Chapter 4 - Distributions of Random Variables

John Mazon

library(DATA606)

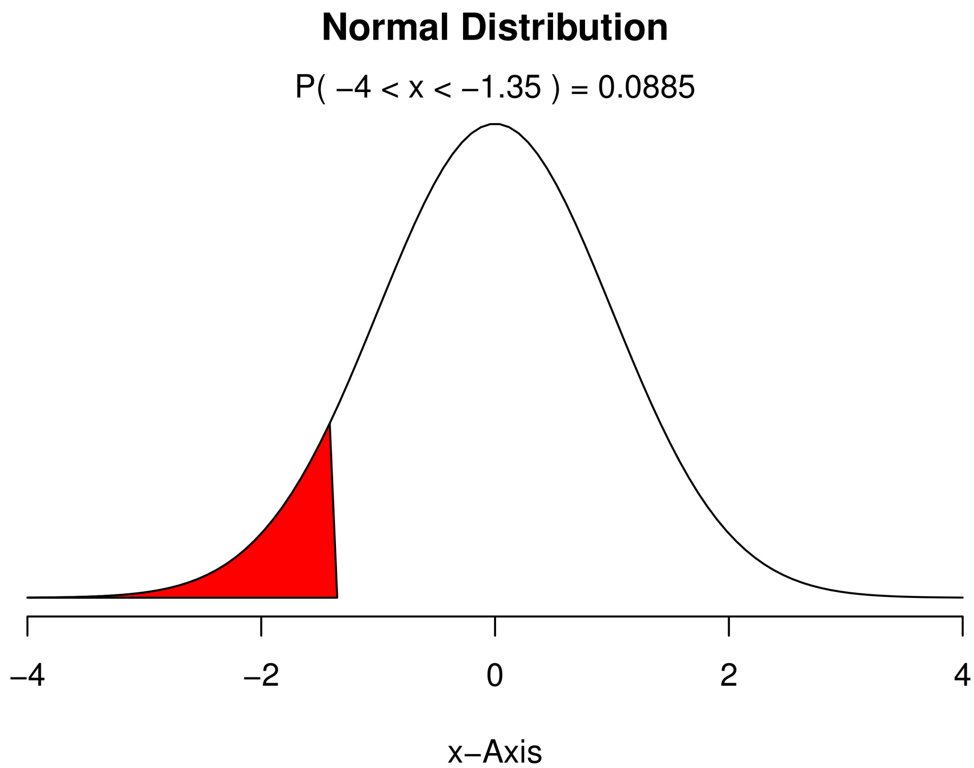
Area under the curve, Part I. (4.1, p. 142) What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

(a) $Z < -1.35$ The percent of a standard normal distribution is 8.85%.

```
library(qualityTools)
```

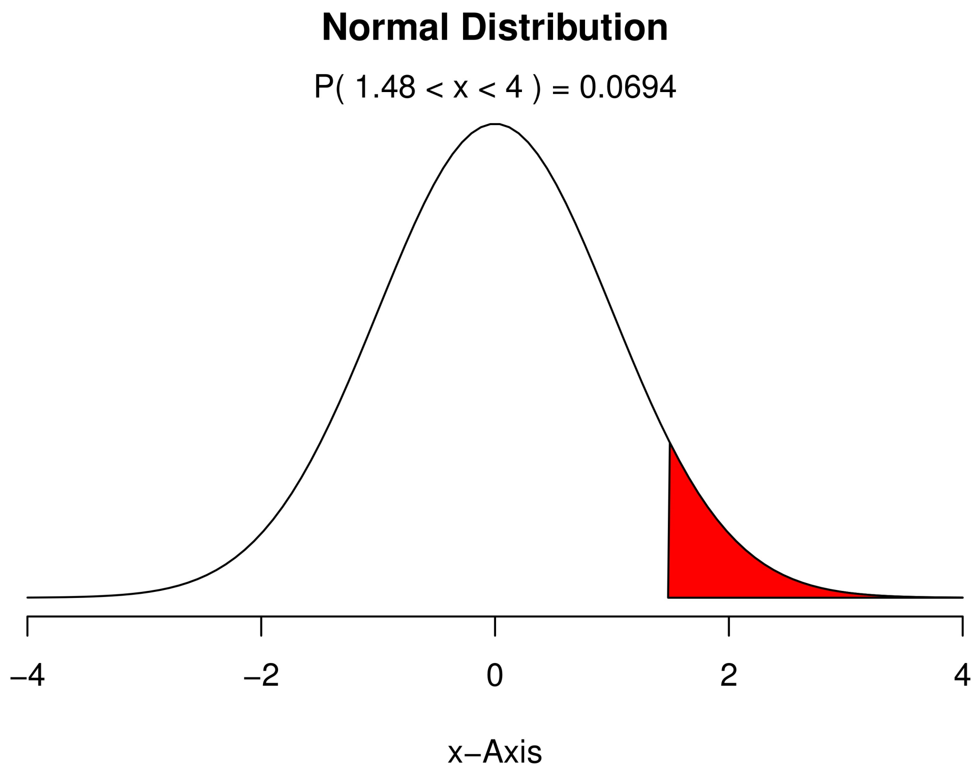
```
## Loading required package: Rsolnp

## Loading required package: MASS

##
## Attaching package: 'qualityTools'

## The following object is masked from 'package:stats':
##
##     sigma
```

```
DATA606::normalPlot(mean = 0, sd = 1, bounds=c(-4,-1.35))
```



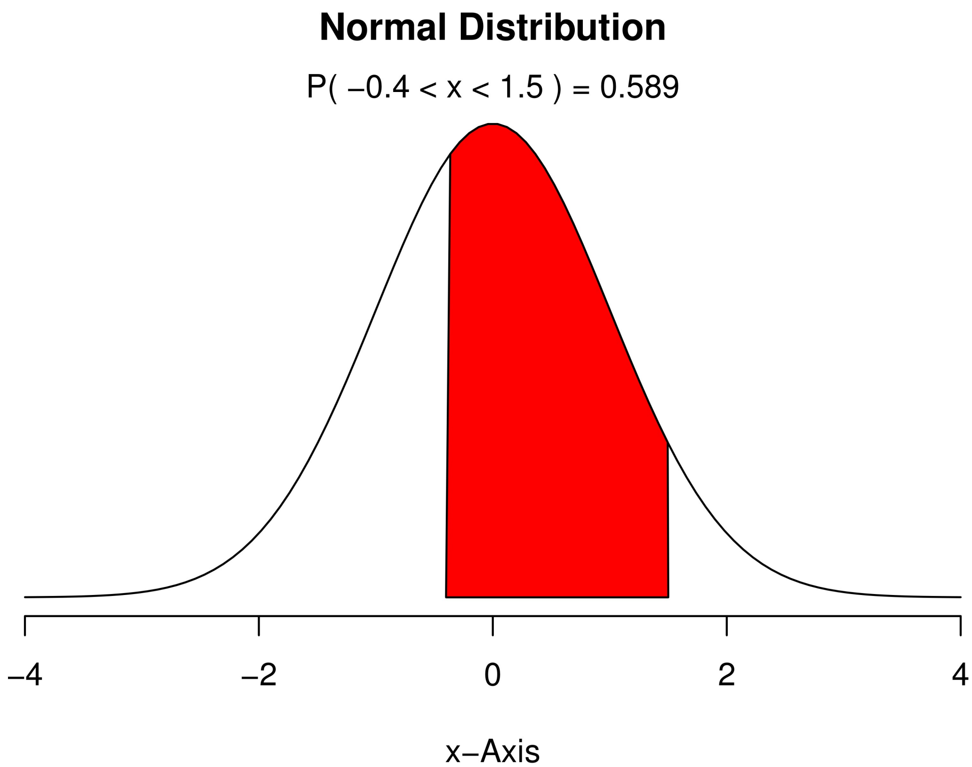
(b) $Z > 1.48$ The percent of a standard normal distribution is 6.94%.

```
library(qualityTools)


DATA606::normalPlot(mean = 0, sd = 1, bounds=c(1.48,4), tails = FALSE)
```



(c) $-0.4 < Z < 1.5$ The percent of a standard normal distribution is 58.9%

```
library(qualityTools)

DATA606::normalPlot(mean = 0, sd = 1, bounds=c(-0.4,1.5), tails = FALSE)
```

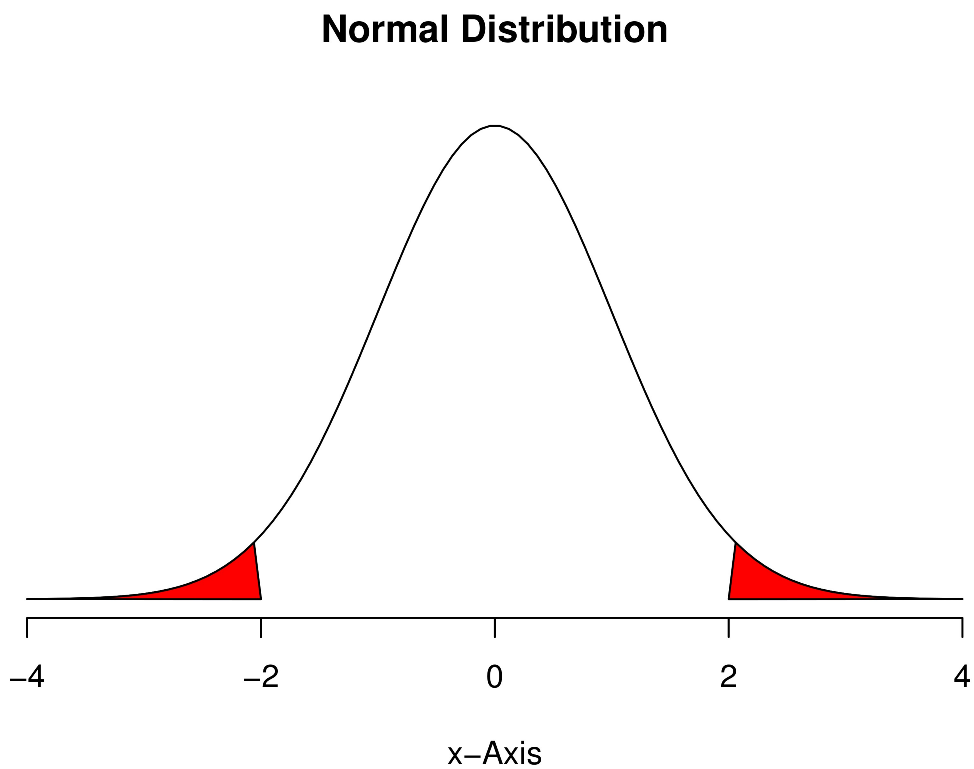


(d) $|Z| > 2$ The percent of a standard normal distribution is 4.55%

```
mu <- 0
sd <- 1
Z <- 2

x <- Z * sd + mu

x1 <- -x
x2 <- x

p1 <- pnorm(x1, mean = 0, sd = 1)
p2 <- 1- pnorm(x2, mean = 0, sd = 1)
p1 +  p2
```

```
## [1] 0.04550026
```

```
DATA606::normalPlot(mean = 0, sd = 1, bounds = c(x1, x2), tails = TRUE)
```



Triathlon times, Part I (4.4, p. 142) 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.

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

```
women_mean=5261
women_sd=807

men_mean=4313
men_sd=583
```

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

```
Leo_z_score<-(4948-men_mean)/men_sd
Leo_z_score
```

```
## [1] 1.089194
```

```
mary_z_score<-(5513-women_mean)/women_sd
mary_z_score
```

```
## [1] 0.3122677
```

(c) Did Leo or Mary rank better in their respective groups? Explain your reasoning. Since Mary is closer to mean, I believe the answer is Mary

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

```
pnorm(Leo_z_score,lower.tail=FALSE)
```

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

13.8%

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

```
pnorm(mary_z_score,lower.tail=FALSE)
```

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

37.7%

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

I believe NO. The performance from leo and mary are relative to other participants of their groups.

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

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
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

(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. According to below calculations, heights approx follow 68-95-99.7% Rule

```
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, 

heights_mean<-mean(heights)

heights_sd<-sd(heights)

#68%
1-2*pnorm(heights_mean+heights_sd,mean=heights_mean,sd=heights_sd,lower=FALSE)
```

```
## [1] 0.6826895
```

```
#95%
1-2*pnorm(heights_mean+2*heights_sd,mean=heights_mean,sd=heights_sd,lower=FALSE)
```

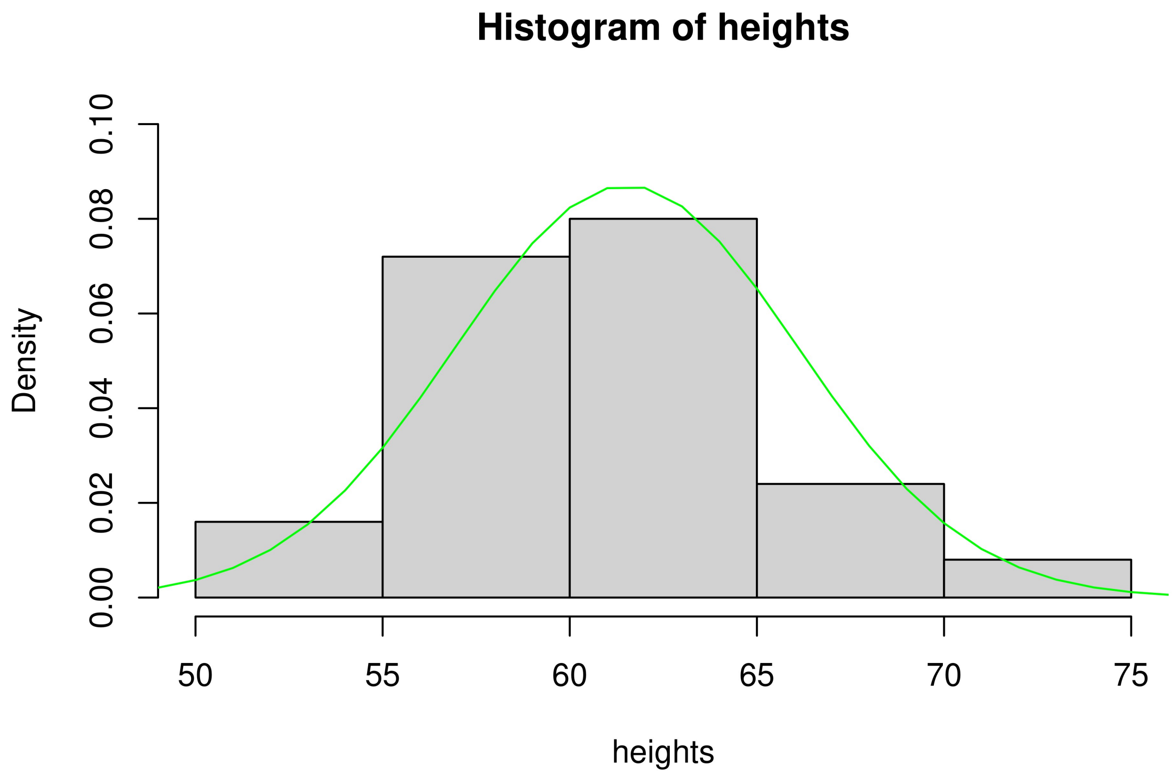
```
## [1] 0.9544997
```

```
#99.7%
1-2*pnorm(heights_mean+3*heights_sd,mean=heights_mean,sd=heights_sd,lower=FALSE)
```

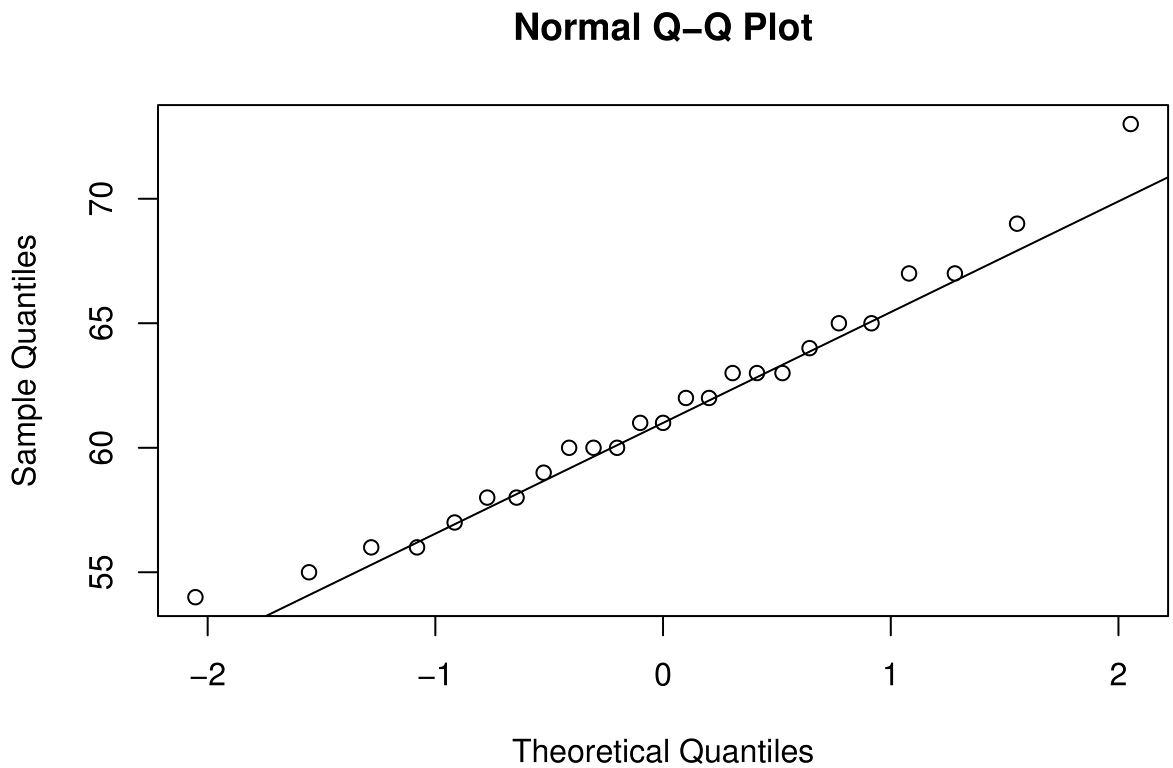
```
## [1] 0.9973002
```

(b) Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below. I believe the data do not appear to follow a normal distribution. There is a slight right skew in the heights data(not symmetric).

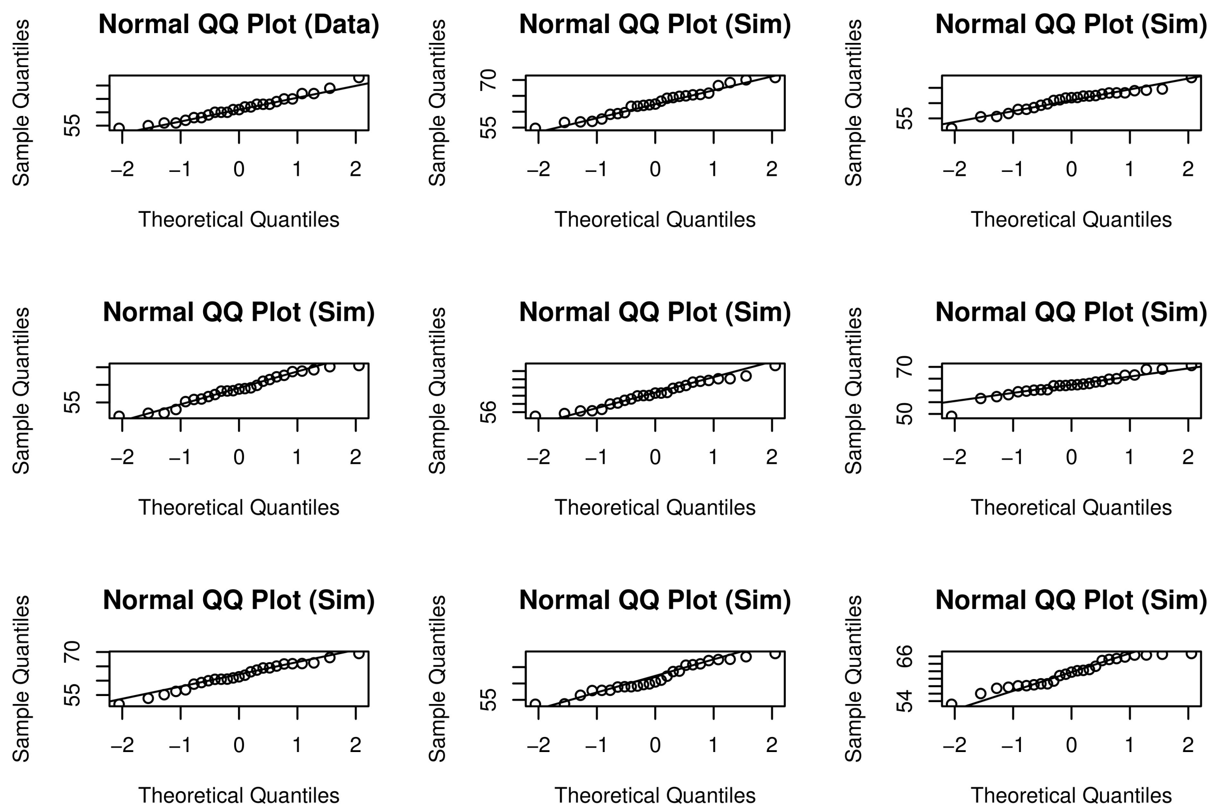
```
heights_mean<-mean(heights)

heights_sd<-sd(heights)

hist(heights, probability = TRUE,ylim = c(0, 0.1))
x <- 40:80
y <- dnorm(x = x, mean = heights_mean, sd = heights_sd)
lines(x = x, y = y, col = "green")
```

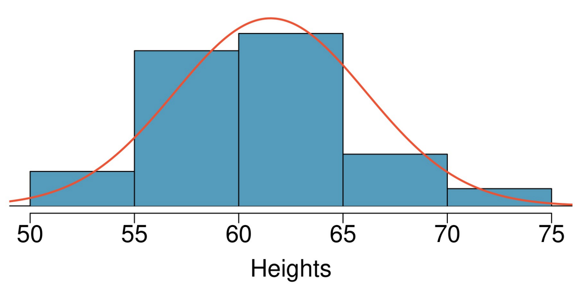


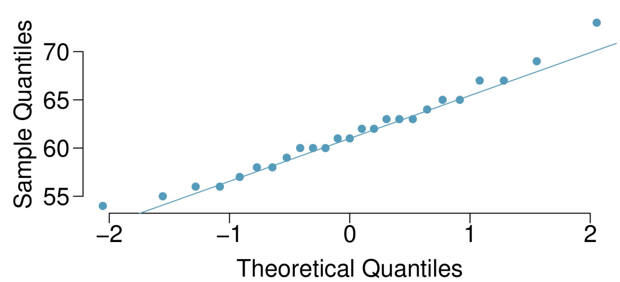

```r
qqnorm(heights)
qqline(heights)
```




```
DATA606::qqnormsim(heights)
```








```
# Use the DATA606::qqnormsim function
```

Defective rate. (4.14, p. 148) 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?

```
(0.98^9)*0.02
```

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

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

```
0.98^100
```

```
## [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?

```
1/0.02
```

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

```
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?

```
1/0.05
```

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

```
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? It appears as possibility/probability becomes greater, mean and sd decrease. In other words a certain ocurrence will repeat more times.

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.

```
dbinom(2,3,0.51)
```

```
## [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.

```
((0.51^2)*0.49)+((0.51^2)*0.49)+((0.51^2)*0.49)
```

```
## [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). This approach would be tedious since the number of possible orderings would be higher.

Serving in volleyball. (4.30, p. 162) 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?

```
p <- .15
n <- 10
k <- 3
P <- choose(n-1, k-1) *(1-p)^(n-k)*p^k
P
```

```
## [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?

10th serve success probability is 0.15, since her serves are independent of each other.

(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?

in question [a] we are calculating a specific probability of having 3 successes out of 10. later in question [b] we are calculating the probability of one successful serve and since the serves are independent, it does not matter what happened in previous serves.