Homework 4

Irene Jacob

2020-09-27

Chapter 4 - Distributions of Random Variables

Area under the curve, Part I.

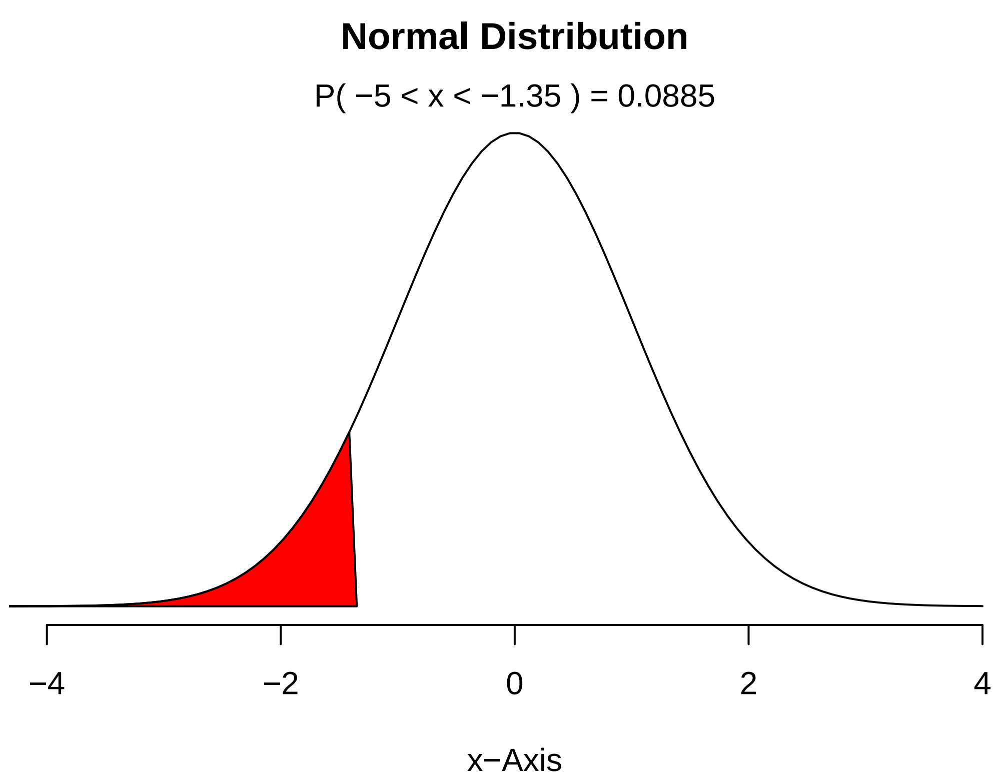
What percent of a standard normal distribution N(= 0, = 1) is found in each region? Be sure to draw a graph.

(a) $Z < -1.35$

```
pnorm(-1.35)
```

```
## [1] 0.08850799
```

```
normalPlot(bounds = c(-5,-1.35))
```

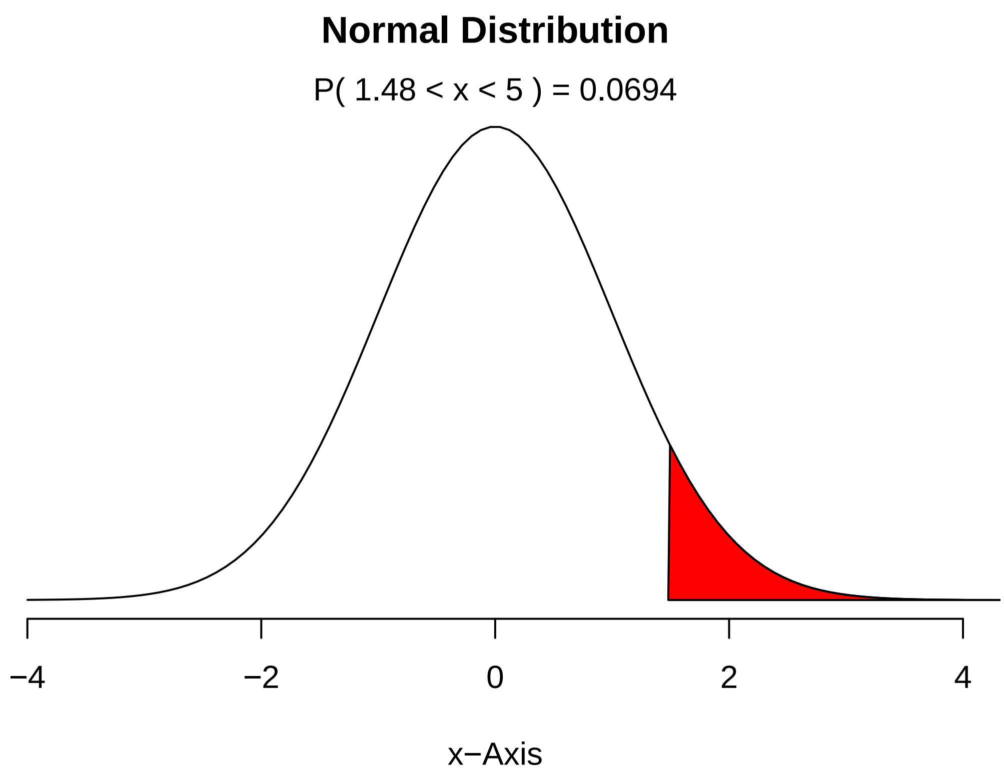


(b) $Z > 1.48$

```r
1-pnorm(1.48)
```

```
## [1] 0.06943662
```

```r
normalPlot(bounds = c(1.48,5))
```

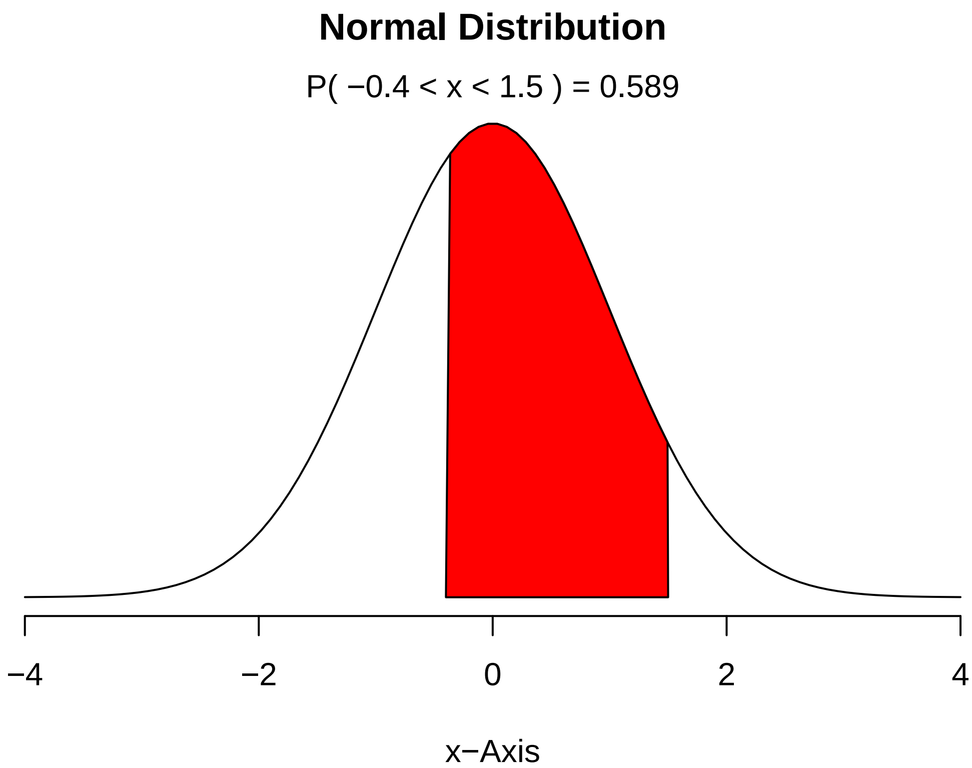


(c) $-0.4 < \mathrm{Z} < 1.5$

```r
pnorm(1.5)-pnorm(-0.4)
```

```
## [1] 0.5886145
```

```r
normalPlot(bounds = c(-0.4,1.5))
```



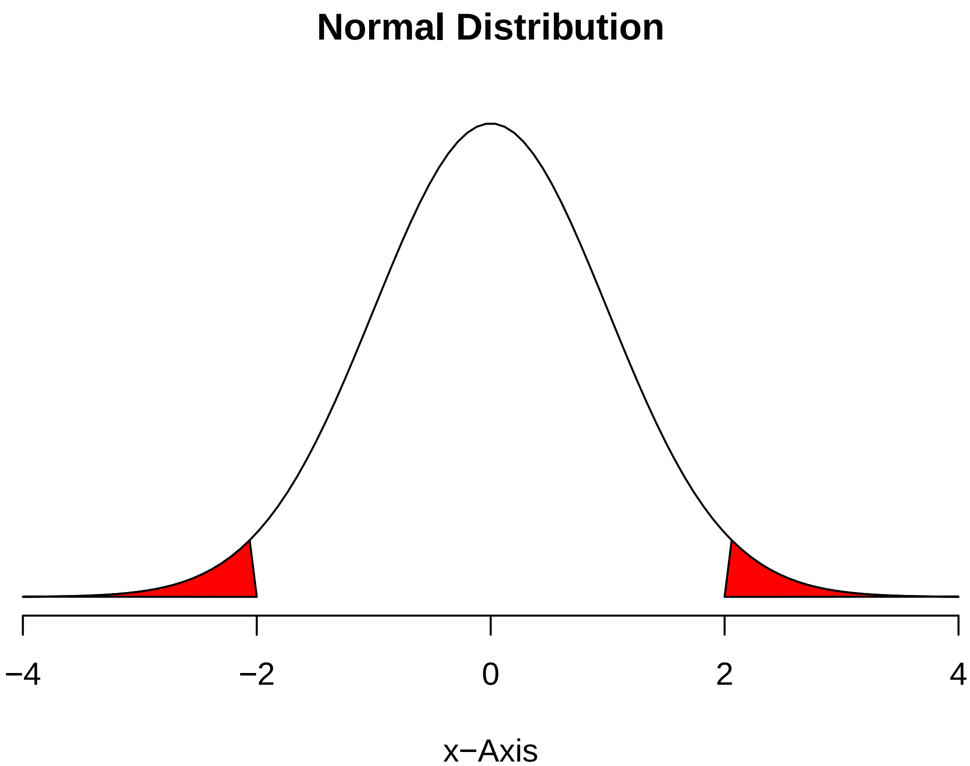
(d) $|Z| > 2$

when modulus is involved it means Z > 2 or Z < -2.

```r
pnorm(2)+pnorm(-2)
```

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

```r
normalPlot(bounds = c(-2,2), tails = TRUE)
```



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 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.

Leo's z score is as follows

```
m_mean <- 4313
m_sd <- 583
leo <- 4948
z_leo <- (leo - m_mean)/m_sd
z_leo
```

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

Mary's z score is as follows

```
w_mean <- 5261
w_sd <- 807
mary <- 5513
z_mary <- (mary-w_mean)/w_sd
z_mary
```

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

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

From the z scores of LEO and MARY displayed below it is clear that MARY performed better.

```
z_leo
```

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

```
z_mary
```

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

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

LEO's z score is 1.089 whereas MARY's is 0.31 so MARY performed better when compared to LEO in their respective groups

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

```
1-pnorm(z_leo)
```

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

From this it is clear that LEO finished faster than 13.8% of his group.

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

```
1-pnorm(z_mary)
```

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

From this it is clear that MARY finished faster than 37.7% than her group

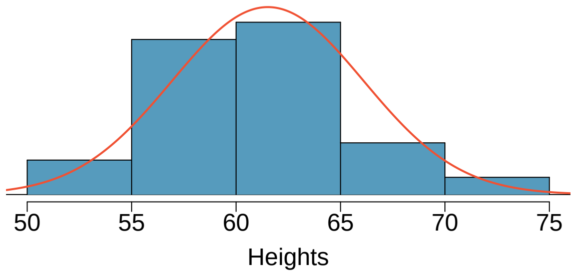
(f) If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

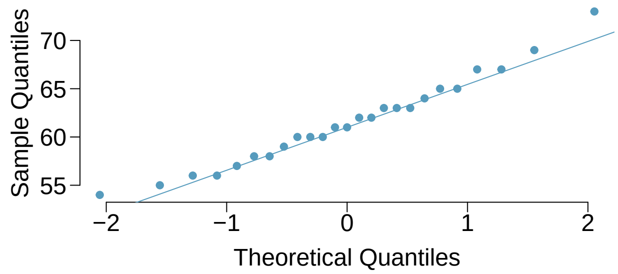
If the distribution was not normal the values will change for the above answers as z score cannot be used on distributions that are not normal.

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.

```
m = mean(heights)
sd = sd(heights)

length(heights[x>=m-sd & x<=m+sd]) / length(heights)          #68
```

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

```
length(heights[x>=m-2*sd & x<=m+2*sd]) / length(heights)      #95
```

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

```
length(heights[x>=m-3*sd & x<=m+3*sd]) / length(heights)      #99
```

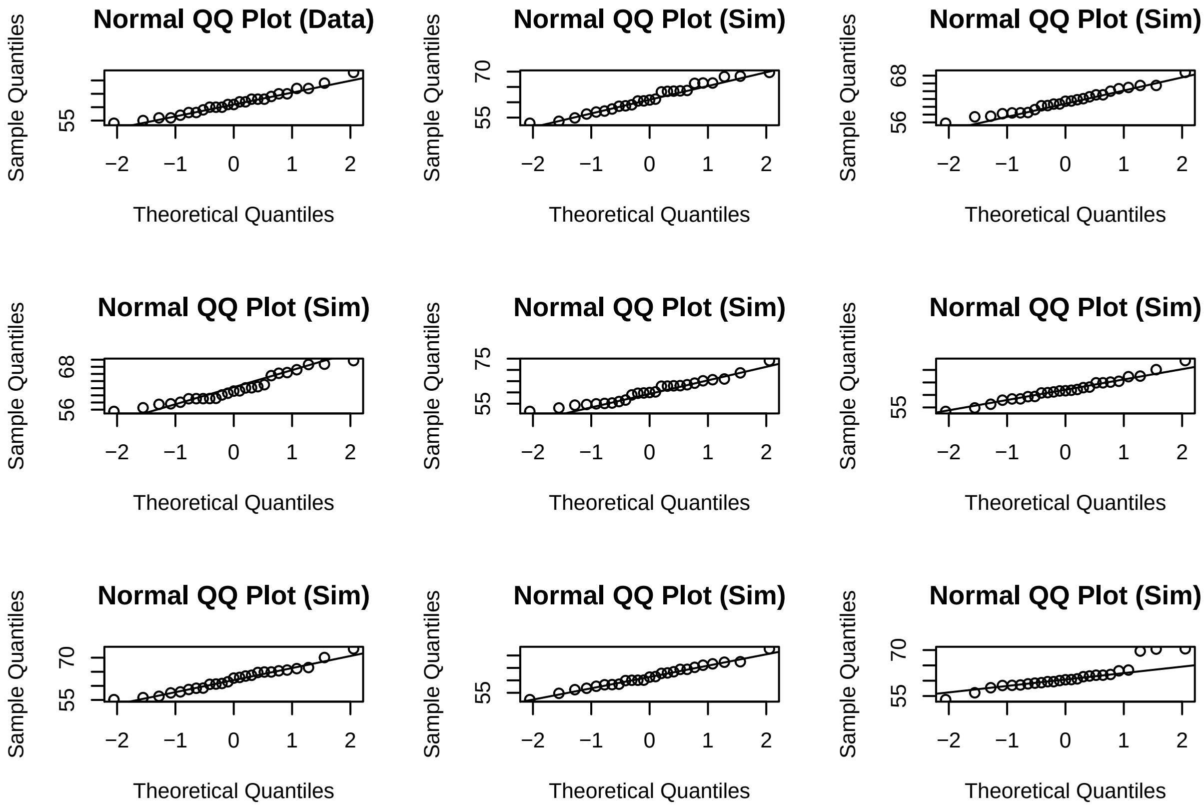
```
## [1] 1
```

From the above results it is clear that it follows 68-95-99.7% Rule.

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

From the QQ plot we can see that almost all the data lies near the central line and the histogram follows the overlaying curve. So the distribution is normal.

```
qqnormsim(heights)
```



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?

```
n <- 10
p <- 0.02
q <- 1-p

Prob_n <- (q^(n-1))*p
Prob_n
```

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

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

```
n <- 100

Prob_n <- q^n
Prob_n
```

```
## [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
m <- 1/p
sd <- sqrt(q/(p^2))
print( paste("Transistors produced before the first with a defect:", m))
```

```
## [1] "Transistors produced before the first with a defect: 50"
```

```r
print( paste("Standard Deviation:", sd))
```

```
## [1] "Standard Deviation: 49.4974746830583"
```

(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
p1 <- 0.05
q1 <- 1-p1

m1 <- 1/p1
sd1 <- sqrt(q1/(p1^2))
print( paste("Transistors produced before the first with a defect:", m1))
```

```
## [1] "Transistors produced before the first with a defect: 20"
```

```r
print( paste("Standard Deviation:", sd1))
```

```
## [1] "Standard Deviation: 19.4935886896179"
```

(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 defect percentage increases the number of transistors proced before a defective one is produced decreases. The standard deviation is largely reduced as well.

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.

```r
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.

P(B)∗P(B)∗P(G)+P(B)∗P(G)∗P(B)+P(G)∗P(B)∗P(B)=0.51*0.51*0.49+0.51*0.49*0.51+0.49*0.51*0.51

```r
0.51*0.51*0.49+0.51*0.49*0.51+0.49*0.51*0.51
```

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

The answers in (a) and (b) are same.

(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).

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

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

There are 56 combinations so it is not easy to use the method used in (b)

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?

```r
dnbinom(7,3,0.15)
```

```
## [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?

Her serves are independent of each other so the probability of 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?

(d) is about probability of a sequence of events whereas (b) is about a single event