Homework 3

Daniel Dittenhafer

September 27, 2015

**3.2 Area under the curve Part II (p158)**

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 > -1.13 First, we define in `R` code the Z score and then use the `pnorm` function to determine the percentage on the left tail. Subtract this value from 1 to find the right tail value.

```
zGt <- -1.13
pGt <- 1 - pnorm(zGt)
pGt
```

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

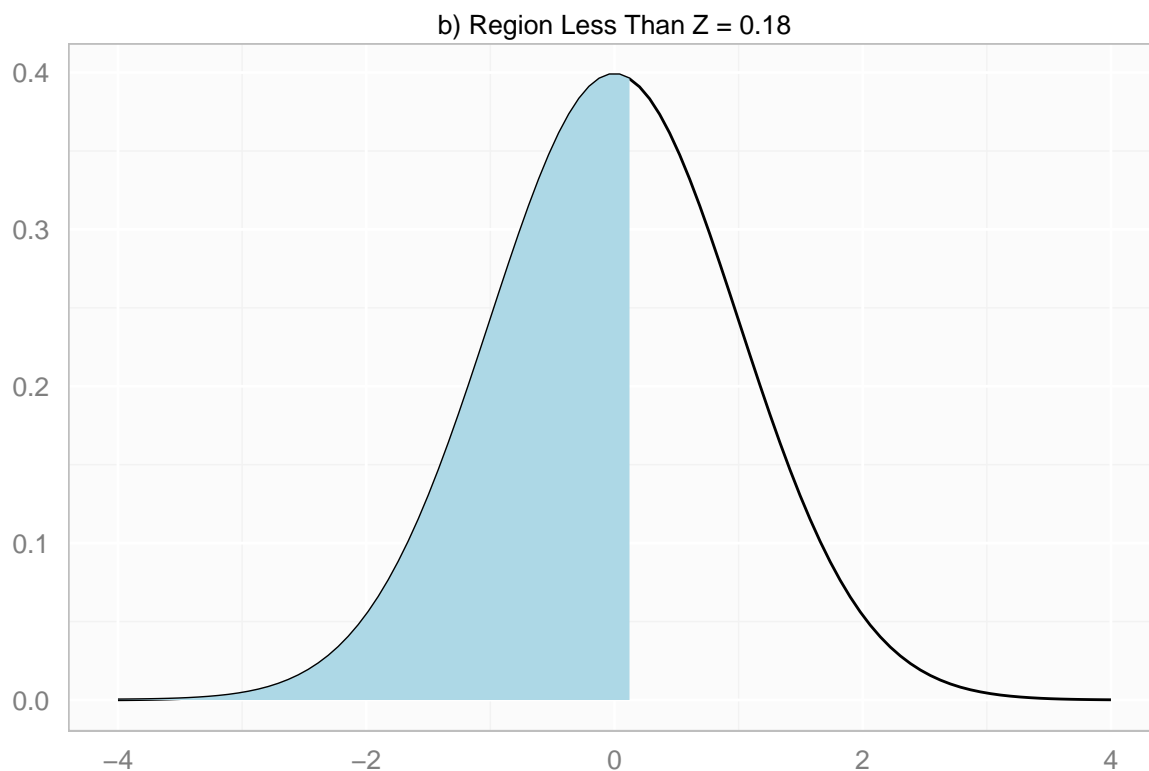The percent of the standard normal distribution found in the region Z > -1.13 is 0.8707619.

a) Region Greater Than Z = −1.13
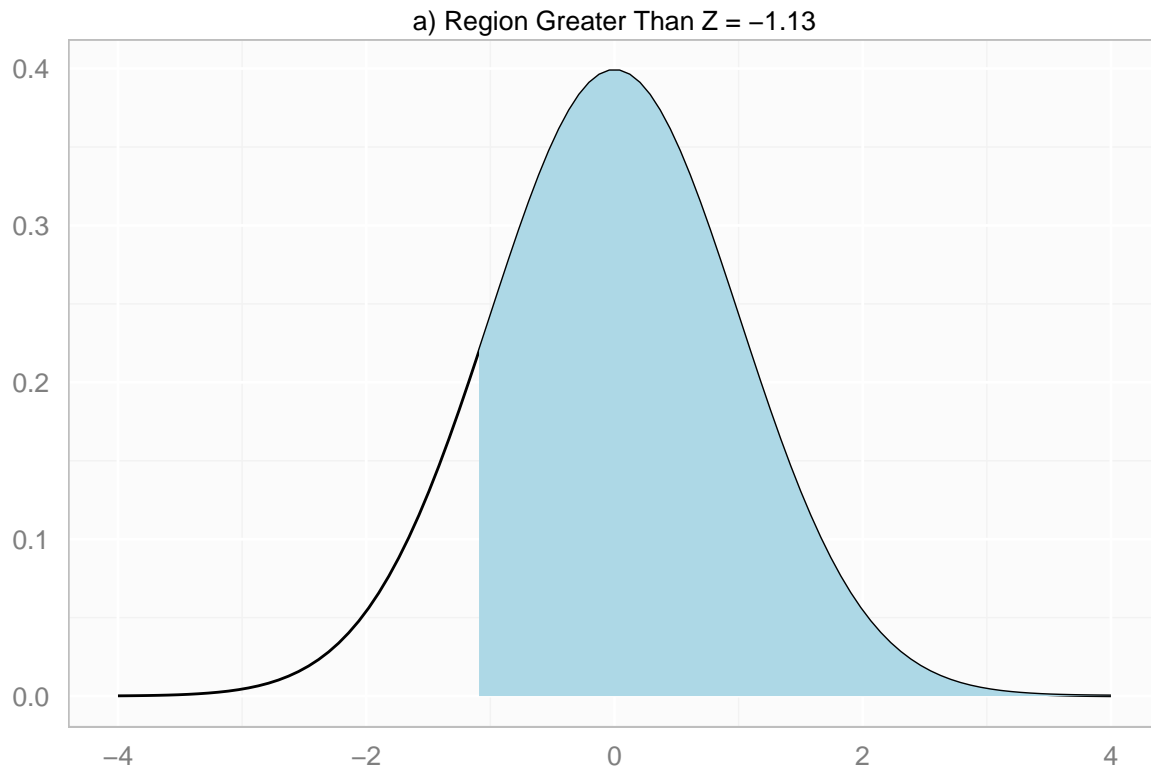
b. Z < 0.18   First, we define in R code the Z score and then use the `pnorm` function to determine the percentage on the left tail (less than).

```
zLt <- 0.18
pLt <- pnorm(zLt)
pLt
```

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

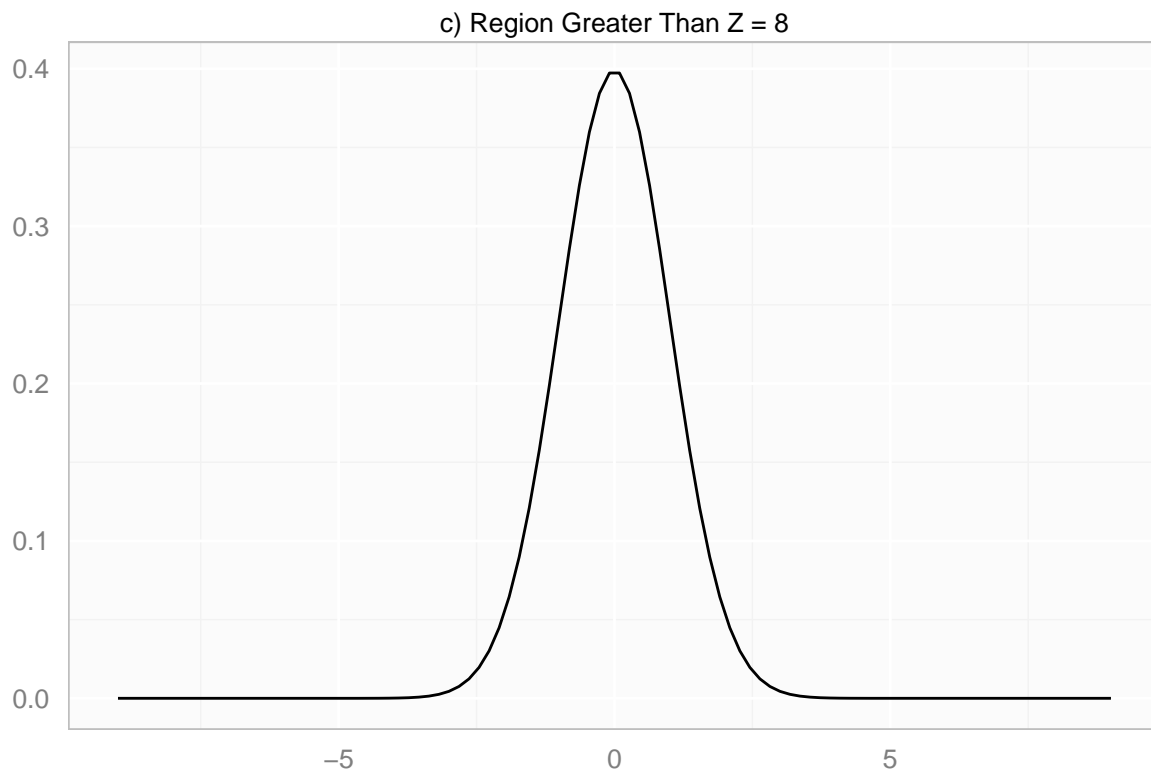The percent of the standard normal distribution found in the region Z < 0.18 is 0.5714237.

b) Region Less Than Z = 0.18

c. Z > 8   Again, we define in R code the Z score and then use the `pnorm` function to determine the percentage on the right tail (greater than).

```
zGt <- 8
pGt <- 1 - pnorm(zGt)
round(pGt, 4)
```

```
## [1] 0
```

The percent of the standard normal distribution found in the region Z > 8 is 0. This particular scenario is so extreme that it doesn't even show on the visualization.

c) Region Greater Than Z = 8

d. **|Z| < 0.5**   Again, we define in `R` code the Z score and then use the `pnorm` function to determine the percentage on the left tail. Due to the absolute value sign on the Z, we are looking for the middle region and subtract the pnorm value from 0.5. Also, this becomes a two tail-like question and we therefore double the value resulting from the `pnorm` subtraction.

```r
zAbs <- 0.5
pAbs <- 2 * (0.5 - pnorm(-1 * zAbs))
round(pAbs, 4)
```

```
## [1] 0.3829
```

The percent of the standard normal distribution found in the region |Z| < 0.5 is 0.3829.

d) Region |Z| < 0.5
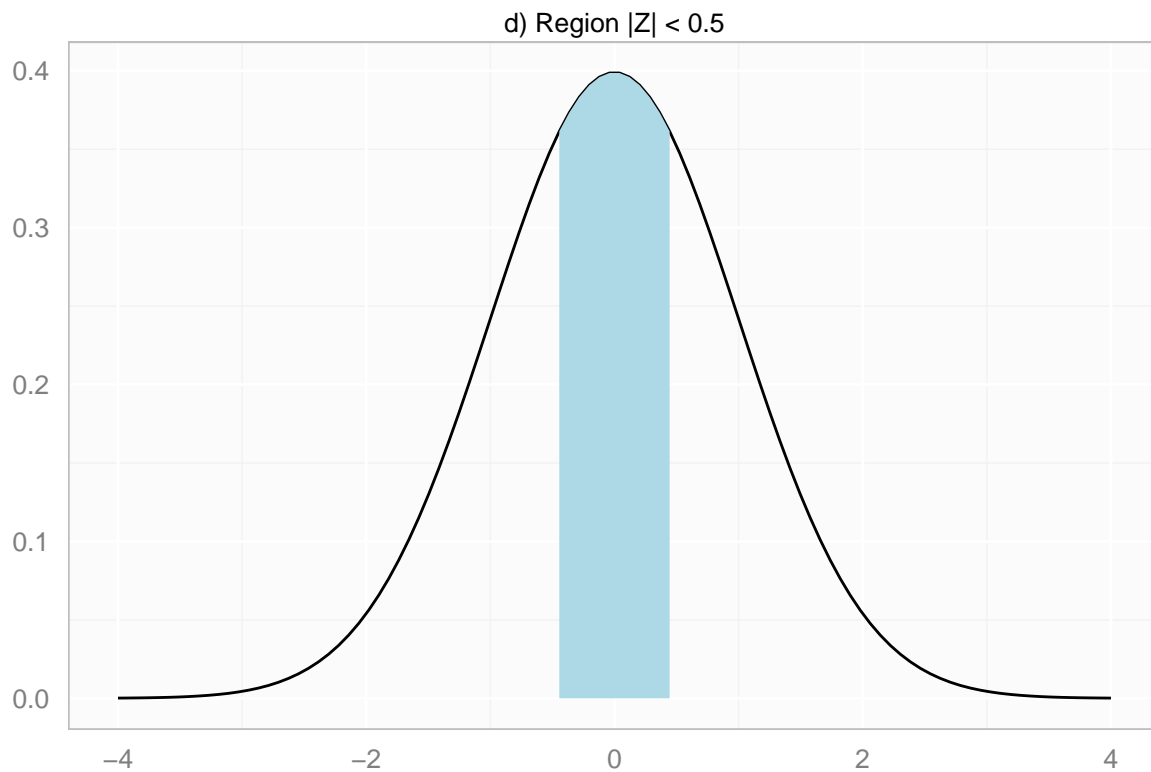
3.4 Triathlon times, Part I (p158)

Racer	Group	Time (sec)
Leo	Men, 30 - 34	4948
Mary	Women, 25 - 29	5513

Group	Mean	Stdev
Men, 30-34	4313	583
Women, 25-29	5261	807

- Normally distributed finishing times for both groups

a) Write down the short-hand for these two normal distributions. The short-hand for these two normal distributions follows:

Men, 30-34: $N(\mu = 4313, \sigma = 583)$

Women, 25-29: $N(\mu = 5261, \sigma = 807)$

b) What are the Z-scores for Leo's and Mary's finishing times? What do these Z-score tell you? Using R, we define the mean, standard deviation and individual times. Then we compute the Z score using the equation $Z = \frac{x - \mu}{\sigma}$ for Leo and Mary.

```
men3034mean <- 4313
men3034sd <- 583
leoTime <- 4948

leoZc <- (leoTime - men3034mean) / men3034sd
leoZc
```

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

```
women3034mean <- 5261
women3034sd <- 807
maryTime <- 5513

maryZc <- (maryTime - women3034mean) / women3034sd
maryZc
```

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

Leo's Z score is 1.09, and Mary's Z score is 0.31. These Z scores tell me how each of the participants faired in comparison to their respective groups.

C) Did Leo or Mary rank better in their respective groups? Explain your reasoning. Mary did better than Leo in their respective groups. Although they are both finishing above the mean time, Mary was much closer to the women's mean than Leo was to the men's mean. Another way to look at this is Mary was much closer to the left tail (lower Z equates to lower finishing time).

d) What percent of the triathletes did Leo finish faster than in his group? Using the `pnorm` function, we can determine the percent of triathletes who finished faster than Leo. By then taking the difference from 1 we can determine the percent of triathletes whom Leo's finished faster than.

```
pFasterThanLeo <- pnorm(leoZc)
pFasterThanLeo
```

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

```
pLeoFasterThan <- 1 - pFasterThanLeo
pLeoFasterThan
```

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

Leo finished faster than 13.8% of the triathletes in his group.

e) What percent of the triatheletes did Mary finish faster than in her group? Again, using the `pnorm` function, we can determine the percent of triathletes who finished faster than Mary. By then taking the difference from 1 we can determine the percent of triathletes whom Mary finished faster than.

```
pFasterThanMary <- pnorm(maryZc)
pFasterThanMary
```

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

```
pMaryFasterThan <- 1 - pFasterThanMary
pMaryFasterThan
```

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

Mary finished faster than 37.74% of the 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.   Most certainly the answers to parts (b) - (e) would change if the distribution of finishing times were not nearly normal. The Z scores and percentages are based on the area under the normal distribution curve. If the curve were not symmetric and were skewed to one side or the other, this would affect the area under the curve at any given finishing time value.

3.18 Heights of femal college students (p161)

```
heights <- c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,65,67,67,69,73)
dfHeights <- data.frame(heights)

meanHeight <- mean(heights)
meanHeight
```

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

```
sdHeight <- sd(heights)
sdHeight
```

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

a) Determine if the heights approximately follow the 68-95-99.7% Rule.   We can check the data versus the normal 68-95-99.7% Rule by selecting out the values that fall into each range, and then divide the number of rows in the range by the total number of rows. To do this, I've defined a function `percentBySd` that we can use as a helper to compute the percentages through repeated calls.

```
percentBySd <- function(data, numSd)
{
  m <- mean(data)
  s <- sd(data)

  sd1Lower <- m - (s * numSd)
  sd1Upper <- m + (s * numSd)

  sdData <- data[sd1Lower < data & data < sd1Upper]
  pSdData <- length(sdData) / length(data)
  return (pSdData)
}
# 1 Standard Deviation
sd1 <- percentBySd(heights, 1) * 100
# 2 Standard Deviation
```

```
sd2 <- percentBySd(heights, 2) * 100
# 3 Standard Deviation
sd3 <- percentBySd(heights, 3) * 100
sdList <- c(sd1, sd2, sd3)
```

It does appear that the height data basically follows the 68-95-99.7% Rule: 68-96-100.

```
## [1]  68  96 100
```

b) Do these data appear to follow the normal distribution? Explain your reasoning usign the graphs provided below. Based on the graphs in the text, the data roughly follows the normal distribution, but with a bit of skew. The best fit normal curve on the histogram is not a perfect fit, but it could be worse. Likewise, the normal probability plot generally follows the normal line, deviating mostly at the upper end.
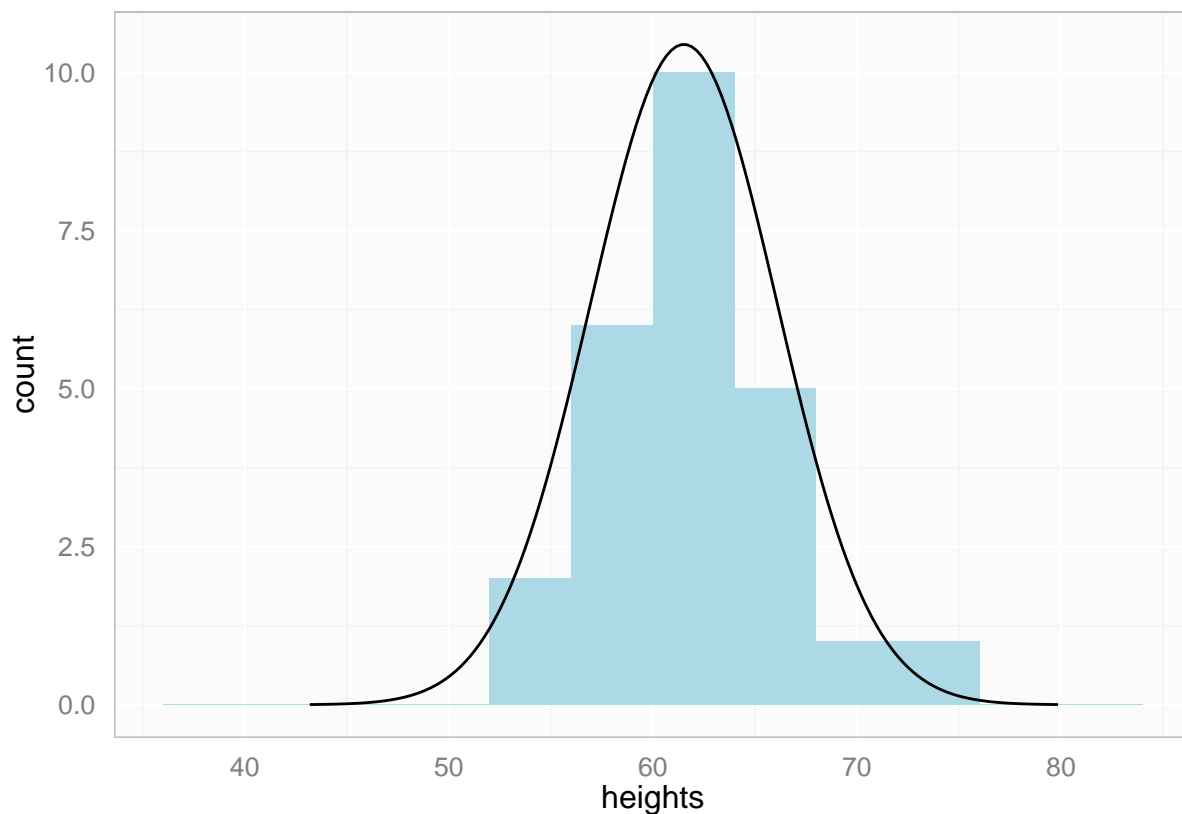
3.22 Defective Rate (p162)

- 2% defective rate
- Production is a random process
- Each widget is independent of the others

a) What is the probability that the 10th transistor produced is the first with a defect? Following the equation given on p143 of the text, $p_n = (1 - p)^{n-1} \times p$, written in R code, we have:

```
defectRate <- 0.02
successRate <-  1 - defectRate
n <- 10

p10 <- successRate^(n-1) * defectRate
p10
```

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

The probability of the 10th transistor produced is the first with a defect is 0.0167.

b) What is the probability that the machine produces no defective transistors in a batch of 100?   We could look at this as asking what the probability is that a defective transitor will be produced in a batch of 100, and then take the complement. This would be $P(n=1) + P(n=2)$ ...

```
p100D <- 0
for(i in 1:100)
{
  p100D <- p100D + (successRate^(i-1) * defectRate)
}
p100C <- 1 - p100D
```

The probability that the machine produces no defector transistors in a batch of 100 is 0.1326.

c) On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?   Using the equation from p143 of the text, the follow R code computes the expected value and standard deviation:

```
expectedVal <- 1 / defectRate
expectedVal
```

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

```
stdevDef <- sqrt( (1 - defectRate) / defectRate^2 )
stdevDef
```

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

I would expect 50 transistors to be produced before the first with a defect. The standard deviation is 49.4974747.

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?   Using the new defective rate, the follow computations, in R help use determine the expected value and standard deviation:

```
defectRate <- 0.05
successRate <-  1 - defectRate
```

```
expectedVal <- 1 / defectRate
expectedVal
```

[1] 20

```
stdevDef <- sqrt( (1 - defectRate) / defectRate^2 )
stdevDef
```

[1] 19.49359

I would expect 20 transistors to be produced before the first with a defect. The standard deviation is 19.4935887.

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 an event decreases the mean and standard deviation of the wait time until the event.

3.38 Male Children (p166)

3.42 Serving in Volleyball (p166)