Homework on Normal Distribution (#3)

Valerie Briot

February 28, 2016

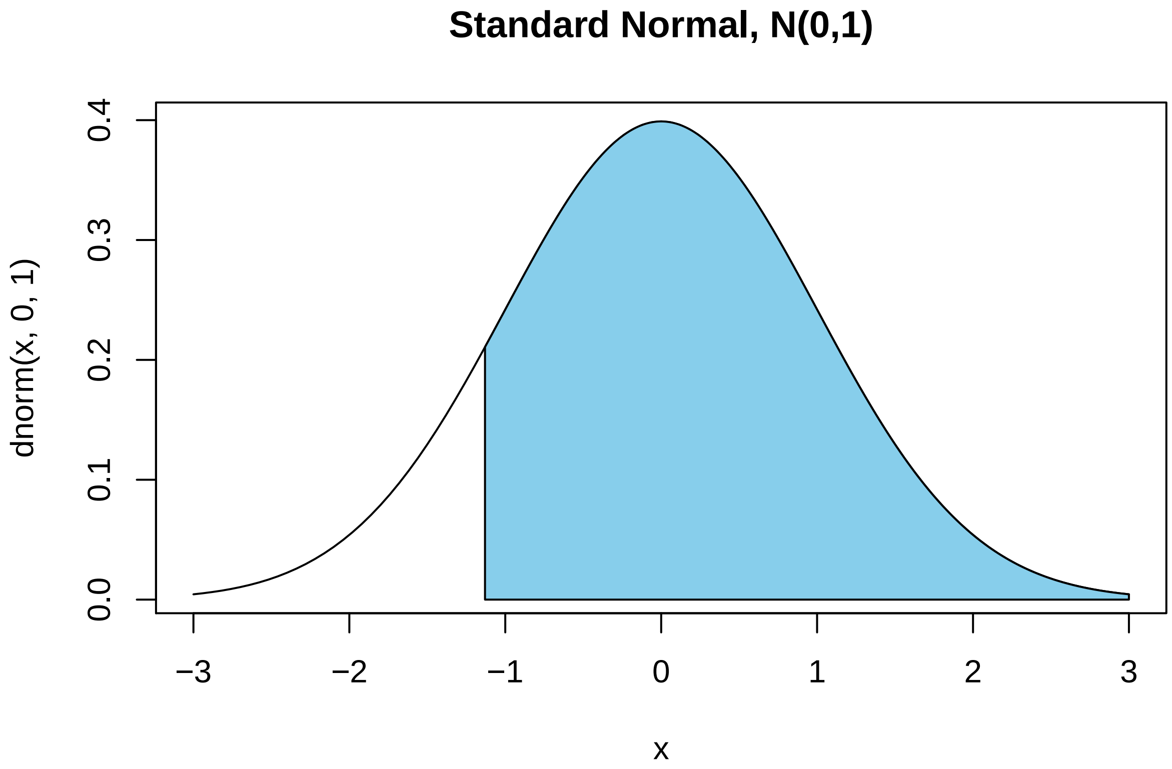
This document pertains to the homework assigned on the chapter on "Normal Distribution".

Exercise 3.2

Considering a standard normal distribution $N(\mu = 0, \sigma = 1)$, what percentage is found in each regions? we will plot a graph and then calcuate percentage using qwe will round to 2 decimal points for each answers.

a) Z > -1.13, First we will plot the graph:

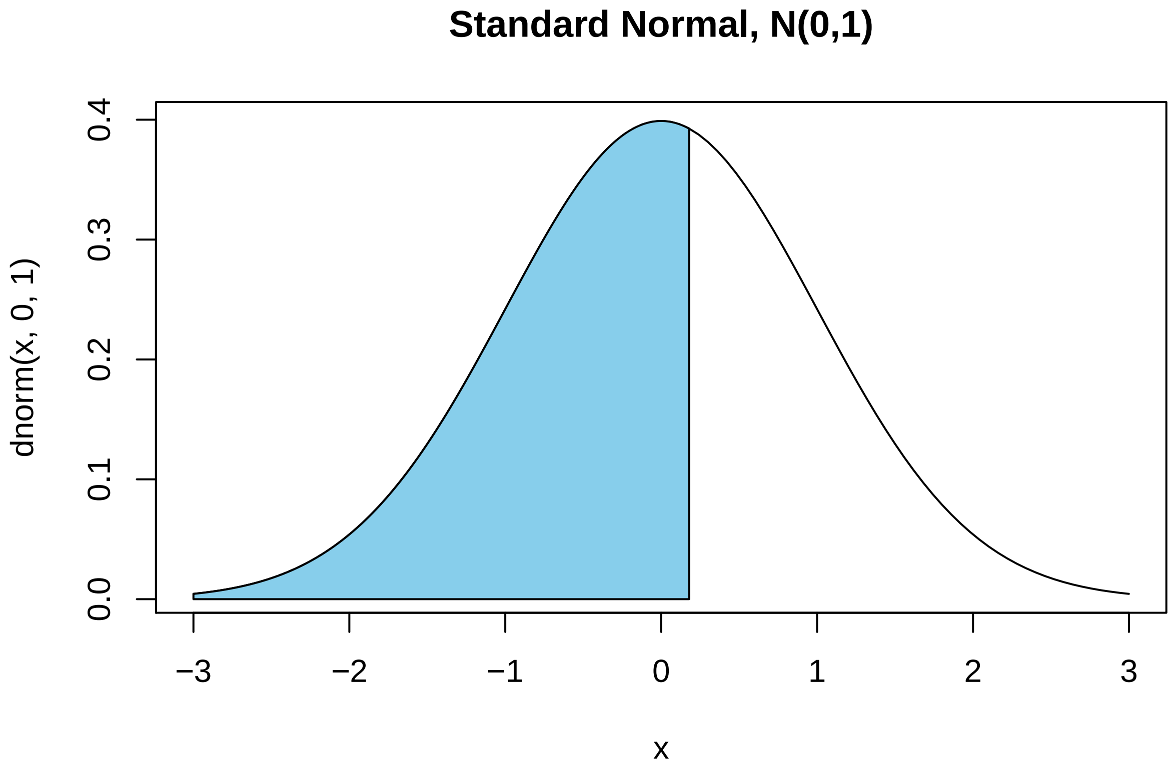
```r
curve(dnorm(x,0,1),xlim=c(-3,3),main="Standard Normal, N(0,1)")

cord.x <- c(-1.13, seq(-1.13,3,0.01),3)
cord.y <- c( 0, dnorm(seq(-1.13,3,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```



P(Z>-1.13) = 0.87

b) $Z < 0.18$

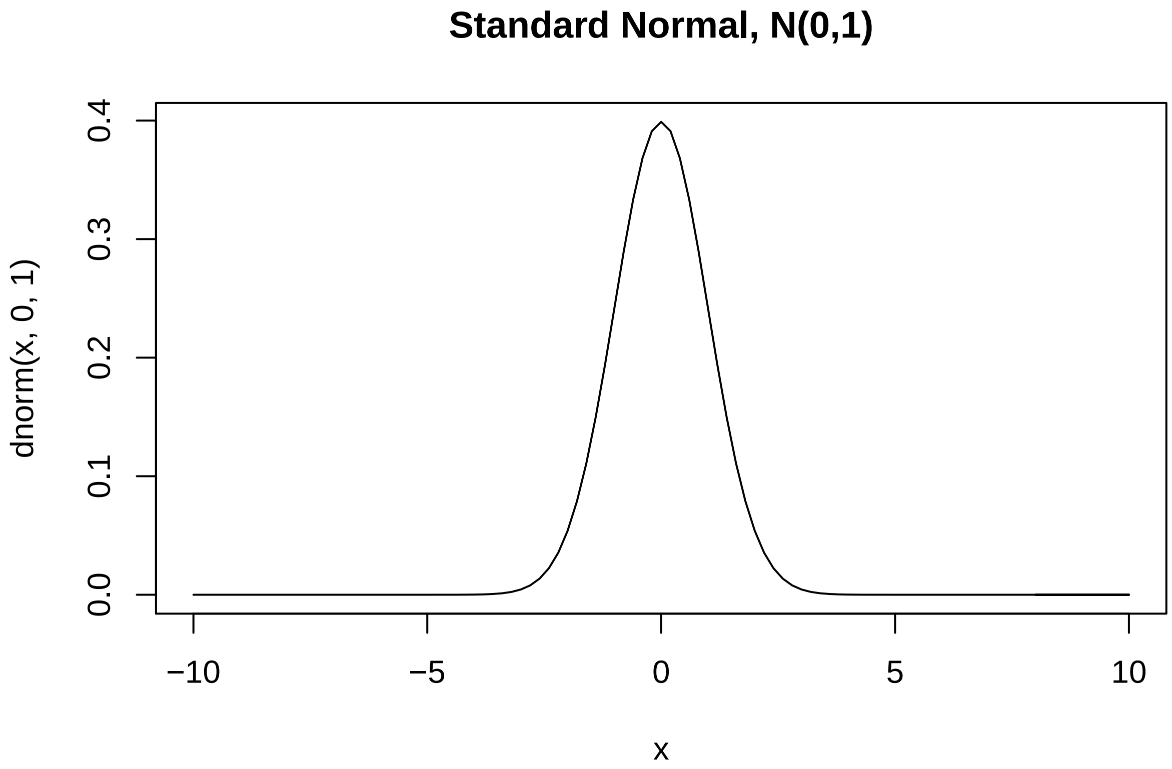
```
curve(dnorm(x,0,1),xlim=c(-3,3),main="Standard Normal, N(0,1)")

cord.x <- c(-3, seq(-3,0.18,0.01),0.18)
cord.y <- c( 0, dnorm(seq(-3,0.18,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```



$P(Z<0.18) = 0.57$

c) $Z > 8$

```
curve(dnorm(x,0,1),xlim=c(-10,10),main="Standard Normal, N(0,1)")

cord.x <- c(8, seq(8,10,0.01),10)
cord.y <- c(0, dnorm(seq(8,10,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```

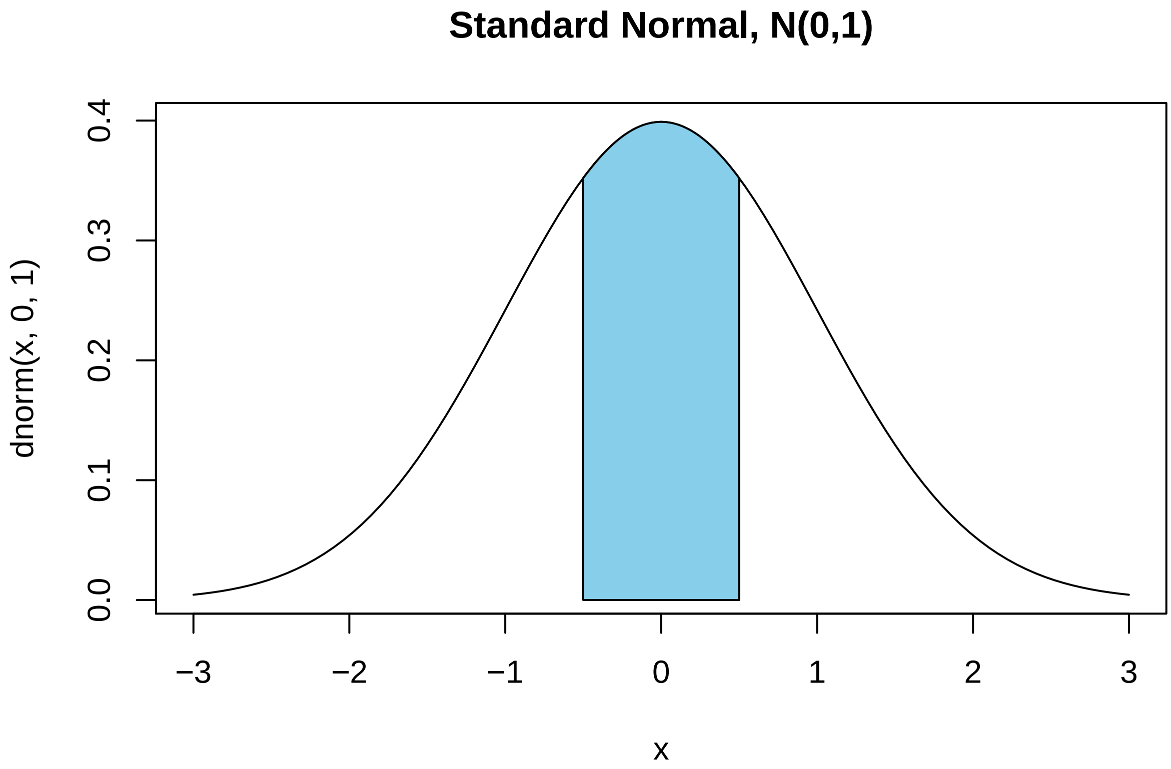


P(Z>8) = 0

This result is not surprizing since we are looking in an area under the curve 8 standard diviations above the mean.

d) |Z| < 0.5

```
curve(dnorm(x,0,1),xlim=c(-3,3),main="Standard Normal, N(0,1)")

cord.x <- c(-0.5, seq(-0.5,0.5,0.01),0.5)
cord.y <- c( 0, dnorm(seq(-0.5,0.5,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```



P(-0.5 < Z < 0.5) = 0.38

Exercise 3.4 - Triathlin times, Part I.

In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo completed in the “Men-Ages30-34” group while Mary competed in the “Women-ages25-29” group. Leo completed the race in 1:22:28 (4948 seconds) Mary completed the race in 1:31:53 (5513 seconds)

The following information is on the performance for Leo and Mary groups:

mean-m-30-34 = 3413, sd-m-30-34 = 583 mean-w-25-29 = 5261, sd-f-25-29 = 807

The distributions of finishing times for both groups is approximately Normal.

a) Write down the short-hand for these two normal distributions ***Male, 30-34*** $N\_m\_30\_34(\mu = 4313, \sigma = 583)$

Women, 25-29 $N\_w\_25\_29(\mu = 5261, \sigma = 807)$

b) What are the z-scores for Leo’s and Mary’s finishing times? What do the score tell us?

```
x_leo <- 4948    #Leo's finishing time
mu_m_30_34 <- 4313
sd_m_30_34 <- 583

x_mary <- 5513   #Mary's finishing time
mu_w_25_29 <- 5261
sd_w_25_29 <- 807

Z_score_f <- function(x, mu, sd){
# Function to calculate z-score
    if(sd != 0)
  {
    z_temp <- round((x - mu) / sd, 2)
  }else
  {
      z_temp <- NA
  }
  return(z_temp)
}

z_score_leo <- Z_score_f(x_leo, mu_m_30_34, sd_m_30_34)
z_score_mary <- Z_score_f(x_mary, mu_w_25_29, sd_w_25_29)
```

Leo's z-score = 1.09

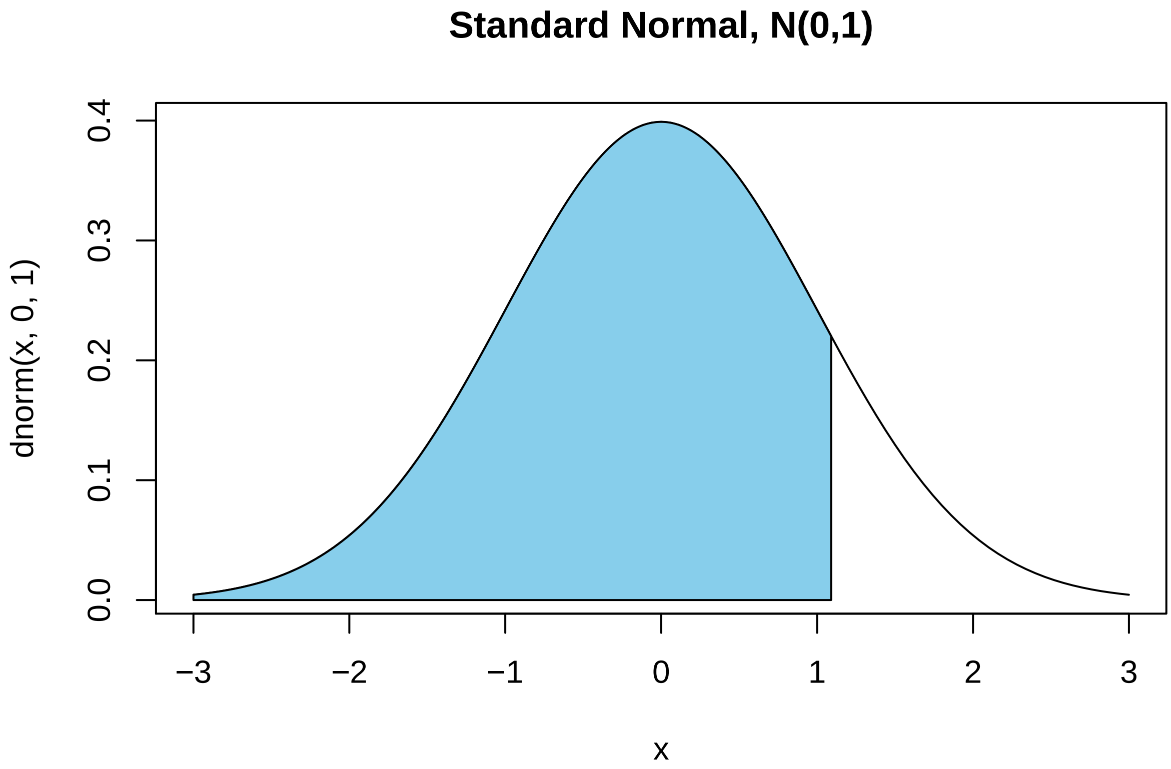
Mary's z-score = 0.31

The z-score is an indication of how far above or below a mean an observation is. In this case, Mary's finishing time is only 0.31 above the mean while Leo's time is 1.09 above the mean.

c) Did Leo or Mary rank better in their respective groups? Leo did better since his z-score is 1 standard deviation above the mean. This will mean that more participants finish worse times. Respectively, Mary finishing times place her only 0.31 standard deviation above the mean.

d) What percentage of triathletes did Leo finish faster than in his group?

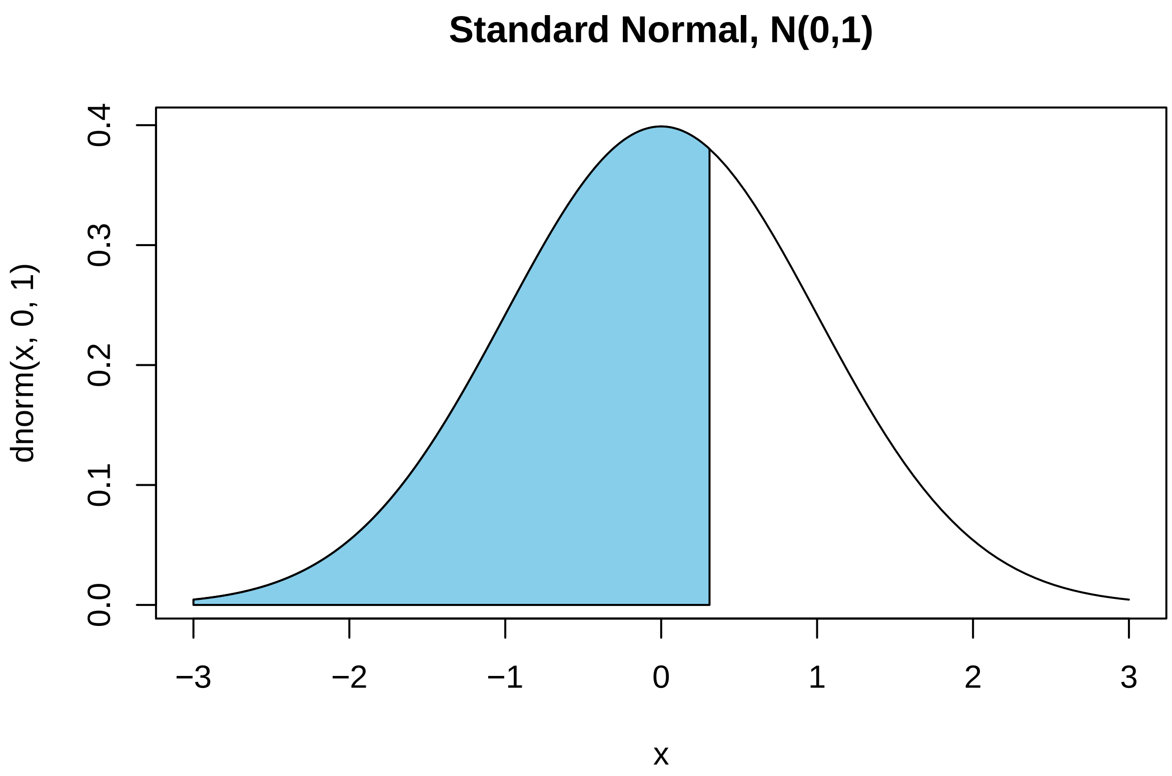
```
curve(dnorm(x,0,1),xlim=c(-3,3),main="Standard Normal, N(0,1)")

cord.x <- c(-3, seq(-3, z_score_leo,0.01),z_score_leo)
cord.y <- c( 0, dnorm(seq(-3, z_score_leo,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```



The percentage of triathletes did Leo finish faster than = 0.86

e) What percentage of triathletes Mary finished faster than in her group?

```
curve(dnorm(x,0,1),xlim=c(-3,3),main="Standard Normal, N(0,1)")

cord.x <- c(-3, seq(-3, z_score_mary,0.01),z_score_mary)
cord.y <- c( 0, dnorm(seq(-3, z_score_mary,0.01)),0)

polygon(cord.x,cord.y,col='skyblue')
```



The percenatge of triathletes Mary finished faster than in her group = 0.62

f) Would could evalute z-score but we could not made any conclusion from these as to percentage since require a normal distribution.

Exercise 3.18 - Heights of femail college students

```
heights_college_female <- c(54, 55, 56, 56, 57, 58, 58, 59, 60, 60, 60, 61, 61, 62, 62, 63, 63, 63, 64,

length(heights_college_female)
```

```
## [1] 25
```

```
mu_heights <- 61.52
st_heights <- 4.58
```

a) Does this data set follow the 68-95-99.7% rule approximately

First we calculate the percentage of observations within a range of 1 standard deviation from mean (below-above)

```
range_1st <- c(mu_heights-st_heights, mu_heights + st_heights)

ratio_1st <- round(((sum(heights_college_female >= range_1st[1] & heights_college_female < range_1st[2]
```

Percentage of observation within 1 standard deviation = 68

Next we calculate the percentage of observations within 2 standar deviations.

```
range_2nd <- c(mu_heights-(2*st_heights), mu_heights + (2*st_heights))

ratio_2nd <- round(((sum(heights_college_female >= range_2nd[1] & heights_college_female < range_2nd[2]
```

Percentage of observation within 2 standard deviations = 96

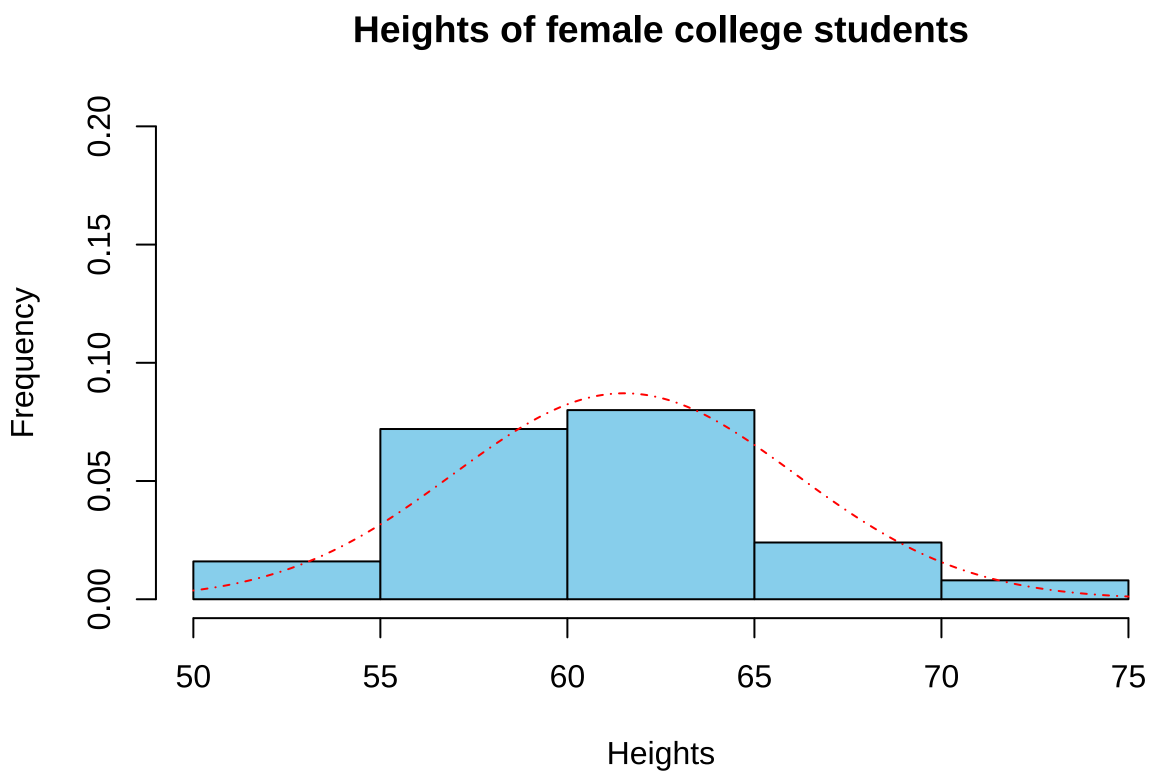
Finally, we calculate the percentage of observations within 3 standard deviations:

```
range_3rd <- c(61.52-(3*4.58), 61.52 + (3*4.58))

ratio_3rd <- round(((sum(heights_college_female >= range_3rd[1] & heights_college_female < range_3rd[2]
```

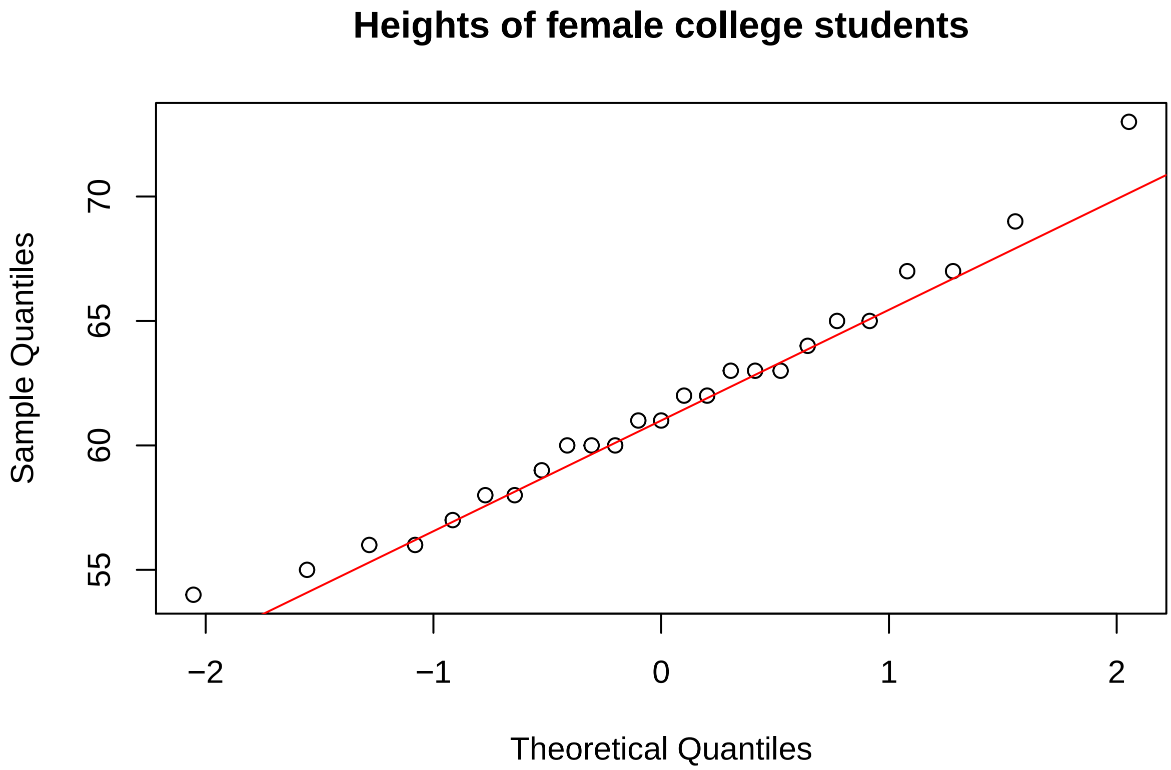
Percentage of observation within 2 standard deviations = 100

b) Do these data follow a normal distribution?

```
hist(heights_college_female, xlim = c(50, 75), ylim = c(0,0.2), breaks = seq(50, 75,5), xlab = "Heights
     main = "Heights of female college students", prob = TRUE, col = "skyblue")

x <- seq(50, 75, by = 0.1)
y <- dnorm(x, mu_heights, st_heights, log=FALSE)
points(x , y , mu_heights, st_heights, type = "l", col = "red")
```



```
qqnorm(heights_college_female, main="Heights of female college students")
qqline(heights_college_female, col = "red")
```



From these plots, the data set appears to have a normal distribution. From the histogram, the distribution is unimodal, appears symetrical. From the qq-plot, we can see that the observation are very cloe to the line.

Exercise 3.22 - Defective Rate

A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others.

a) What is the probability of the 10th transitor produced is the first with a defect?

```
p <- 0.02
```

The probability of the 10th transitor produced is a defect can be calculated as (1-p)*(1-p)....(1-p)*p, where there is 9 non-defective part. This scenario follows a geometric distribution. Hence, with n=10, p=0.02 $(1-p)^{n-1}p$

```
(1-p)^9 * p
```

```
## [1] 0.01667496
```

b) What is the probability that the machine produces no defective transistor in a batch of 100? This would mean that we would have (1-p) for 100th observation, hence $(1-p)^{100}$

```
(1-p)^100
```

```
## [1] 0.1326196
```

c) On average, how many transistors would you expect to be produced before the first with defect? and what would be the standard deviation?

Since we have a geometric distribution, average before 1st defect = mean given by 1/p and $\sigma = \sqrt{\frac{1-p}{p^2}}$

```
mu <- 1/p
sd <- ((1-p)/p^2)^.5
```

average = 50 sd = 49.4974747

d) Another machine that produces transistors has a 5% defective rage

```
p2 <- 0.05
mu2 <- 1/p2
sd2 <- ((1-p2)/p2^2)^.5
```

average = 20 sd = 19.4935887

e) Increasing the probability of an event reduces the mean and standard deviation of the wait time until success.

Exercise - Male Children

The probability of having a boy is 0.51. Suppose a couple plans to have 3 kids.

a) Use binomial model to calculate probability that 2 of them will be boys

```
p <- 0.51

n <- 3
k <- 2

# probability of 2 of 3 kids are boys.  Choosing 2 out of 3

choose(n,k)*p^k*(1-p)^(n-k)
```

```
## [1] 0.382347
```

b) We will write all the possibilities as permutation of 3 letters for example BBG (boy, boy, girl) Possibilities = GBB, BBG, BGB we would have 3 possibilities, which is same as result of choosing 2 out of 3.

for each possibility probability = $p^2(1-p)$

```
3*(p^2*(1-p))
```

```
## [1] 0.382347
```

Answer from part a) and b) are the same as expected.

c) Using the binomial model as in part a) is only a matter of using the formula and mapping the appropriate value. If we were to write all the combinations, as we have larger numbers it will become tedious very fast to write all possible combinations.

Exercise 3.42 - Serving in Volleyball

a player has 15% chance of making the serve. Suppose that the serves are independent of each others.

a) What is the probability that on the 10th trial she will make her 3rd successful serve?

This is model by a negative binomial model.

```
n<-10
k<-3
p<- 0.15

# negative binomial distribution

choose(n-1,k-1) * p^k * (1-p)^(n-k)
```

```
## [1] 0.03895012
```

b) suppose she has made 2 successful attemp in 9 attempts. What is the probability that the 10th serve will be successful? probability will be the one of making a successful serve = 0.15
c) The discrepency between a) and b) is that in b), we are given the condition that we have 2 successful serve already in the previous 9th. we are only considering the 10th serve.