HW2

Ahsanul Choudhury

September 25, 2016

3.2 Area under the curve, Part II. What percent of a standard normal distribution N(µ = 0, = 1) is found in each region? Be sure to draw a graph.

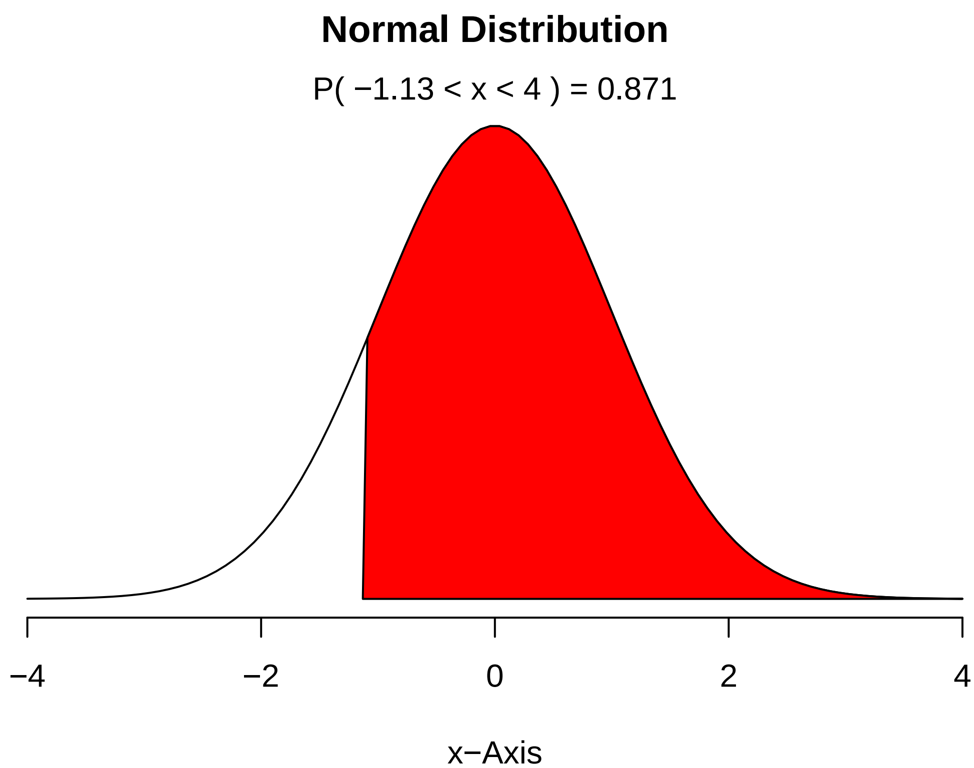
(a) Z > -1.13 (b) Z < 0.18 (c) Z > 8 (d) |Z| < 0.5

Sol (a)

```r
1- pnorm(-1.13)
```

```
## [1] 0.8707619
```

```r
normalPlot(bounds=(c(-1.13, 4)))
```

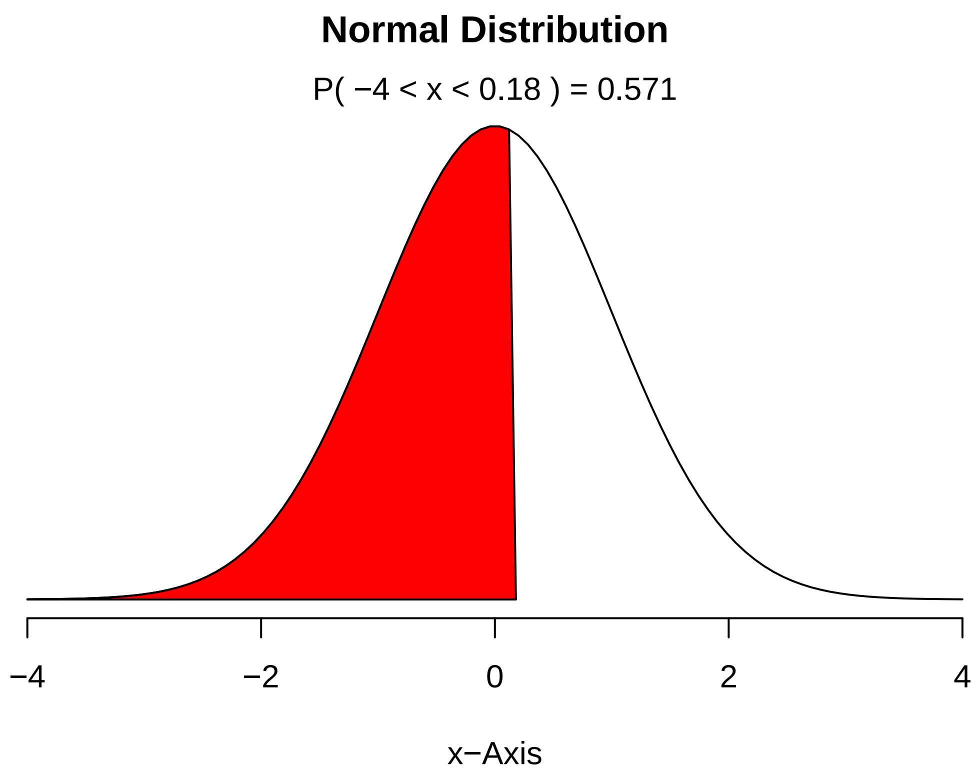


sol (b)

```r
pnorm(0.18)
```

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

```r
normalPlot(bounds = (c(-4, 0.18)))
```

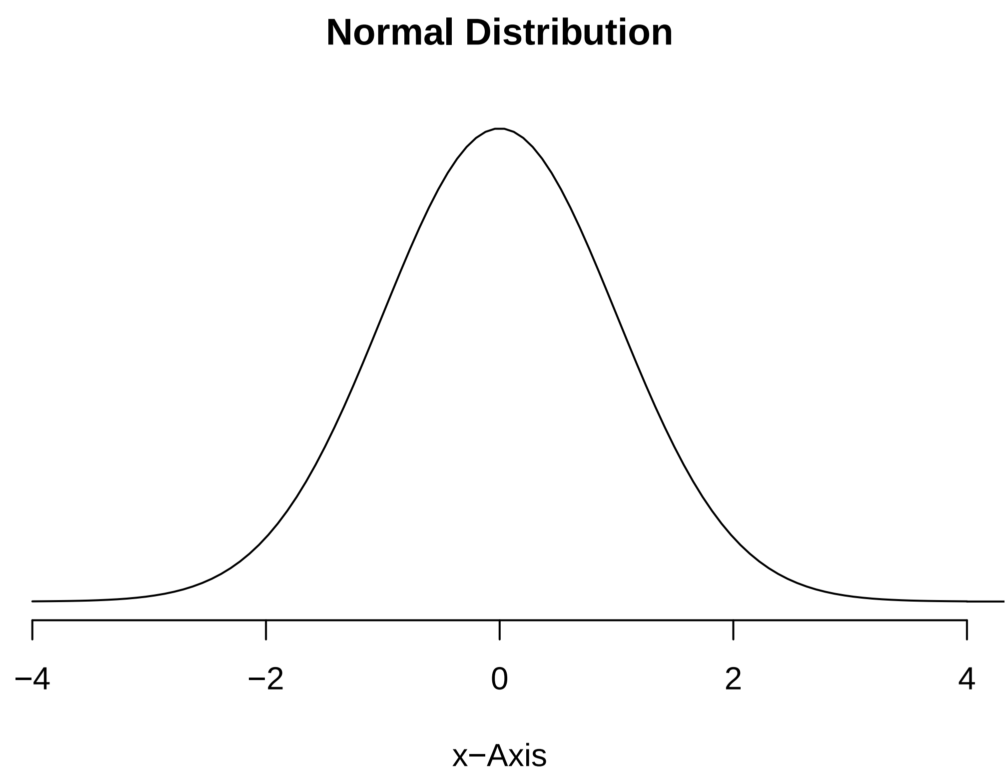


sol (c)

```r
1- pnorm(8)
```

```
## [1] 6.661338e-16
```

```r
normalPlot(bounds=c(8, 4))
```



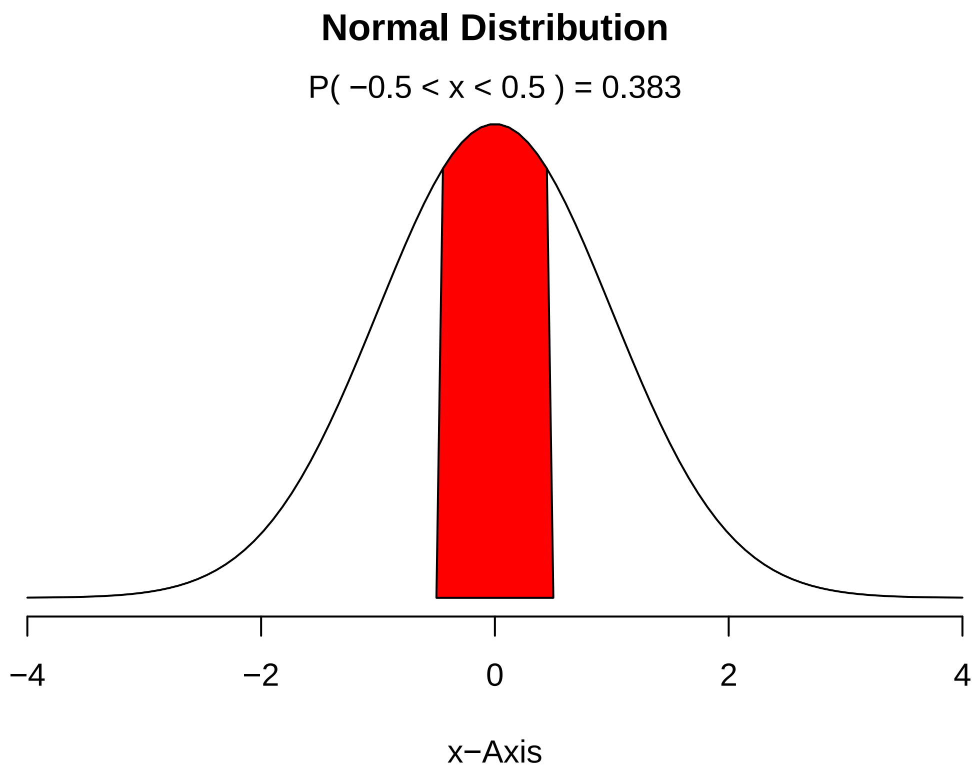
sol (d)

```
area <- 2* (0.5 - pnorm(-1 * 0.5))
area
```

```
## [1] 0.3829249
```

```
lowerbound <- qnorm(.5-(area/2))
upperbound <- qnorm(.5+(area/2))


normalPlot(bounds =c(lowerbound, upperbound))
```



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 astandard deviation of 583 seconds.

. The finishing times of the Women, Ages 25 - 29 group has a mean of 5261 seconds with astandard deviation of 807 seconds.

. The distributions of finishing times for both groups are approximately Normal. Remember: a better performance corresponds to a faster finish.

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

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

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

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

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

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

Answer: (a)Men, (30-34): N(?? = 4313, ?? = 583) Women, (25-29): N(?? = 5261, ?? = 807)

(b)

```r
leogroupmean <- 4313
leogroupsd <- 583
leo <- 4948
leoz <- round(((leo - leogroupmean)/leogroupsd), 3)
leoz
```

```
## [1] 1.089
```

```r
marygroupmean <- 5261
marygroupsd <- 807
mary <- 5513
maryz <- round(((mary- marygroupmean)/marygroupsd), 3)
maryz
```

```
## [1] 0.312
```

Leo's z score is 1.089 and Mary's z score is 0.312, this tells us Leo finished 1.089 standard deviation above his group and mary finished 0.312 standard deviation above hers.

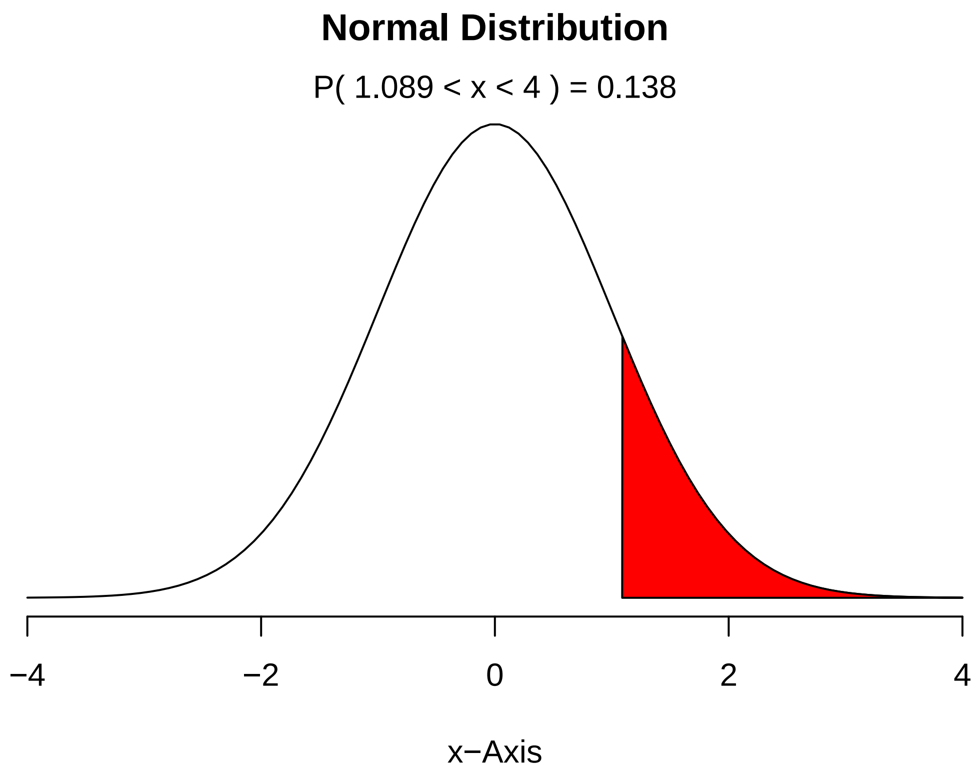
(c) Mary ranked better, they both finised above their respective group's mean but Mary's z score in closer to the mean and in this normal distribution the closer to the mean is the better.

(d)

```r
leofaster <- (round(1 - pnorm(leoz), 4))
leofaster
```

```
## [1] 0.1381
```

```r
normalPlot(mean = 0, sd = 1, bounds = c(leoz, 4))
```



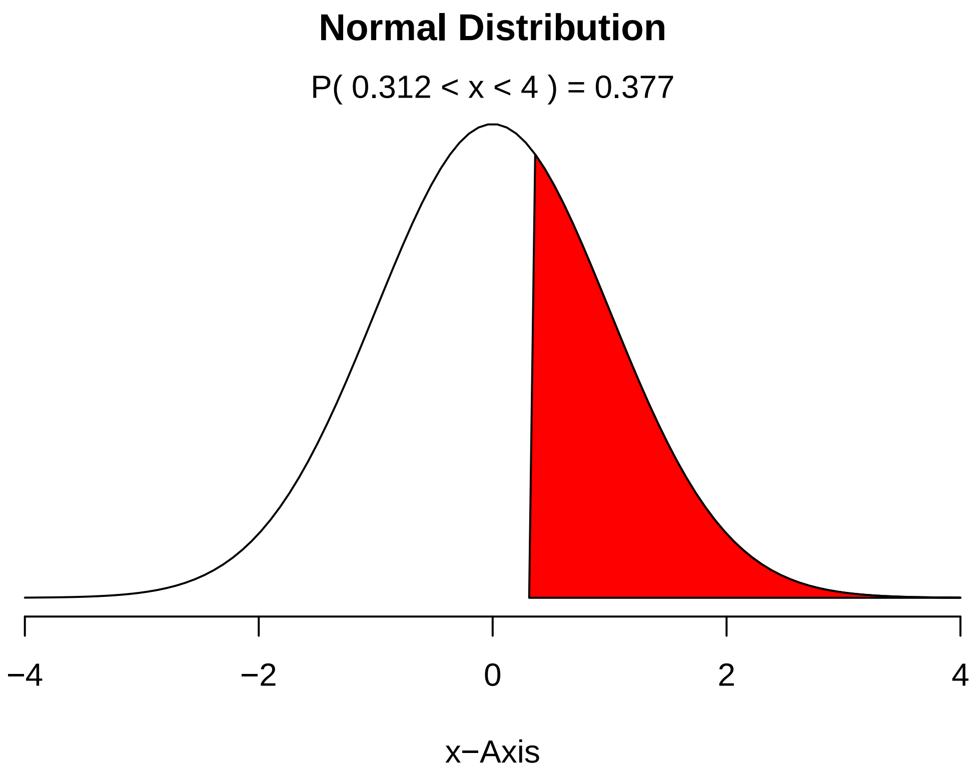
Leo was faster than 13.8% triathletes in his group

(e)

```
maryfaster <- (round(1 - pnorm(maryz), 4))
maryfaster
```

```
## [1] 0.3775
```

```
normalPlot(mean = 0, sd = 1, bounds = c(maryz, 4))
```



Mary was faster than 37.75% triathletes in her group

(f) If the distributions of finishing times are not nearly normal my answers to parts (b) - (e) would definately change, normal probability calculation will not apply. the use of z score required a normal distribution.

3.18 Heights of female college students. Below are heights of 25 female college students. (textbook, page:161)

(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.

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

Solution:

(a)

```
heights <- c(54, 55, 56, 56, 57, 58, 58, 59, 60, 60, 60, 61, 61, 62, 62, 63, 63, 63, 64, 65, 65, 67, 67

meanheights <- mean(heights)
meanheights
```

```
## [1] 61.52
```

```r
heightsd <- sd(heights)
heightsd
```

```
## [1] 4.583667
```

```r
length(heights[heights < meanheights  +heightsd & heights > meanheights - heightsd]) /length(heights)
```

```
## [1] 0.68
```

```r
length(heights[heights  < meanheights + 2 * heightsd & heights > meanheights - 2 * heightsd]) /length(he
```

```
## [1] 0.96
```

```r
length (heights[heights < meanheights + 3 * heightsd & heights > meanheights - 3 * heightsd]) /length(he
```

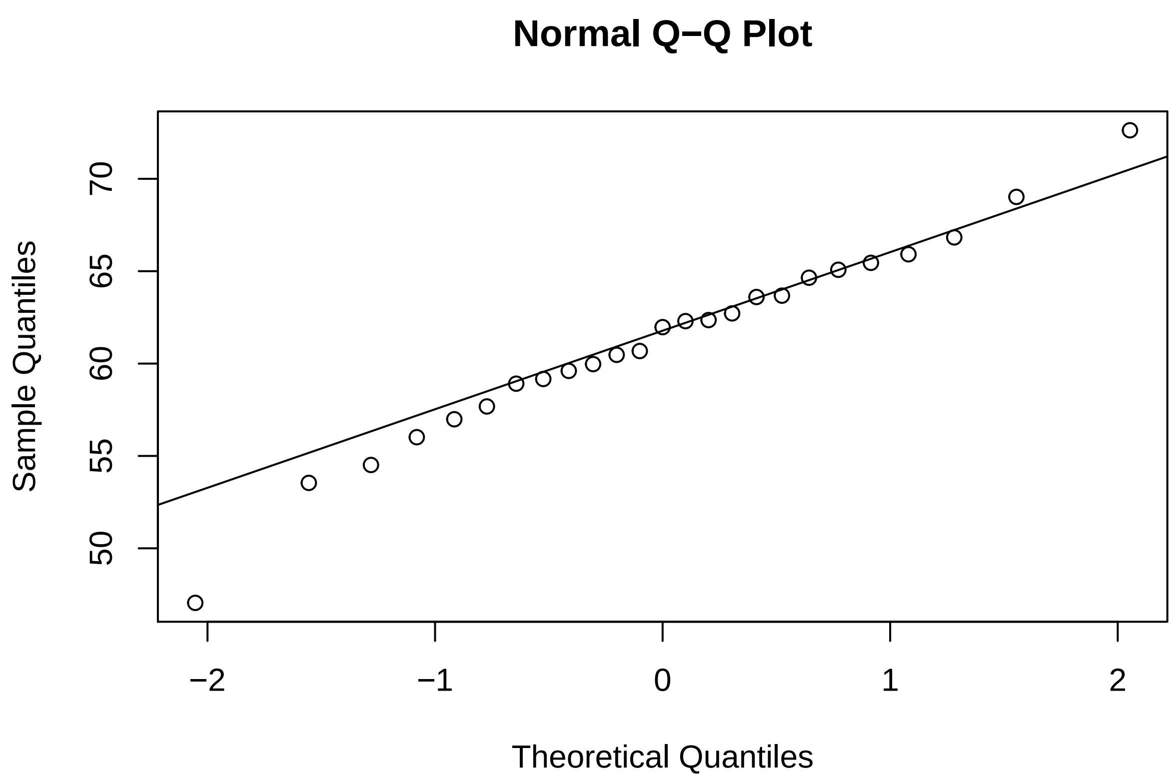
```
## [1] 1
```

Heights appears to approximately follow the 68-95-99.7% Rule.

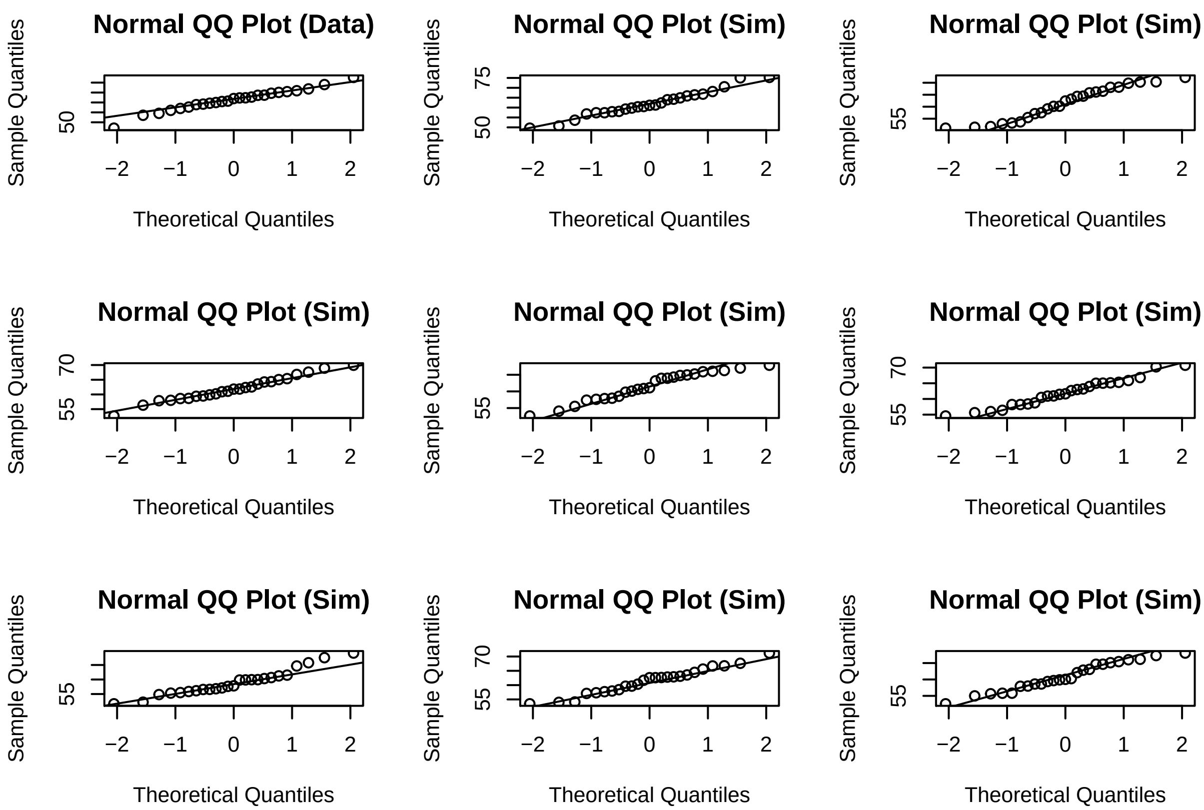
(b)

```r
simheight <- rnorm(length(heights), meanheights, heightsd)

qqnorm(simheight)
qqline(simheight)
```



```r
qqnormsim(simheight)
qqline(simheight)
```



Data appear to follow normat distribution. All the simulated data appears to fall within normal line.

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?

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

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

(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?

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

Solution

(a)

```
p <- .02
n <- 10
(1 - p)^(n - 1) * p
```

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

Probability of 10th to be first defective is 0.01667496

(b)

```
n <- 100
(1 - p)**n
```

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

probability if no defective in 100 is 0.1326796

(c)

```
1/p
```

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

```
sqrt((1-p)/(p^2))
```

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

On ave 50 produced to get first defective and standard deviation is 49.49747

(d)

```
p <- 0.05
1/p
```

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

```
sqrt((1-p)/(p^2))
```

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

On snd machine ave 20 produced to get first defective and standard deviation is 19.49359

(e) Increasing probability decreases the mean and standard deviation.

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.

(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.

(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)

```
dbinom(2, size=3, prob=0.51)
```

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

Probability of two boys are 0.382347

(b)

```
combo1 <- c("girl", "boy", "boy")
combo2 <- c("boy", "girl", "boy")
combo3 <- c("boy", "boy", "gir1")
df <- data.frame(combo1, combo2, combo3)
df
```

```
##   combo1 combo2 combo3
## 1   girl    boy    boy
## 2    boy   girl    boy
## 3    boy    boy   gir1
```

```
boyprob <- 0.51

((1 - boyprob) * boyprob * boyprob)*3
```

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

Answer to (a) and (b) matches

(c)

```
factorial(8) / (factorial(3)*factorial(8 - 3))
```

```
## [1] 56
```

There will be possible 56 different combinitions and the formula will be longer if we go with approach in (b), the approach in (a) will be simpler.

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?

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

(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?

Solution (a)

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

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

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

The probability is 0.3895012

(b) The answer will be 0.15, because she has to be successful in 1 try and thats the probability of her success.

(c)There are no discrepancy, on solution (a) we are finding the probability that on tenth try she is going to have 3rd successful hit where in solution (b) we are assuming she already has two successful hits.