Chapter 4 - Distributions of Random Variables

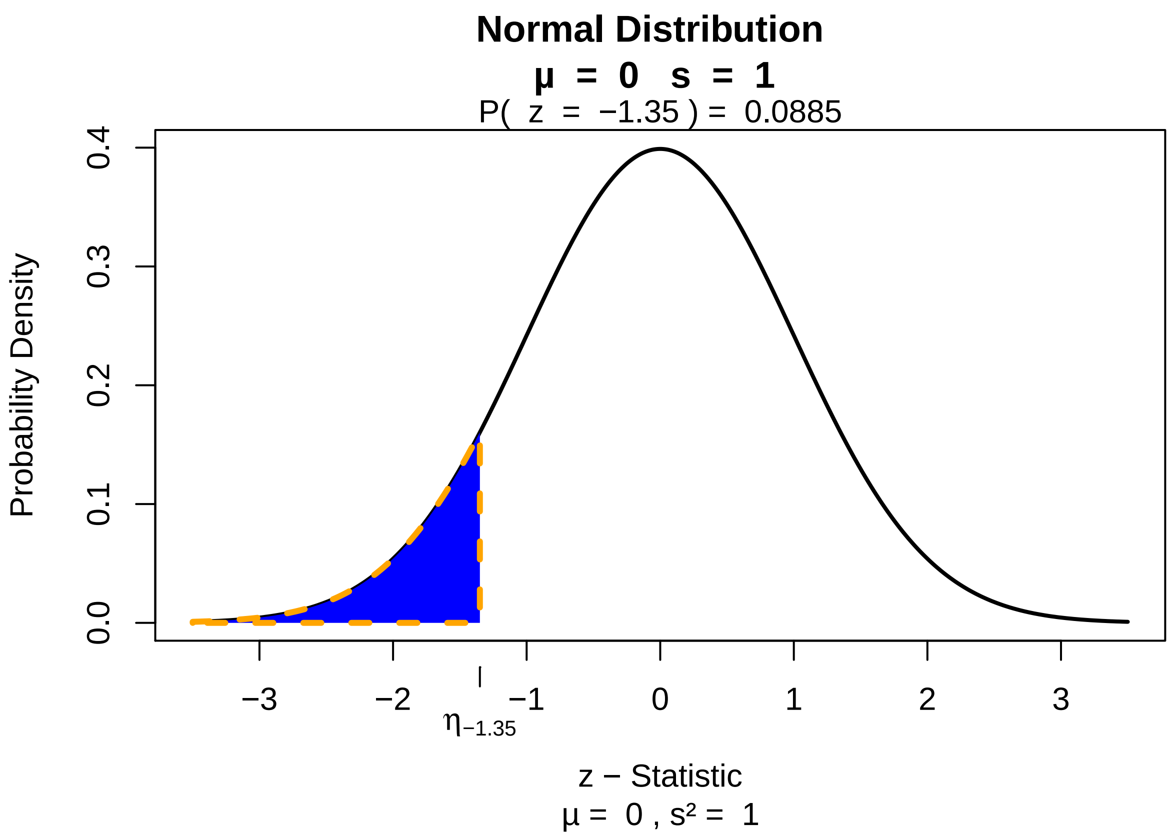
Zachary Safir

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.

```
library(visualize)
```

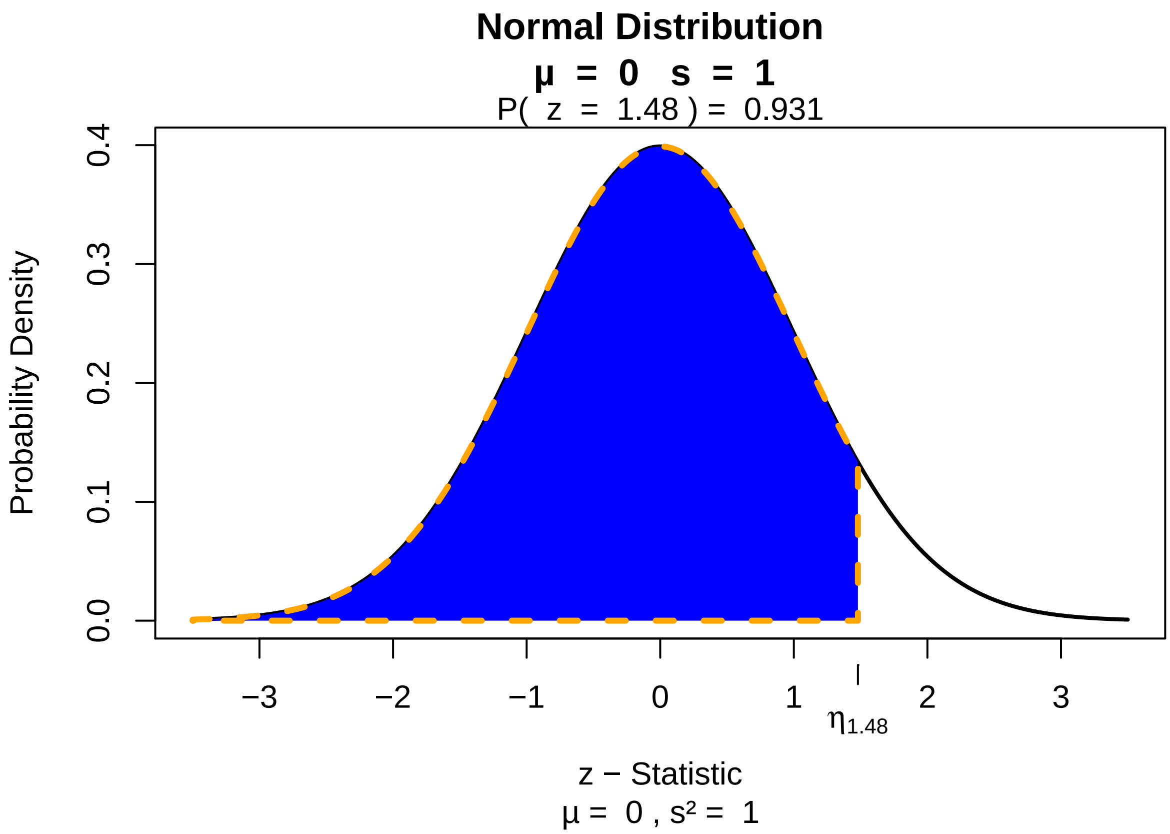
(a) $Z < -1.35$

```
visualize.norm(stat=-1.35,mu=0,sd=1)
```



(b) $Z > 1.48$

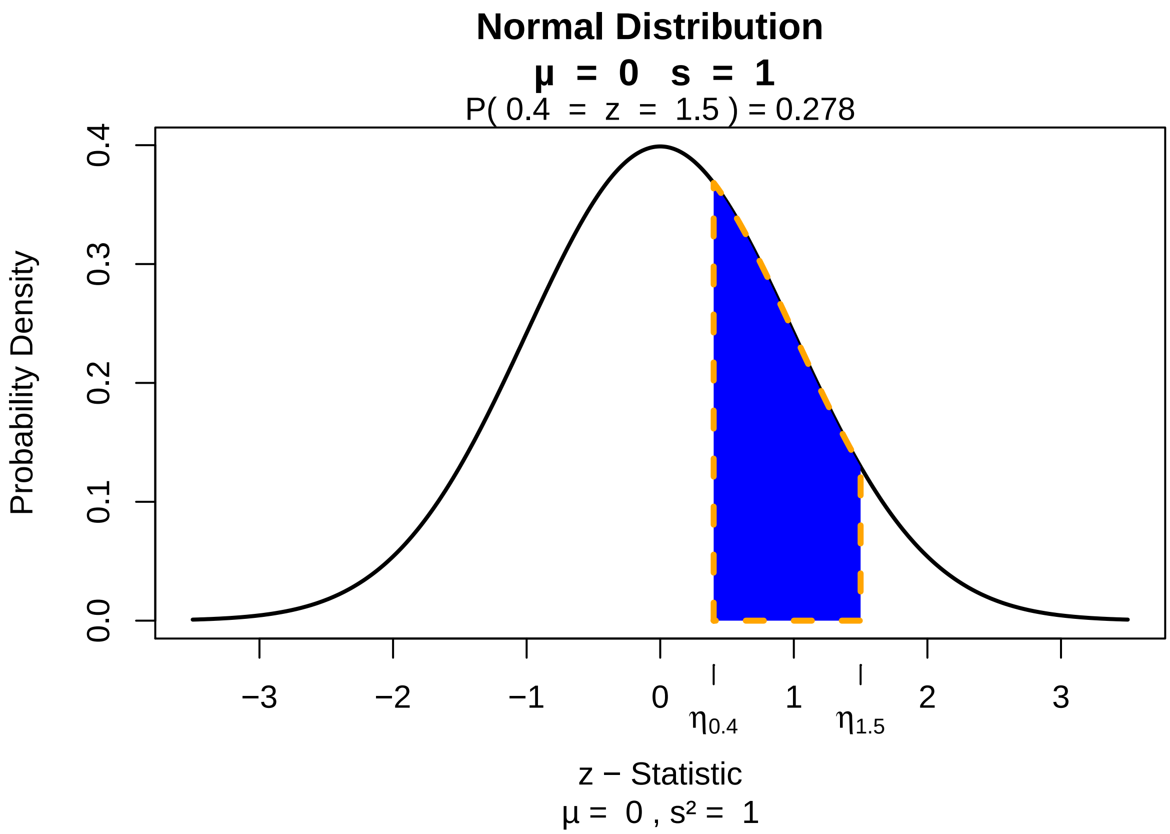
```
visualize.norm(stat=1.48,mu=0,sd=1)
```



(c)

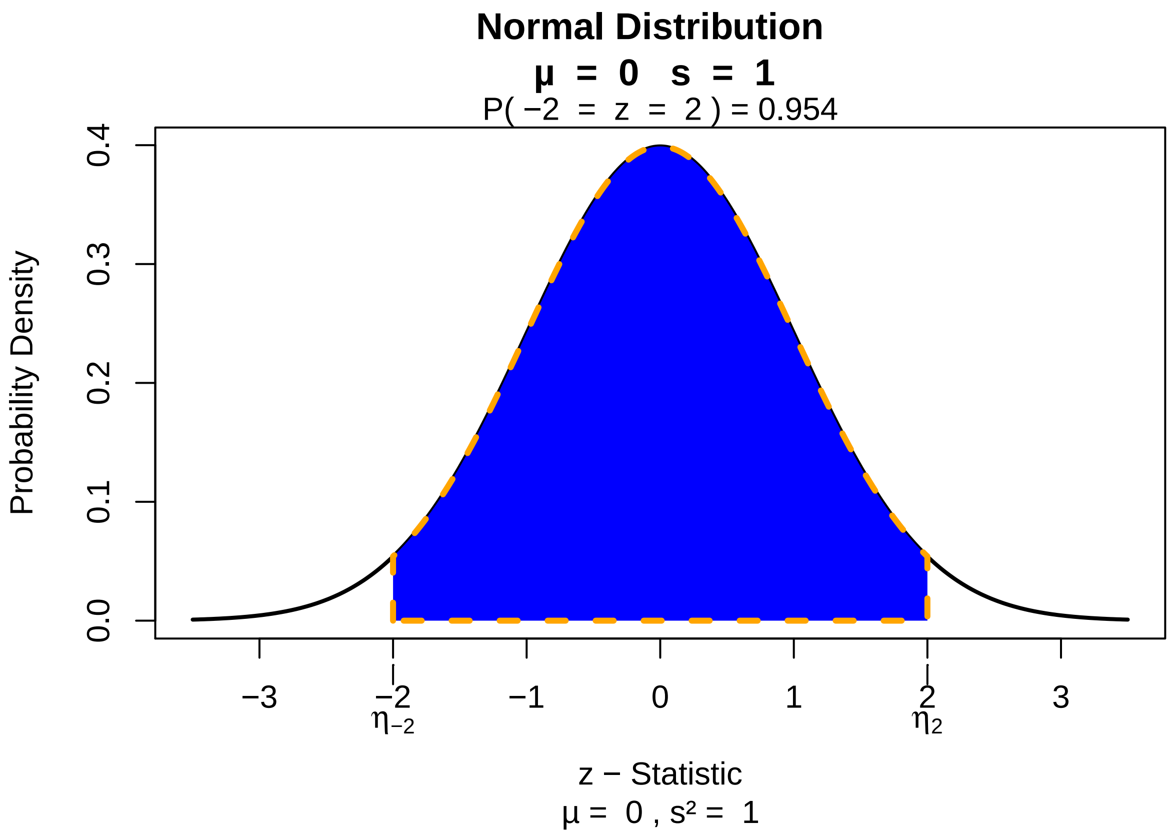
$-0.4 < Z < 1.5$

```
visualize.norm(stat=c(0.4,1.5),mu=0,sd=1,section="bounded")
```



(d) $|Z| > 2$

```
visualize.norm(stat=c(-2,2),mu=0,sd=1,section="bounded")
```


Normal Distribution
μ = 0 s = 1
P(−2 = z = 2) = 0.954
Probability Density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
η−2
η2
z − Statistic
μ = 0 , s² = 1

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.

Men, Ages 30 - 34: $\mu = 4313, \sigma = 583$

Women, Ages 25 - 29 : $\mu = 5261, \sigma = 807$

(b) What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you? $Z = \frac{x-\mu}{\sigma}$ is our formula Leo is , $Z = \frac{4948-4313}{583}$

```
(4948-4313)/583
```

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

Mary is, $Z = \frac{5513-5261}{807}$

```
(5513-5261)/807
```

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

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

```
1-pnorm(4948, mean = 4313, sd=583)
```

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

```
1-pnorm(5513, mean = 5261, sd=807)
```

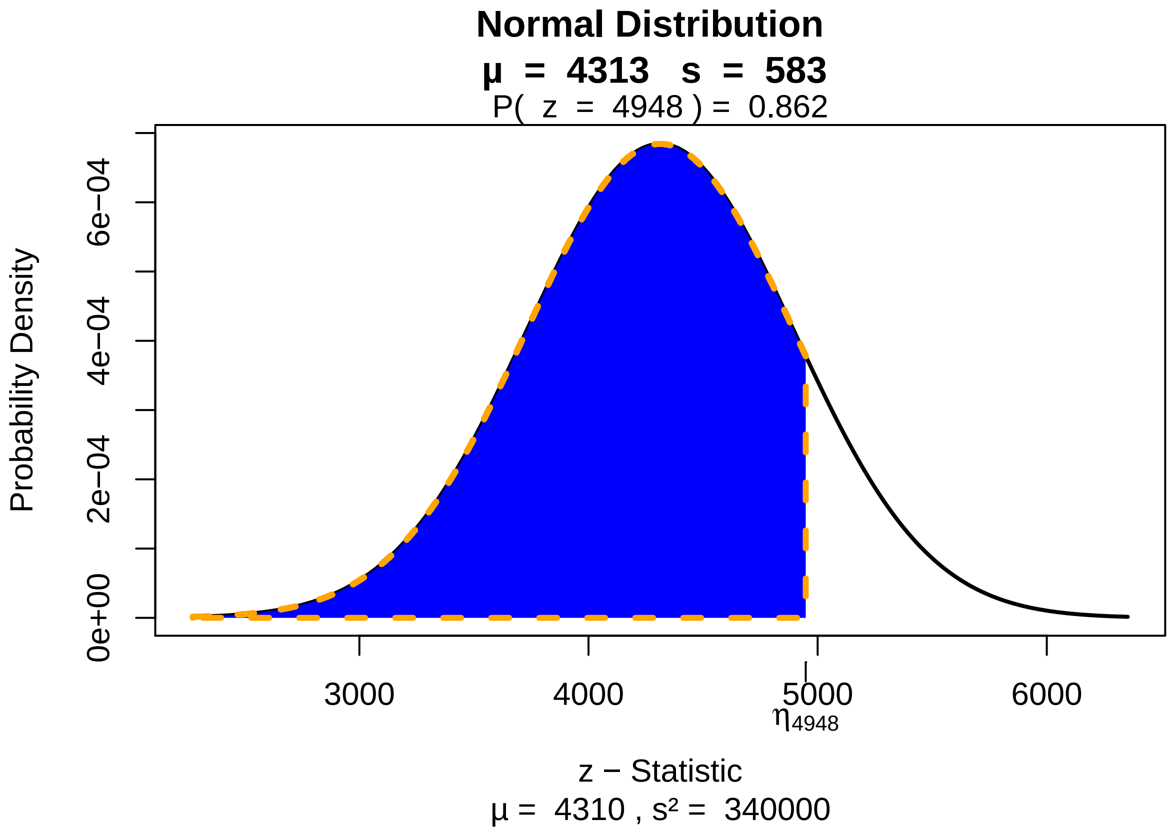
```
## [1] 0.3774186
```

Leo ranked much better than Mary. Leo was in the top 13.8 percent of his group whereas Mary was in the top 37.7 percent.

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

13.8 percent.

```r
visualize.norm(stat = 4948, mu=4313, sd =583, section = "lower")
```



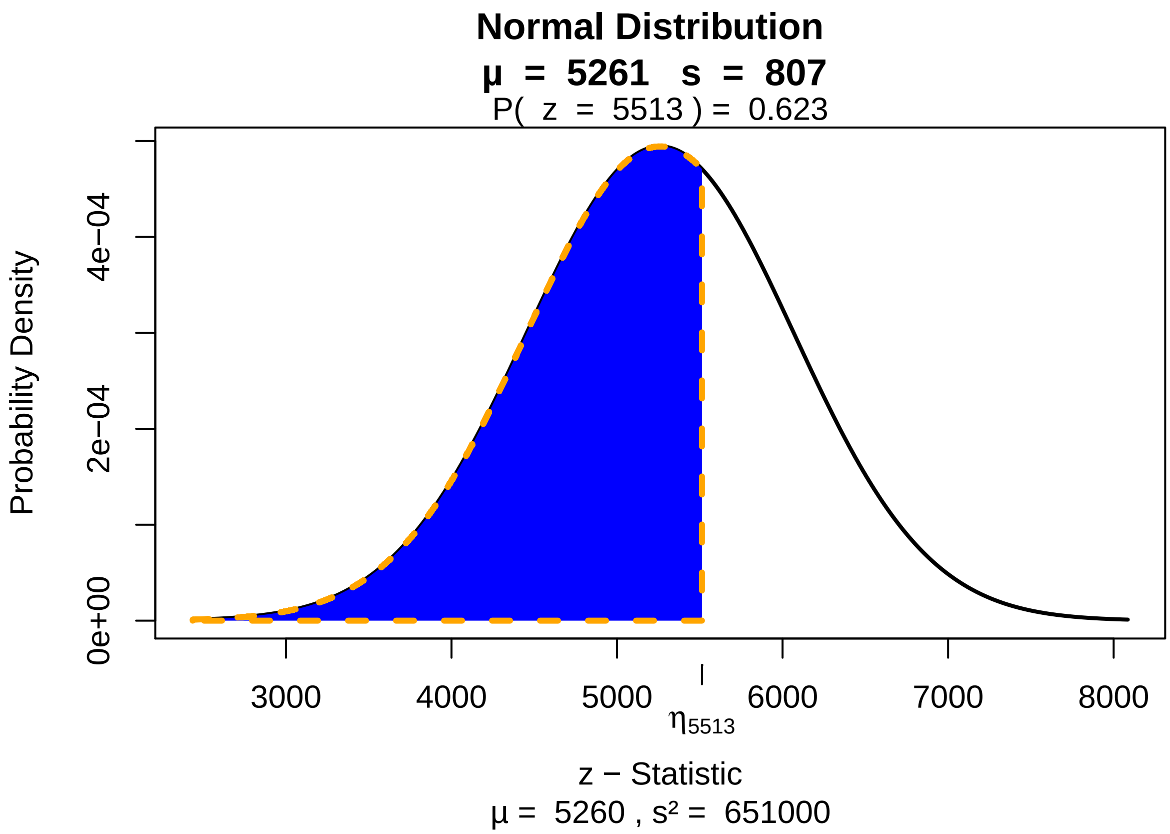

```r
pnorm(4948, mean = 4313, sd=583)
```

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

86 percent.

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

```r
visualize.norm(stat = 5513, mu=5261, sd =807, section = "lower")
```




```r
pnorm(5513, mean = 5261, sd=807)
```

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

62 percent.

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

It depends on how skewed the data is. If we were to assume that the current data set is normal, then we would need to add extreme values to the data that could possible change the means and standard deviation. In doing so, some of the answers above could changed based on variables we add.

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.

```
#Using one standard deviation above and below the mean.

pnorm(66.1, mean=61.52, sd=4.58) - pnorm(56.94, mean=61.52, sd=4.58)
```

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

```
#Using two standard deviations above and below the mean.

pnorm(70.68, mean=61.52, sd=4.58) - pnorm(52.36, mean=61.52, sd=4.58)
```

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

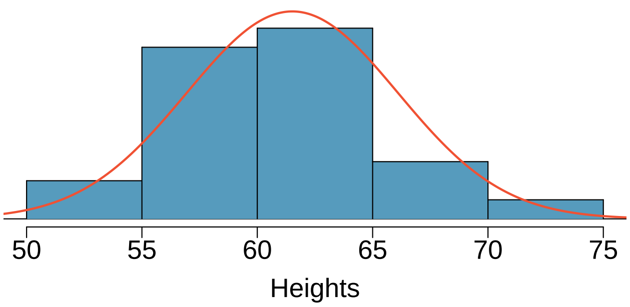
```
#Using three standard deviations above and below the mean.

pnorm(75.26, mean=61.52, sd=4.58) - pnorm(47.78, mean=61.52, sd=4.58)
```

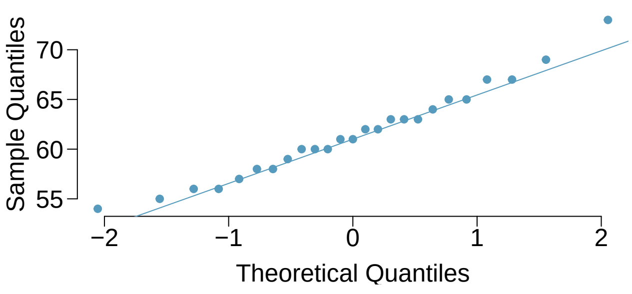
```
## [1] 0.9973002
```

Yes, the heights follow the rule as shown above.

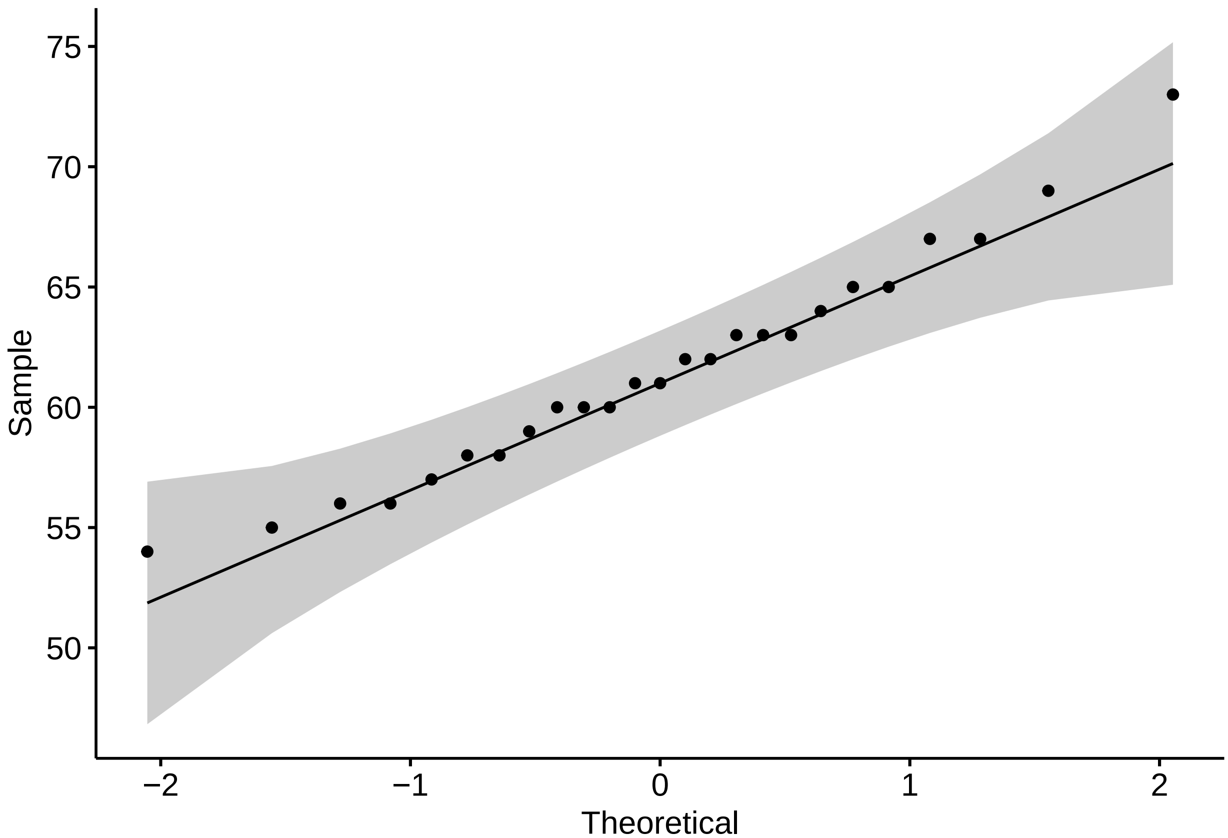
(b) Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.

It's mostly normal, skewing in the right direction, most likely due to the extreme 73 height value alone.






```
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, 

ggpubr::ggqqplot(heights)
```



75
70
65
60
55
50
Sample
−2
−1
0
1
2
Theoretical

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?

```r
#Built in stats function

dgeom(9,0.02)
```

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

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

```r
dbinom(0, 100, .02)
```

```
## [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?

```r
# How many we expect
1/.02
```

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

```r
# Standard Deviation
sqrt((1-.02)/(.02*.02))
```

```
## [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?

```r
# How many we expect
1/.05
```

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

```r
# Standard Deviation

sqrt((1-.05)/(.05*.05))
```

```
## [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?

As the probability increases, it appears that the wait time decreases.

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,.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.

The three possible cases are BBG,BGBG,GBB

```
# Additon rule

(.51 *.51 * .49) + (.51 *.49 * .51) + (.49 *.51 * .51)
```

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

As we see above, the values 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).

Finding every possible combination, calculating them out, and adding them all together will take much more time than using the built in R functions that let us immediately calculate the value we want using only three values we already know.

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

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

We are supposing that the serves are independent of each other, therefore the probability stays the same at 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?

The first scenario was a joint probability whereas the second scenario was asking about a single trial. With each serve independent of the rest, successfully pitching does not increase your chance of success on the following pitch.