CUNY MSDA 606 Spring 2016 Ch. 3 Graded Problems

James Topor

Chapter 3, Problems 3.2 (see normalPlot), 3.4, 3.18 (use qqnormsim from lab 3), 3.22, 3.38, 3.42

3.2 (see normalPlot)

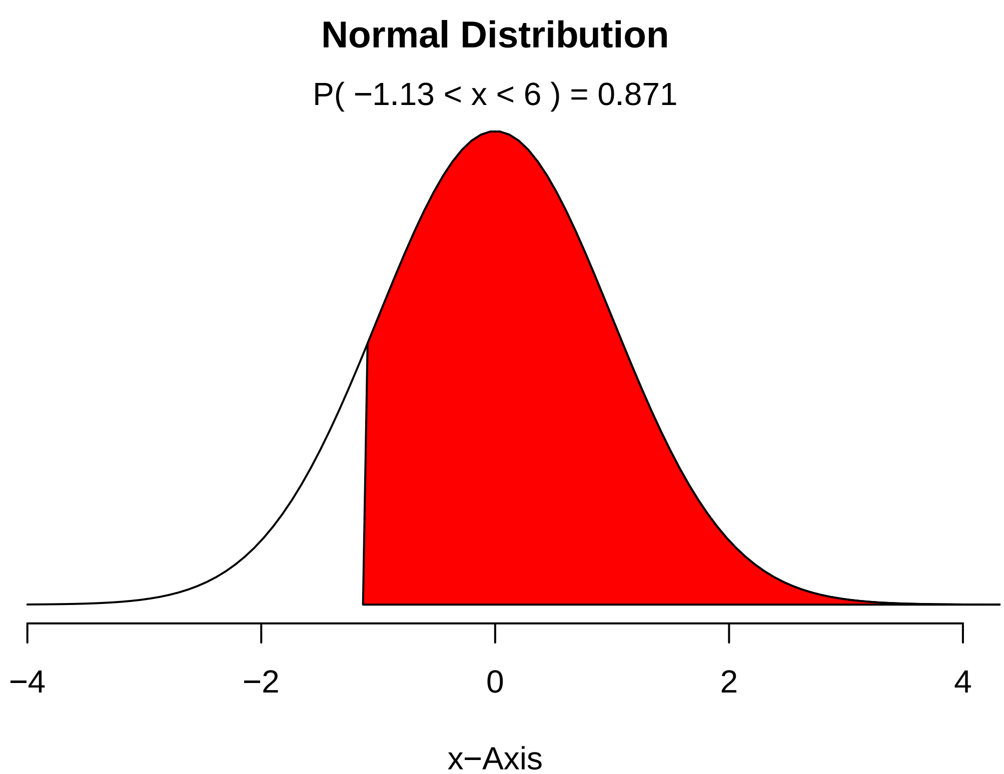
What percent of a standard normal distribution N(mu = 0; stddev = 1) is found in each region? Be sure to draw a graph.

a) Z > -1.13

```r
1 - pnorm(-1.13, 0, 1)
```

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

```r
normalPlot(0, 1, bounds = c(-1.13, 6))
```

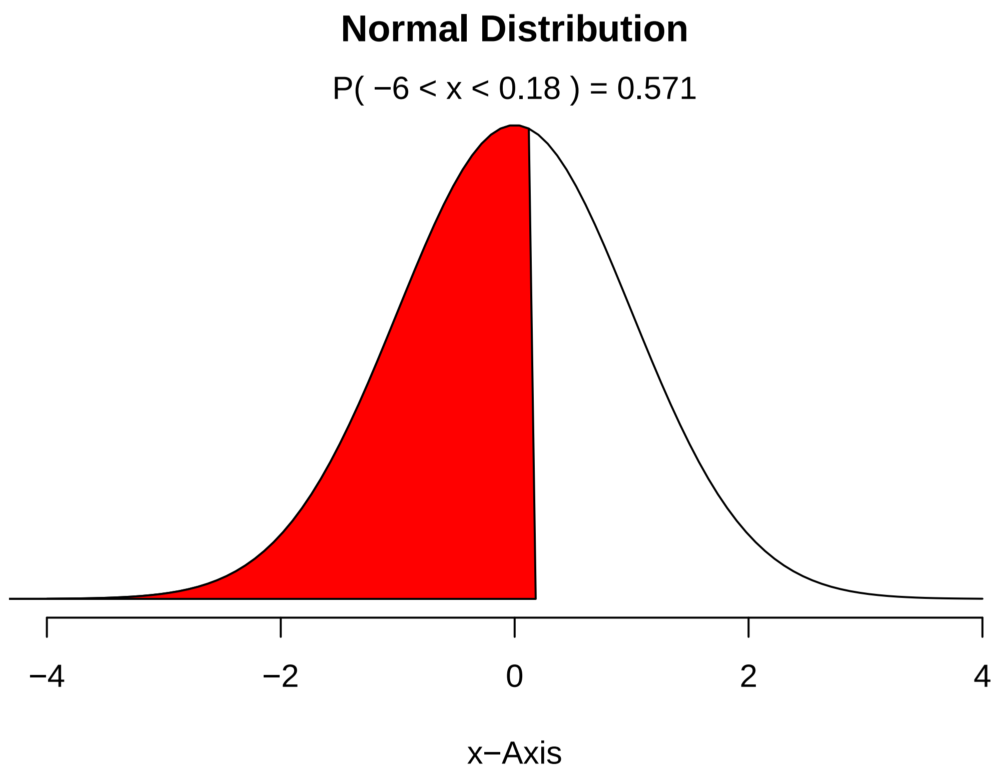


b) Z < 0.18

```r
pnorm(.18, 0, 1)
```

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

```r
normalPlot(0, 1, bounds = c(-6, .18))
```

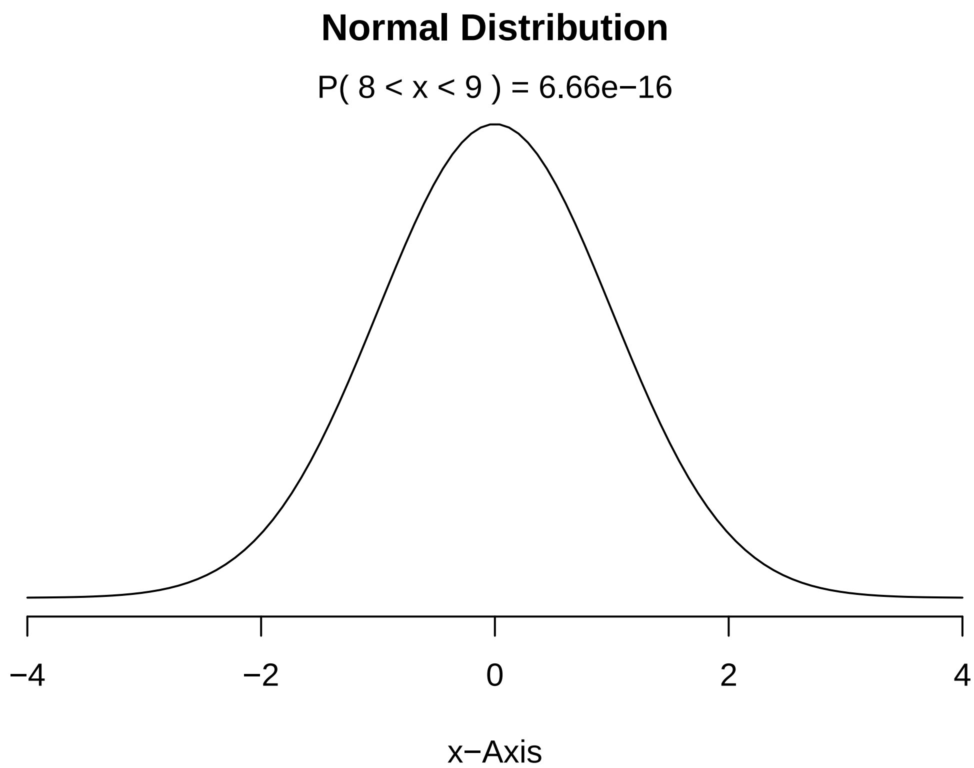


c) $Z > 8$

```r
1 - pnorm(8, 0, 1)
```

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

```r
normalPlot(0, 1, bounds = c(8, 9))
```

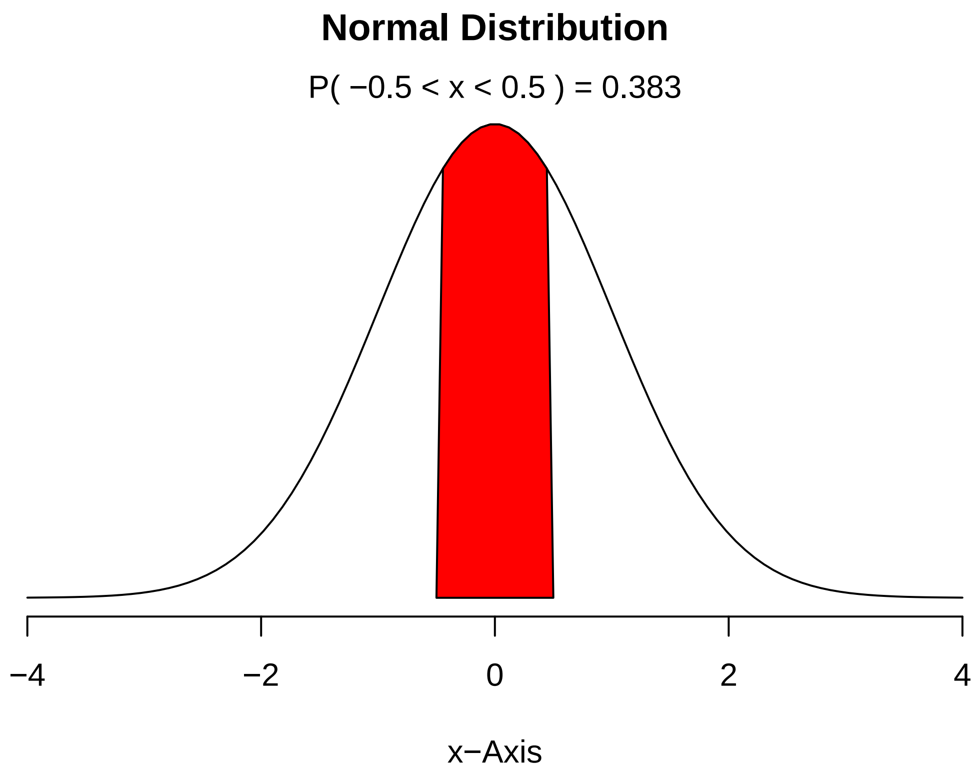


d) $|Z| < 0.5$

```
pnorm(.5, 0, 1) - pnorm(-.5, 0, 1)
```

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

```
normalPlot(0, 1, bounds = c(-.5, .5))
```



3.4 Men, Ages 30 - 34 group has a mean of 4313 seconds with a standard deviation of 583 seconds, distribution normal.

Women, Ages 25 - 29 group has a mean of 5261 seconds with a standard deviation of 807 seconds, distribution normal.

Leo = 4948
Mary = 5513

a) Short-hand for these two normal distributions:

Men(30 - 34) = N(mu = 4313, stddev = 583)
Women(25-29) = N(mu = 5261, stddev = 807)

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

Z(Leo) = (4948 - 4313) / 583:

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

Z(Mary) = (5513 - 5261) / 807:

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

The Z scores tell us that Leo finished 1.089 standard deviations above the mean finish time for his gender/age group while Mary finished .312 standard deviations above the mean finish time for her gender/age group.

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

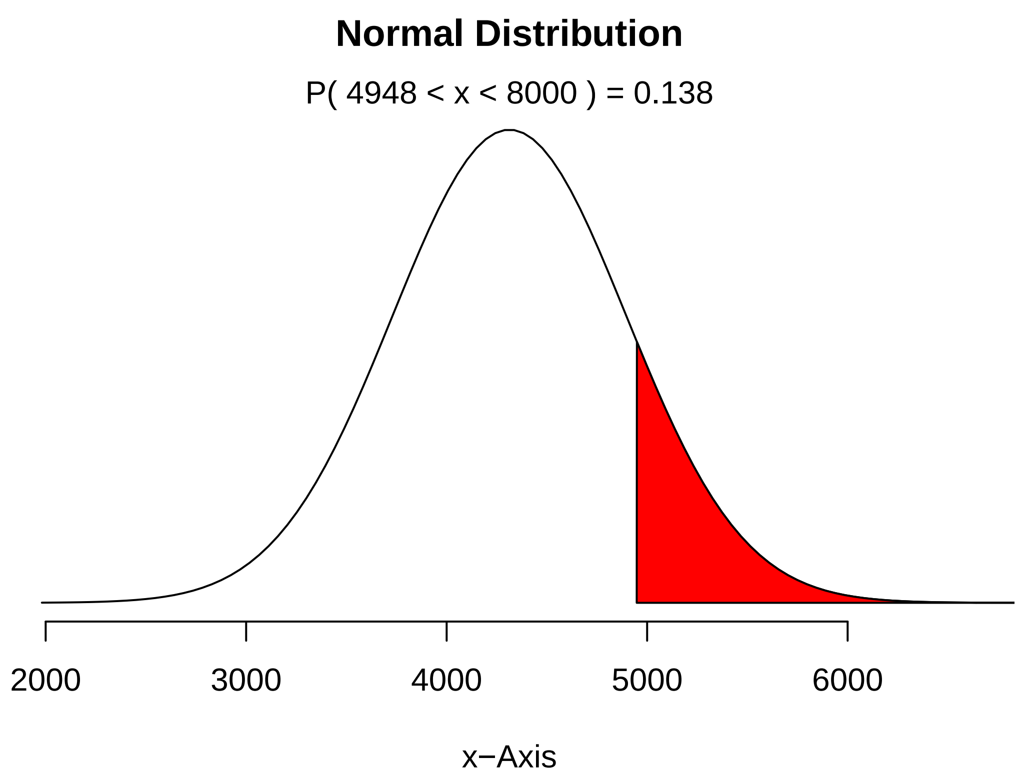
Since the Z scores represent difference between the finisher's time from the mean finishing time for a triathalon, a lower Z score is indicative of a relatively better finishing time within one's competitive group. In other words, the lower your Z score, the better you performed relative to your peer group. As such, Mary had a better finish time within her gender/age group than did Leo despite the fact that Leo had a faster overall finishing time.

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

```
1 - pnorm(4948, 4313, 583)
```

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

```
normalPlot(4313, 583, bounds = c(4948, 8000))
```



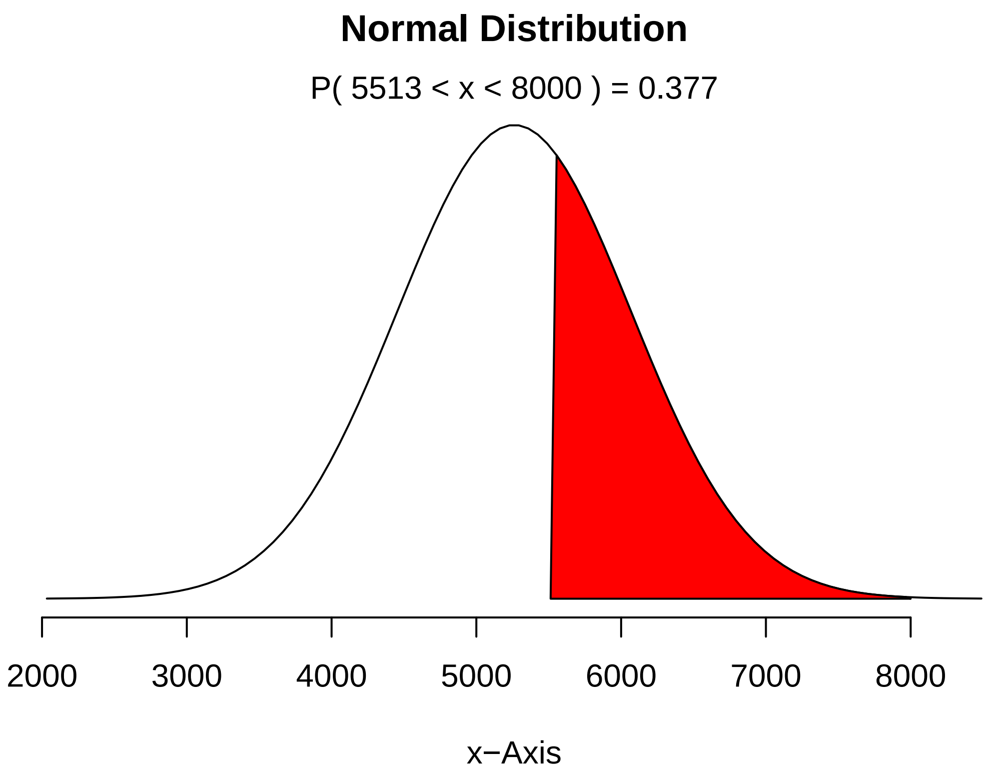
Leo finished ahead of 13.8% of triathletes in his group

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

```
1 - pnorm(5513, 5261, 807)
```

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

```r
normalPlot(5261, 807, bounds = c(5513, 8000))
```



Mary finished ahead of 37.7% of triathletes 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.

The answers to **b)** would not change because we can compute Z scores for non-normal distributions.

Similarly, the answer for **c)** would also remain valid since it is relying solely on the Z scores to conclude that Mary performed better within her group than did Leo.

However, the answers to **d)** and **e)** would not be discernible since we wouldn't be able to calculate percentiles using normal distribution assumptions.

3.18 (use qqnormsim from lab 3)

a) Is the 68, 95, 99 rule followed?

```r
fHgt <- 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(fHgt)
```

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

```r
fhMean <- mean(fHgt)
fhStd <- sd(fHgt)
```

.68 test

```r
pnorm( (fhMean + fhStd), fhMean, fhStd) - pnorm((fhMean - fhStd), fhMean, fhStd)
```

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

.95 test

```r
pnorm( (fhMean + (2 * fhStd)), fhMean, fhStd) - pnorm((fhMean - (2 * fhStd)), fhMean, fhStd)
```

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

.99 test

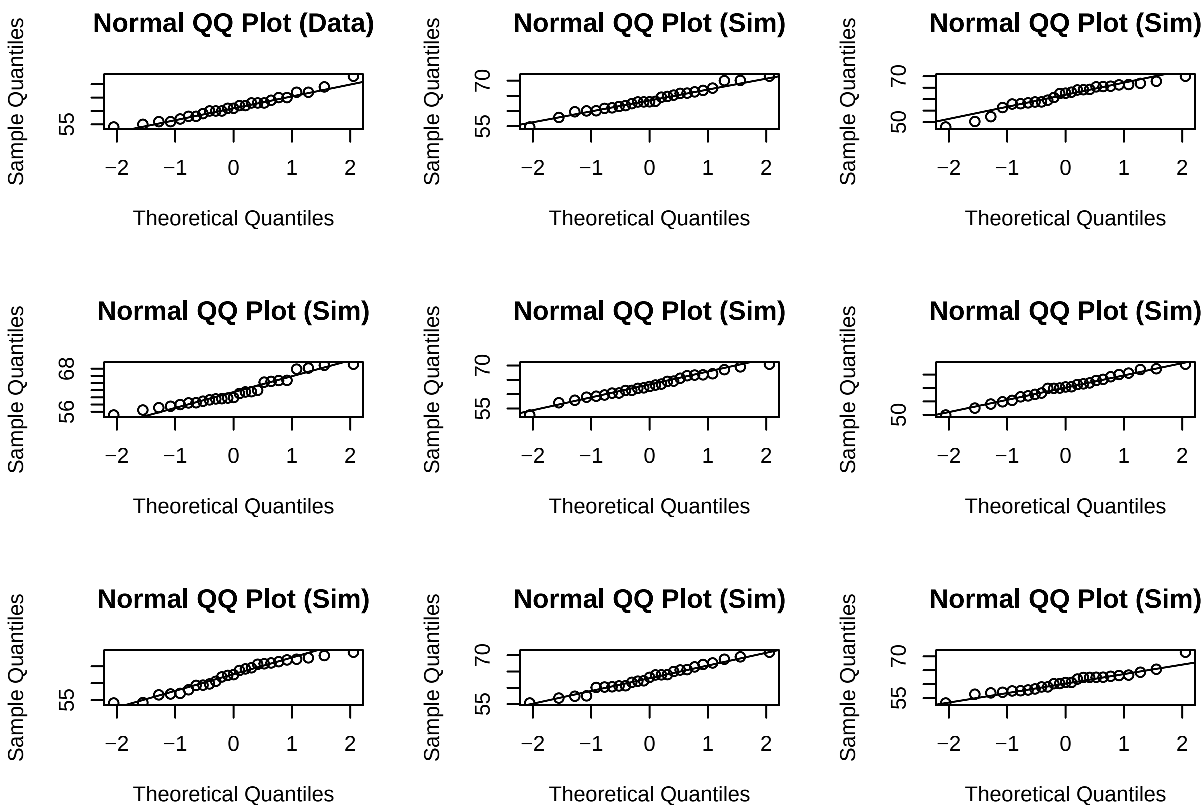
```r
pnorm( (fhMean + (3 * fhStd)), fhMean, fhStd) - pnorm((fhMean - (3 * fhStd)), fhMean, fhStd)
```

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

These results demonstrate that the heights approximately follow the 68-95-99.7% Rule.

b) Yes the data appear to follow a normal distribution. The distribution is approximately bell shaped as shown in the histogram and **qqnormsim** plot (see below) shows that the empirical data are reasonably close to fitting a the line that represents a theoretically ideal normal distribution. Additionally, as shown in the output of R's **summary** function above, the median value is approximately equal to the mean which indicates that we are not dealing with a meaningfully skewed distribution.

```r
qqnormsim(fHgt)
```



3.22 2% defect rate

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

```r
.98^9 * .02
```

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

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

```r
.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?

$1/\text{p} = 1/.02 =$

```r
1/.02
```

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

What is the standard deviation?

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

```r
sqrt( (1 - .02) / (.02^2) )
```

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

d) 5% defective: how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation?

1/p = 1/.05 =

```r
1/.05
```

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

What is the standard deviation?

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

```r
sqrt( (1 - .05) / (.05^2) )
```

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

e) Increasing the probability of an event means the event is less rare. This decreases both the mean and standard deviation of the wait time until success.

3.38

P(boy) = .51, couple plan to have 3 kids

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

We have 3 independent trials:

```r
n <- 3
k <- 2
p <- .51

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

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

So the probability of two of the three children being boys is .382347

b) Write out all possible orderings of 3 children / 2 boys:

bbg = .51 * .51 * .49
bgb = .51 * .49 * .51
gbb = .51 * .51 * .49

Probs = 3 * (.51 * .51 * .49)

```
3 * (.51 * .51 * .49)
```

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

The answers from 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).

The approach from part **b)** would require us to write out every possible combination of ways to have 3 boys out of a total of 8 children and then calculate their respective probabilities. The total number of possible combinations is:

```
choose(8,3)
```

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

Writing out such a large number of combinations would be very tedious and prone to errors. As such, it would be preferable to use the approach from part **a)** to solve this problem.

3.42

p = .15, events independent

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

```
n <- 10
k <- 3
p <- .15

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

Since the events are independent, the probability of any particular serve is .15. As such, the probability of her 10th serve being successful 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?

In part **a)** we were interested in a dependent scenario, specifically the case in which the server has previously had 2 successful serves out of 9 and we are interested in finding the probability that the 10th serve will be her third successful serve overall.

In part **b)** we are asked for the independent probability that that single serve will be successful.

As such, the two probabilities we were asked to find are not identical, with one representing the probability of a chain of events occurring and the other being the probability of a single event.