Hw3_Random_variables

Shyam BV

September 25, 2016

```
library(utils)
library("combinat")
```

```
##
## Attaching package: 'combinat'
```

```
## The following object is masked from 'package:utils':
##
##     combn
```

```
library(ggplot2)
load("C:/CUNY/Courses/CUNY-repository/606/Week3-09_Sep2016/Lab3/more/bdims.RData")
```

3.2 Area under the curve, Part II. What percent of a standard normal distribution $N(\mu = 0, ! = 1)$ is found in each region? Be sure to draw a graph. (a) Z > −1.13 (b) Z < 0.18 (c) Z > 8 (d) |Z| < 0.5

```
#a

pa <- ggplot(data.frame(x = c(-3, 3)), aes(x = x)) +
  stat_function(fun = dnorm)

funcShaded <- function(x) {
  y <- dnorm(x, mean = 0, sd = 1)
  y[x < -1.13] <- NA
  return(y)
}
pa <- pa + stat_function(fun=funcShaded, geom="area", fill="#84CA72", alpha=0.2)
pa
```
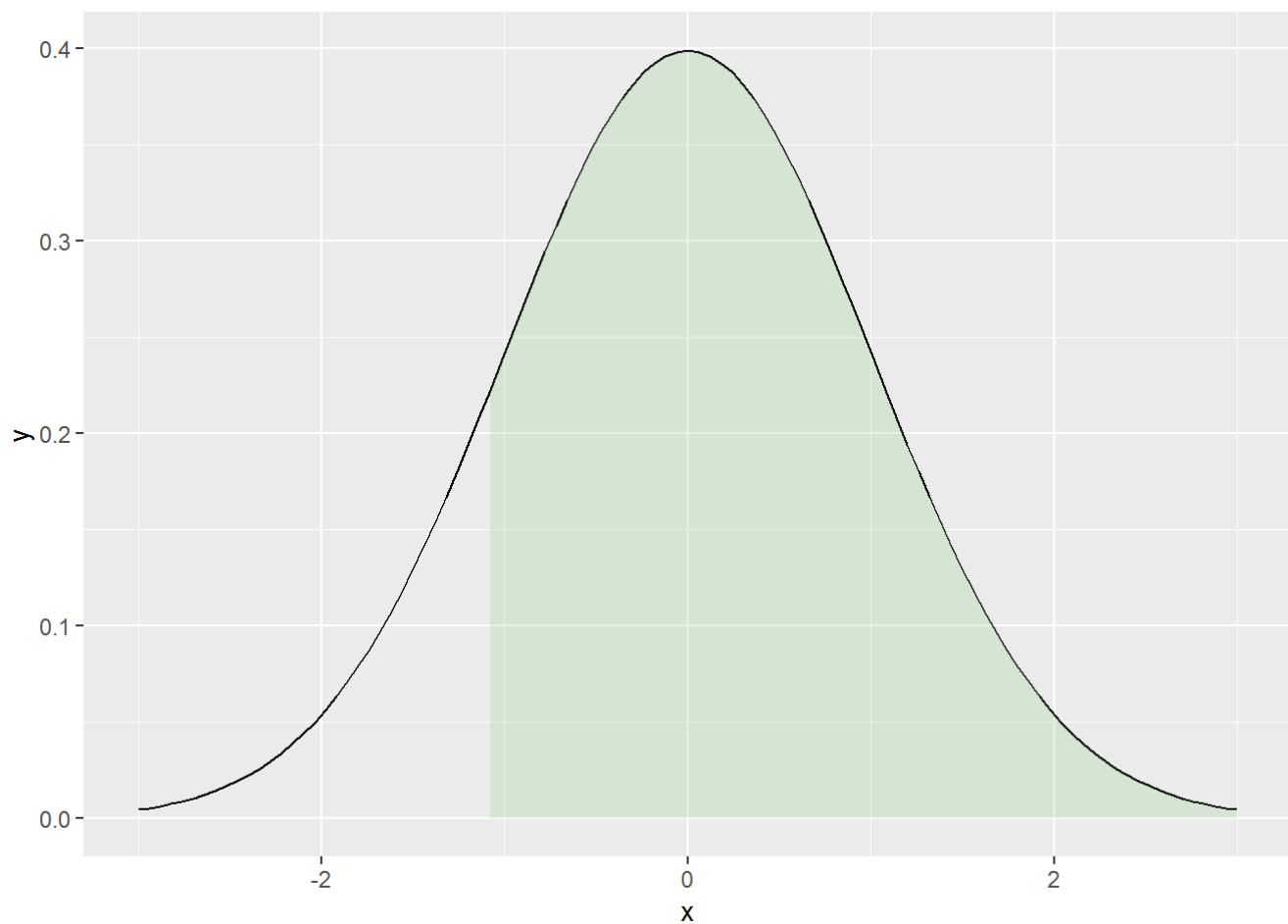
```
#Probability of the shaded area
1-pnorm(-1.13,0,1)
```
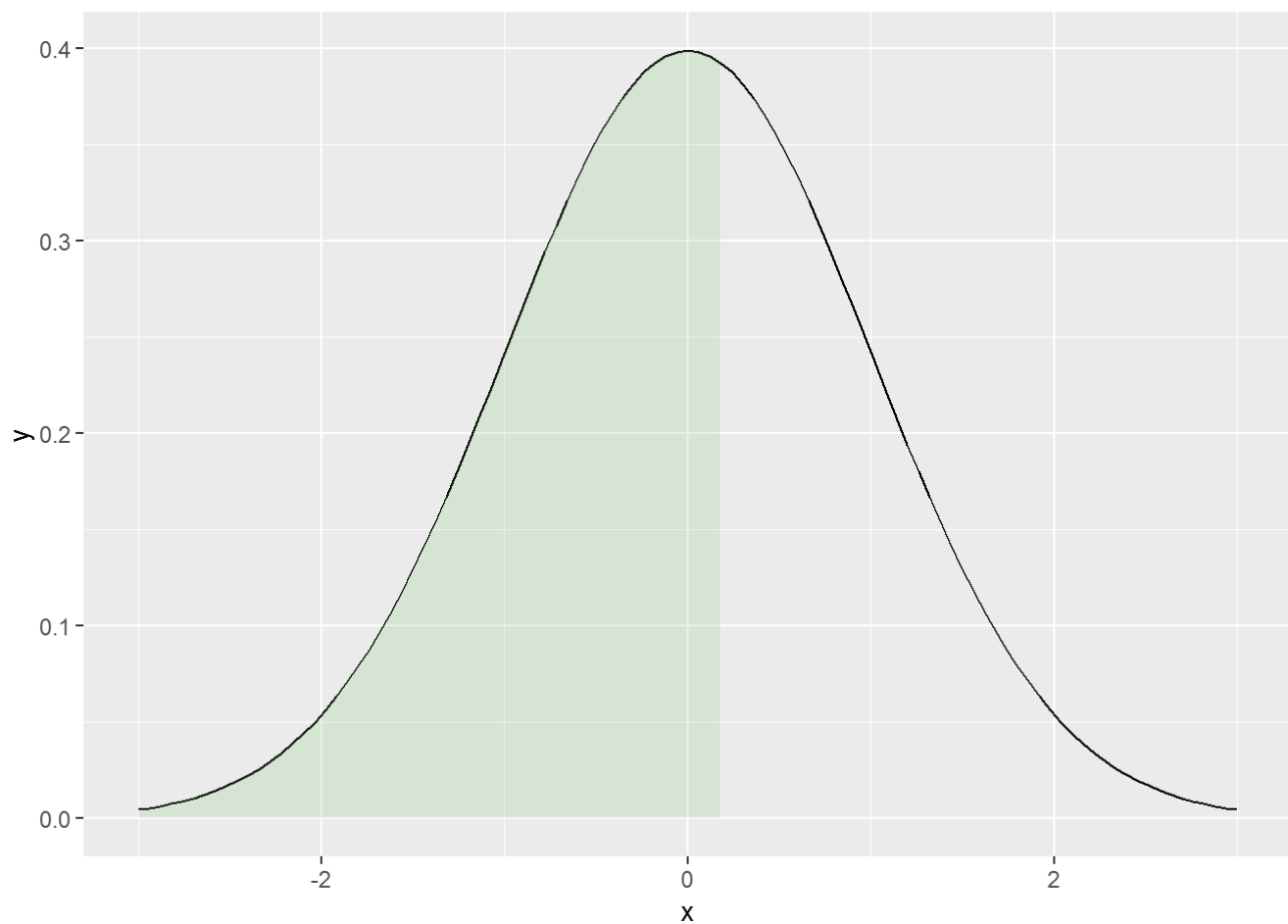
```
## [1] 0.8707619
```

```
#b <.18

pb <- ggplot(data.frame(x = c(-3, 3)), aes(x = x)) +
  stat_function(fun = dnorm)

funcShaded <- function(x) {
  y <- dnorm(x, mean = 0, sd = 1)
  y[x > .18] <- NA
  return(y)
}
pb <- pb + stat_function(fun=funcShaded, geom="area", fill="#84CA72", alpha=0.2)
pb
```
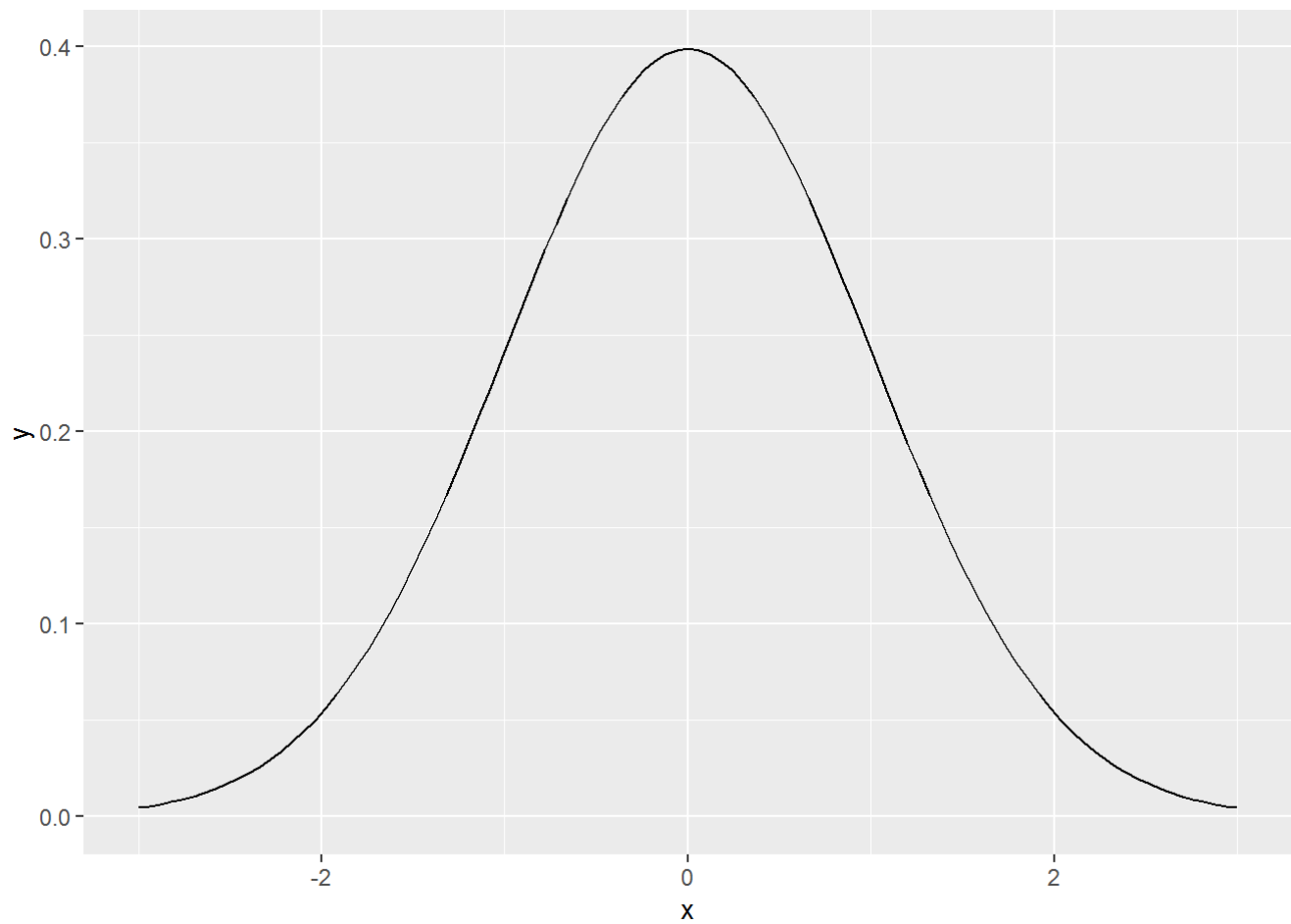
```
#Probability of the shaded area
pnorm(.18,0,1)
```

```
## [1] 0.5714237
```

```
#c - -0.4 < Z < 1.5

funcShaded <- function(x) {
  y <- dnorm(x, mean = 0, sd = 1)
  y[x < 8 ] <- NA
  return(y)
}

pc <- ggplot(data.frame(x = c(-3, 3)), aes(x = x)) +
  stat_function(fun = dnorm)+ stat_function(fun=funcShaded, geom="area", fill="#84CA72", alpha=0.2

pc
```

```
#Probability of the shaded area
pnorm(8,0,1)
```

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

```
#d |Z| >2

funcShaded <- function(x) {
  y <- dnorm(x, mean = 0, sd = 1)
  y[ x < 0.5] <- NA

  return(y)
}
funcShaded1 <- function(x) {
  y <- dnorm(x, mean = 0, sd = 1)
  y[ x > -.5] <- NA

  return(y)
}

pd <- ggplot(data.frame(x = c(-3, 3)), aes(x = x)) +
  stat_function(fun = dnorm)+ stat_function(fun=funcShaded, geom="area", fill="#84CA72", alpha=0.2
+
  stat_function(fun=funcShaded1, geom="area", fill="#84CA72", alpha=0.2)

pd
```
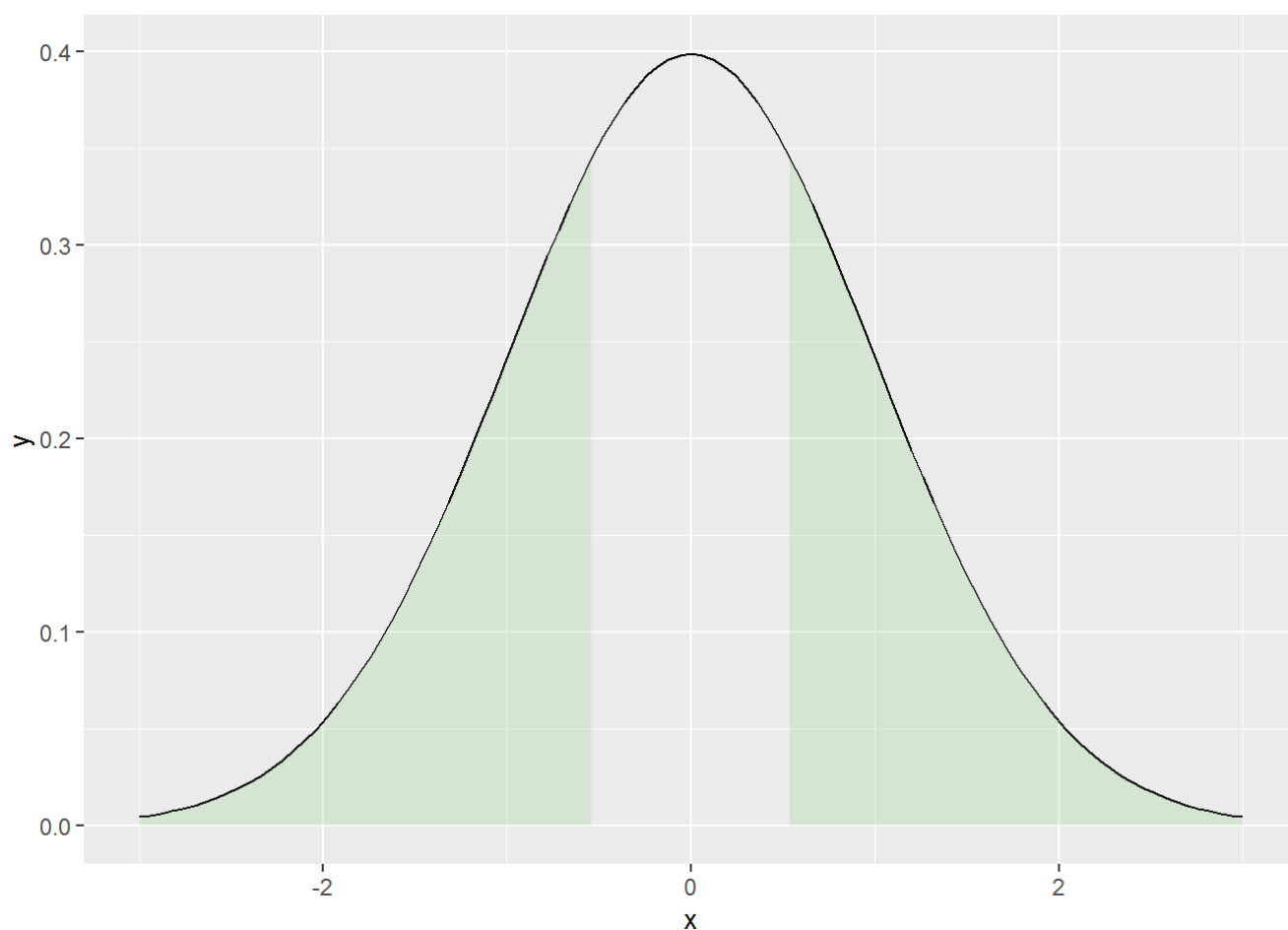


```
#Probability of the shaded area
pnorm(-.5,0,1) +(1-pnorm(.5,0,1))
```

```
## [1] 0.6170751
```

3.4 Triathlon 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 competed in the Men, Ages 30 - 34 group while Mary competed in the Women, Ages 25 - 29 group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups: • The finishing times of the Men, Ages 30 - 34 group has a mean of 4313 seconds with a standard deviation of 583 seconds. • The finishing times of the Women, Ages 25 - 29 group has a mean of 5261 seconds with a standard deviation of 807 seconds. • The distributions of finishing times for both groups are approximately Normal. Remember: a better performance corresponds to a faster finish.

```
leo <- 4948
menmean =4313
mensd=583

mary <- 5513
wmenmean =5261
wmensd=807
```

a. Write down the short-hand for these two normal distributions.

```
#Leo is in below percentile
pnorm(leo,menmean,mensd)
```

```
## [1] 0.8619658
```

```
#Mary is in below percentile
pnorm(mary,wmenmean,wmensd)
```

```
## [1] 0.6225814
```

b. What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?

```
#Leo is around 1.1 Standard deviations away from mean. So it is little bit away from the mean.
zleo <- (leo-menmean)/mensd

#Mary is around .31 Standard deviations away from mean. So it is near the mean.
zmary <- (mary-wmenmean)/wmensd
```

c. Did Leo or Mary rank better in their respective groups? Explain your reasoning.

```
#Leo Percentile
pnorm(leo,menmean,mensd)
```

```
## [1] 0.8619658
```

```
#Mary Percentile
pnorm(mary,wmenmean,wmensd)
```

```
## [1] 0.6225814
```

```
#It shows that Leo has better percentile than Mary.
```

d. What percent of the triathletes did Leo finish faster than in his group?

```
1-pnorm(leo,menmean,mensd)
```

```
## [1] 0.1380342
```

```
#Around 14 percentage finished faster than Leo.
```

e. What percent of the triathletes did Mary finish faster than in her group?

```
1-pnorm(mary,wmenmean,wmensd)
```

```
## [1] 0.3774186
```

```
#Around 38 percentage finished faster than Mary.
```

f. If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

No. If the distribution are not normal, the answer (b) and (e) will remain the same.

3.18 Heights of female college students. Below are heights of 25 female college students.

a. The mean height is 61.52 inches with a standard deviation of 4.58 inches. Use this information to determine if the heights approximately follow the 68-95-99.7% Rule.

```
femaleheights <- c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,65,67,67,69,73)

summary(femaleheights)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   54.00   58.00   61.00   61.52   64.00   73.00
```

```
meanfheight <- mean(femaleheights)

#Only
length(femaleheights)*.685
```

```
## [1] 17.125
```

```
#=
length(femaleheights[femaleheights < (meanfheight+sd(femaleheights)) & femaleheights > (meanfhei
ght-sd(femaleheights))])
```

```
## [1] 17
```

```
#Only
length(femaleheights)*.95
```

```
## [1] 23.75
```

```
#=
length(femaleheights[femaleheights < (meanfheight+sd(femaleheights)*2) & femaleheights > (meanfh
eight-sd(femaleheights)*2)])
```
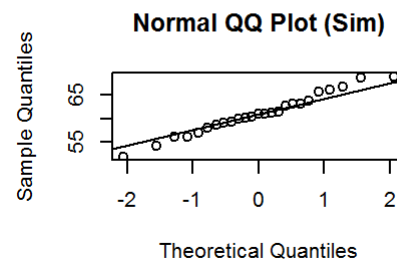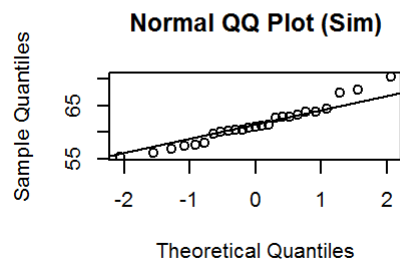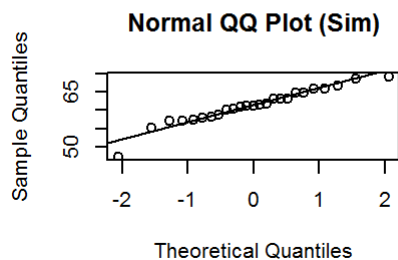
```
## [1] 24
```
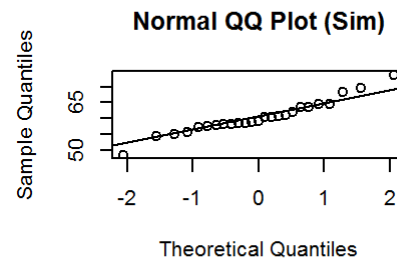
```
#Only
length(femaleheights)*.997
```

```
## [1] 24.925
```

```
#=
length(femaleheights[femaleheights < (meanfheight+sd(femaleheights)*3) & femaleheights > (meanfh
eight-sd(femaleheights)*3)])
```

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

b. Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.

```
qqnormsim(femaleheights)
```

Normal QQ Plot (Data)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

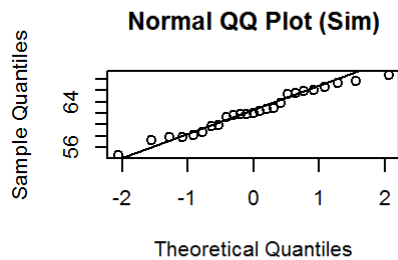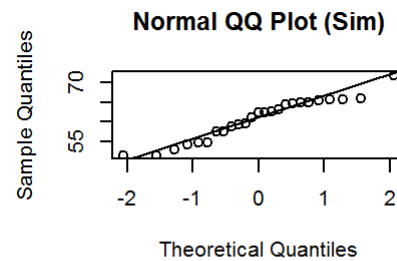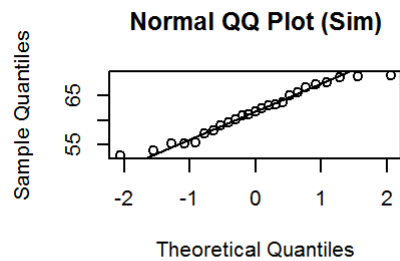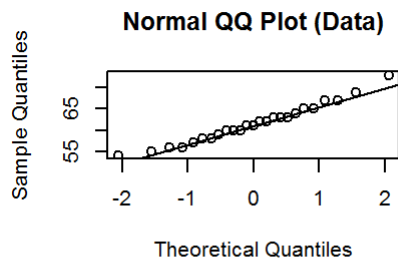Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

Normal QQ Plot (Sim)

Sample Quantiles

Theoretical Quantiles

```
hist(femaleheights,probability = TRUE )
lines(50:75,dnorm(50:75,mean(femaleheights),sd(femaleheights)), col="blue")
```

Histogram of femaleheights
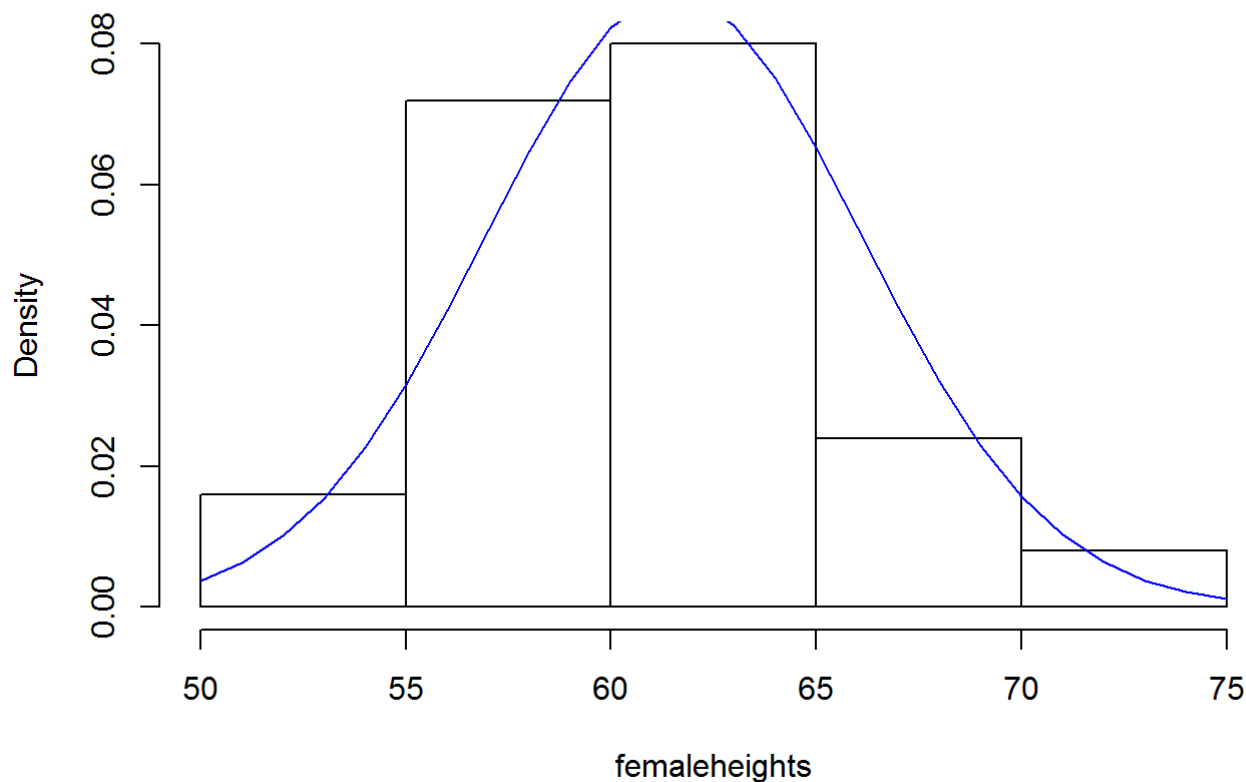

```
#Above charts shows that the heights are normally distributed. It matches with the randomly gene
rated normal distribution charts.
```

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 that the 10th transistor produced is the first with a defect?

```
(1 - .02)^9*.02
```

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

  b. What is the probability that the machine produces no defective transistors in a batch of 100?

```
(1 - .02)^100
```

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

  c. On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?

```
#mean. We would expect around 50 transistors
1/0.02
```

```
## [1] 50
```

```
#sd
sqrt((1-0.02)/0.02^2)
```

```
## [1] 49.49747
```

d. Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation?

```
#mean. We would expect around 20 transistors
1/0.05
```

```
## [1] 20
```

```
#sd
sqrt((1-0.05)/0.05^2)
```

```
## [1] 19.49359
```

e. Based on your answers to parts (c) and (d), how does increasing the probability of an event affect the mean and standard deviation of the wait time until success?

Increasing the probability of success, will inversely affect the mean and mean. So if the probability is 50%, then the mean and SD will be less. So we might expect failure very frequently.

3.38 Male children. While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

a. Use the binomial model to calculate the probability that two of them will be boys.

```
n=3
k=2
p=.51

#probability that two of them will be boys are

dbinom(k,n,p)
```

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

b. Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers

from parts (a) and (b) match.

```
# All possible outcomes

permn(c("boy","boy","girl"))
```

```
## [[1]]
## [1] "boy"  "boy"  "girl"
##
## [[2]]
## [1] "boy"  "girl" "boy"
##
## [[3]]
## [1] "girl" "boy"  "boy"
##
## [[4]]
## [1] "girl" "boy"  "boy"
##
## [[5]]
## [1] "boy"  "girl" "boy"
##
## [[6]]
## [1] "boy"  "boy"  "girl"
```

```
#As we are counting for six times, we need to multiply by 6. Then we have counted two times. So
 we need to divide by 2. Or unique outcomes
(p*p*(1-p))*(6/2)
```

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

c. If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly
describe why the approach from part (b) would be more tedious than the approach from part (a).

```
#Actual probability of getting 3 boys out of 8 kids is 20.98
dbinom(3,8,.51)
```

```
## [1] 0.2098355
```

```
#By regular method

(p*p*p*(1-p)*(1-p)*(1-p)*(1-p)*(1-p))*length(combn(c("boy","boy","boy","girl","girl","girl","gir
l","girl"),3))
```

```
## [1] 0.6295065
```

```
# It is more tedious to do manually. Because we need to calculate the combinations and then mult
iply with the probability of each event
```

3.42 Serving in volleyball. A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other.

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

```
#Probability of 3rd successful serve in
dbinom(2,9,.15)
```

```
## [1] 0.2596674
```

```
n <- 10     #number of attempts
k <- 3      #successful serves
p <- 0.15   #successful serve prob


factorial(n - 1) / (factorial(k-1) * (factorial(n - k))) * p^k * (1-p)^(n-k)
```

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

b. Suppose she has made two successful serves in nine attempts. What is the probability that her 10th serve will be successful?

As these serves are independent of each other, the probability of any one individual serve is 15%.

c. Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be different. Can you explain the reason for this discrepancy?

Probability of single event is different than the probability of combinations. (a) is about future serves. (b) is about single successful event.