xlab="Height",
             main="Probability distribution of height in Vietnamese women")
```



```
In [4]: # => nu cao 1.6m thi xac suat se la
print("Probability of height = 160 cm")
print(dnorm(160, mean=156, sd = 4.6))
```

```
[1] "Probability of height = 160 cm"
[1] 0.05942343
```

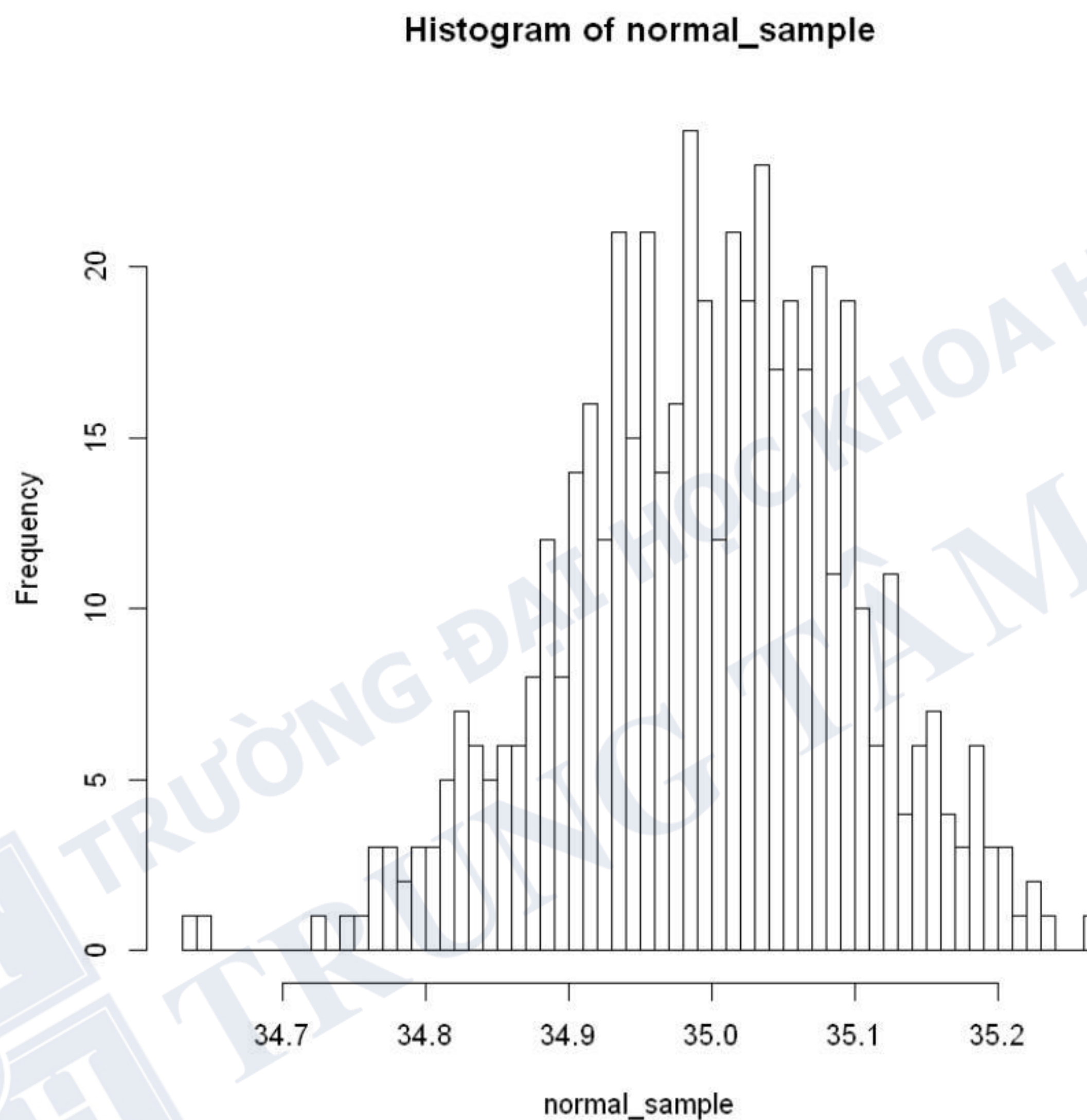
Exercise 3: Tạo mẫu theo mean và sd - Normal Distribution



```
In [5]: # Create a sample from normal distribution.  
# In this case, the first argument (500) means the sample size.  
normal_sample <- rnorm(500, mean = 35, sd = 0.1)  
print("10 of 500 samples:")  
print(normal_sample[1:10])
```

```
[1] "10 of 500 samples:"  
[1] 35.07887 34.83340 35.07103 35.04684 34.86194 34.81262 34.91434 34.91602  
[9] 34.87954 35.18348
```

```
In [6]: # Histogram of normal_sample  
hist(normal_sample, breaks = 50)
```



```
In [7]: # Calculate the mean and standard deviation of normal_sample  
mean_ns <- mean(normal_sample)  
print(paste("Mean:", mean_ns))  
sd_ns <- sd(normal_sample)  
print(paste("Sd:", sd_ns))
```

```
[1] "Mean: 34.9956021612918"  
[1] "Sd: 0.101465892233408"
```

Exercise 4: Binormal Distribution

Vấn đề 1: Trắc nghiệm


```
In [15]: # English Multiple Choices
p1 <- dbinom(4, size=12, prob=0.2) # having 5 choices => prob = 1/5 = 0.2
print("Probability of having four correct answers:")
print(p1)
```

```
[1] "Probability of having four correct answers:"
[1] 0.1328756
```

```
In [16]: # probability of having four or less correct answers
p2 <- pbinom(4, size=12, prob=0.2)
print("Probability of having four or less correct answers:")
print(p2)
```

```
[1] "Probability of having four or less correct answers:"
[1] 0.9274445
```

Vấn đề 2: Dùng thuốc

```
In [10]: # symptomatic relief
p3 <- dbinom(7, size=10, prob=0.8) # having 80%
print("probability that this medication is effective:")
print(p3)
```

```
[1] "probability that this medication is effective:"
[1] 0.2013266
```

Vấn đề 3: Bệnh đau tim

```
In [11]: # probability that all survive
p4 <- dbinom(0, size = 5, prob=0.04) # 0 person dies
print("probability that all survive:")
print(p4)
```

```
[1] "probability that all survive:"
[1] 0.8153727
```

Exercise 5: Thi cuối kỳ - Normal Distribution

- Giả sử rằng điểm kiểm tra cuối kỳ môn tiếng Anh phù hợp với Normal Distribution. Với điểm kiểm tra trung bình là 72 và độ lệch chuẩn là 15.2.
- Cho biết tỷ lệ học sinh đạt 84 điểm trở lên trong kỳ thi là bao nhiêu?
- Giả sử có 100 học sinh => hãy tạo ra các mẫu và vẽ histogram của các mẫu này.



```
In [12]: p1 <- pnorm(84, mean=72, sd=15.2, lower.tail=FALSE)
# lower.tail = FALSE because scoring higher than 84
print("percentage of students scoring 84 or more:")
print(p1)
```

```
[1] "percentage of students scoring 84 or more:"
[1] 0.2149176
```

```
In [13]: y <- rnorm(100, mean = 72, sd = 15.2)
print("Samples:")
print(y)
```

```
[1] "Samples:"
[1] 102.33561 65.60468 81.00817 65.92936 60.65557 101.84296 81.01092
[8] 43.22057 61.41912 101.44457 58.67254 82.27025 55.69274 77.93847
[15] 33.23416 77.15242 86.70350 54.17459 59.12522 75.42939 59.48972
[22] 50.44877 80.34405 88.15638 59.48791 60.66169 79.60534 72.86993
[29] 58.38543 86.41628 66.49856 48.48507 66.45320 90.83690 84.93536
[36] 90.25531 88.04111 68.35127 52.36957 78.61668 70.10469 88.71384
[43] 60.08993 57.63950 95.29893 77.22669 84.10596 91.80926 80.14703
[50] 80.84058 68.70491 85.55882 74.10527 83.64792 84.64275 78.54831
[57] 87.90771 68.89071 77.30575 79.47959 79.63994 96.65169 76.46846
[64] 92.77649 62.10460 67.90650 96.56519 93.75153 69.33361 72.02220
[71] 79.65199 67.72970 78.57242 92.46829 73.35105 33.47551 72.35855
[78] 70.07954 86.99874 49.63026 75.07945 40.48964 61.07996 74.14959
[85] 69.39322 86.08667 69.82757 55.21574 51.05279 51.54251 75.05063
[92] 85.38351 88.68532 90.80797 65.13284 74.48401 58.36210 86.14118
[99] 88.88529 80.24359
```

```
In [14]: hist(y, main = "Students' Scores" )
```

