Chapter 4 - Distributions of Random Variables

Gabriel Campos

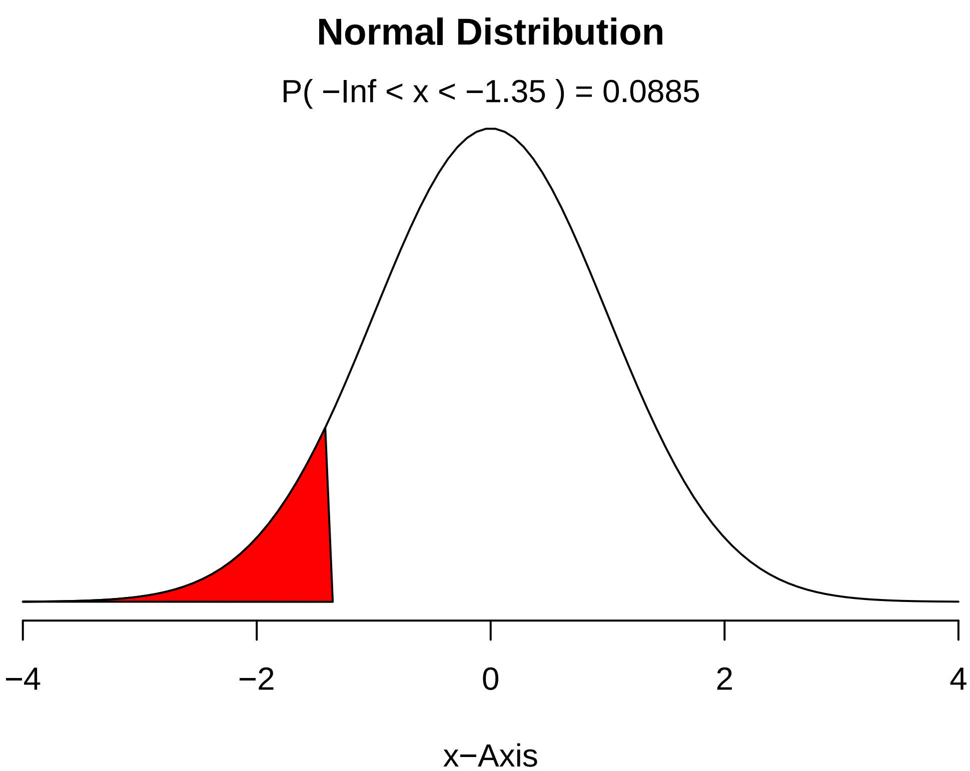
Question 1

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) $Z < -1.35$
(b) $Z > 1.48$
(c) $-0.4 < Z < 1.5$
(d) $|Z| > 2$

(a)

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

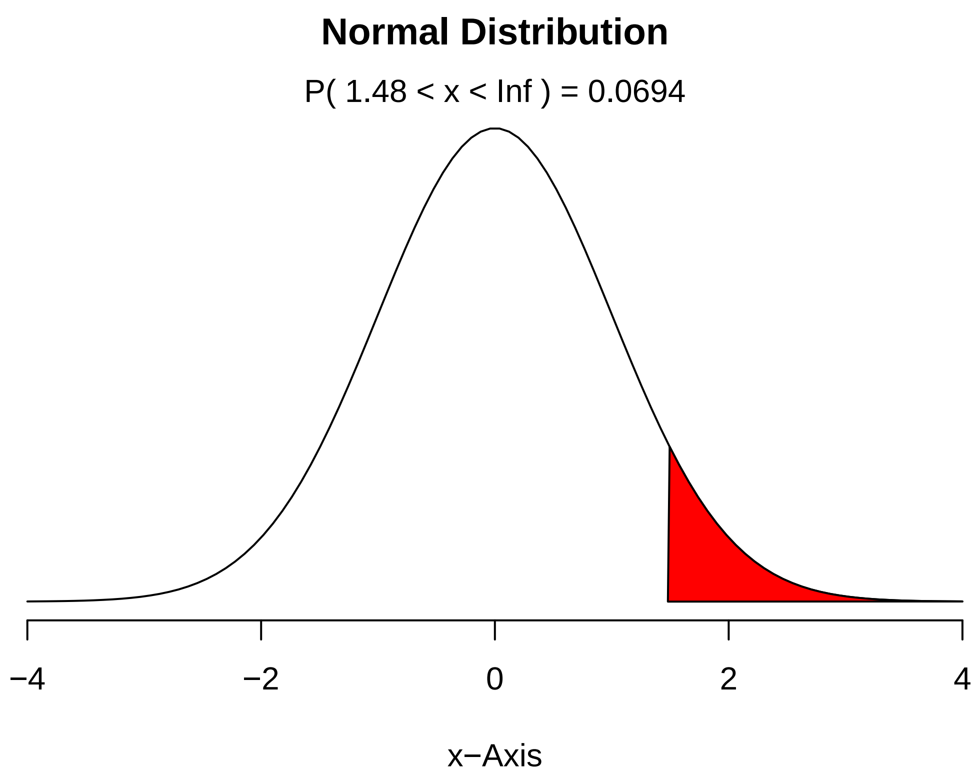


(b)

```r
#$Z > 1.48$
1-pnorm(1.48)
```

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

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

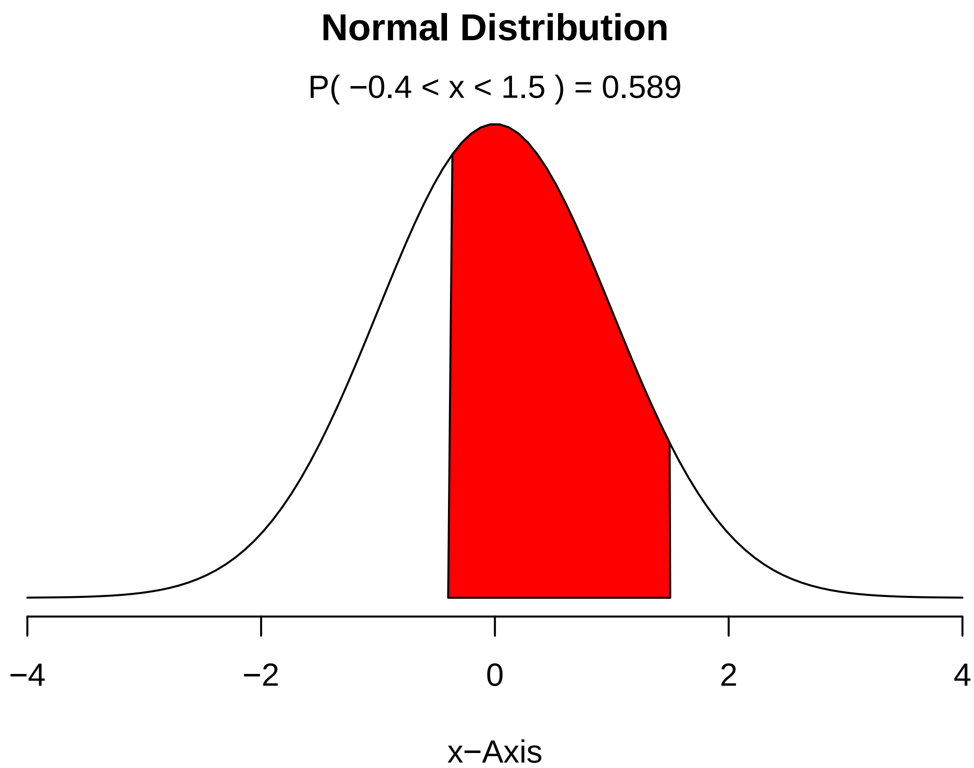


(c)

```r
#$-0.4 < Z < 1.5$
pnorm(1.5)-pnorm(-0.4)
```

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

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

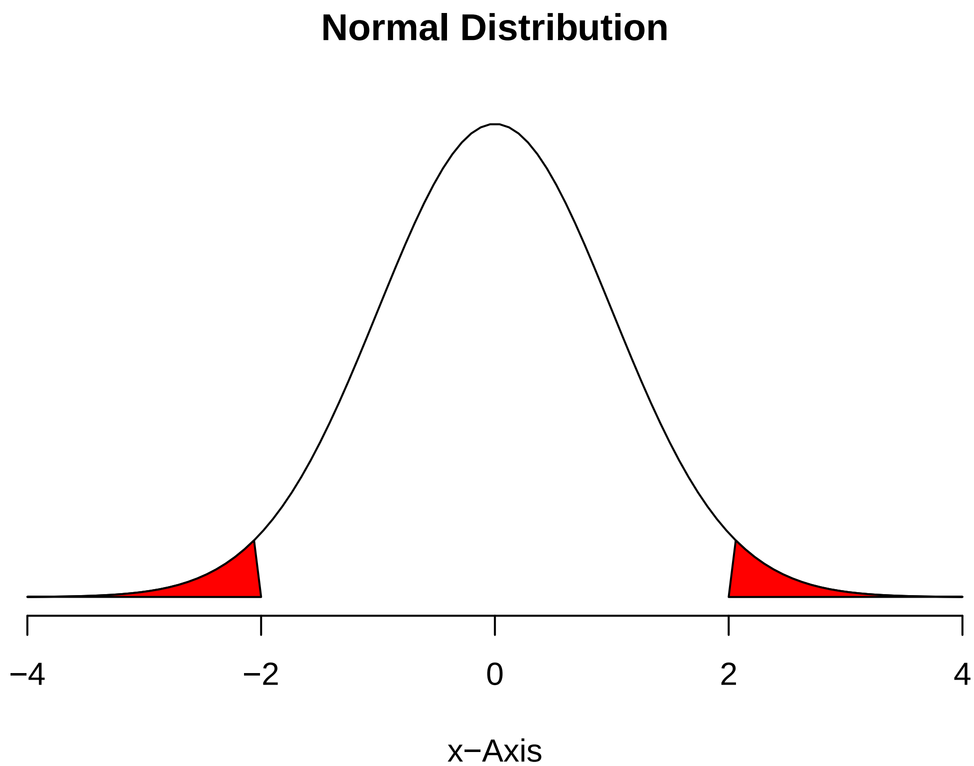


(d)

```
#$|Z| > 2$
pnorm(-2)+(1-pnorm(2))
```

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

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


Normal Distribution
−4
−2
0
2
4
x−Axis

Question 2

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 run times ages 30-34) M: N(μ=4313, σ=583)

(b)

What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?
((Leo or Mary time in seconds)-μ) / σ

```r
#Leo_Z
(4948 - 4313)/583
```

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

```r
#Mary_Z
(5513 - 5261)/807
```

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

** Scores show the standard deviation of each friend (Leo or Mary) with their respective groups mean**

(c)

Did Leo or Mary rank better in their respective groups? Explain your reasoning.
Mary's lower Z-score means she does not deviate as much as Leo. Leo's higher Z-score means that fewer people ran slower than him, making him the slower runner with respect to groups of the two

(d)

What percent of the triathletes did Leo finish faster than in his group?
Leo finished faster that 14% of his group

```r
pnorm(1.089194,lower.tail=FALSE)
```

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

(e)

What percent of the triathletes did Mary finish faster than in her group?
Mary ran faster than 38% of her group

```r
pnorm(0.3122677,lower.tail=FALSE)
```

```
## [1] 0.3774185
```

(f)

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

Our Z-scores will remain the same but without a nearly normal distribution we would not be able to calculate the percentiles with them

---

Question 3

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. **68% the data falls within +/- 1 standard deviation from mean**
95% the data falls within +/- 2 standard deviation from mean
99.7% the data falls within +/- 3 standard deviation from mean

```
# 68%
pnorm((61.52+4.58),61.52,4.58)-pnorm((61.52-4.58),61.52,4.58)
```

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

```
#95%
pnorm((61.52+4.58+4.58),61.52,4.58)-pnorm((61.52-4.58-4.58),61.52,4.58)
```

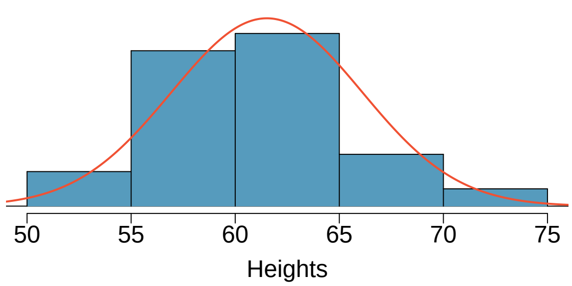
```
## [1] 0.9544997
```

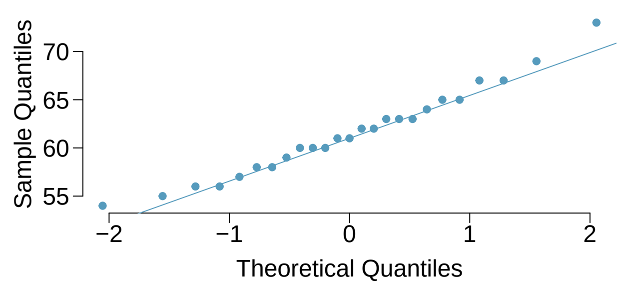
```
#99.7%
pnorm((61.52+4.58+4.58+4.58),61.52,4.58)-pnorm((61.52-4.58-4.58-4.58),61.52,4.58)
```

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

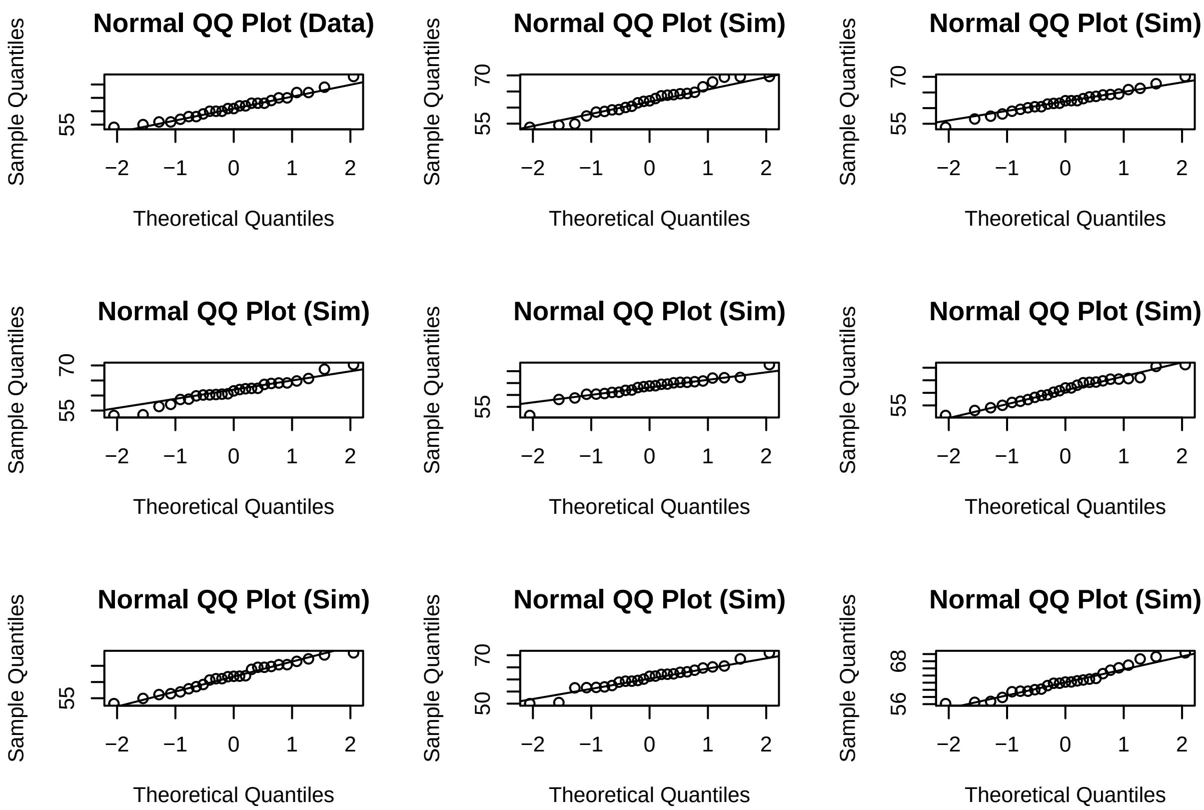
(b)

Do these data appear to follow a normal distribution? Explain your reasoning using the graphs provided below.
Fluctuation of values is minimal. The histogram closely follows the normal density overlay and the scatter plot (with exception of one point in the tail) is very close to maintaining a straight line. all of this supports that the distribution is very nearly normal





```r
# Use the DATA606::qqnormsim function
#?qqnormsim
qqnormsim(heights)
```



Question 4

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?

This can also be viewed as *what is the probability that the first 9 transistors do not have defects?*:

```
# probability of now defect 1-P(defects)
defects<- 0.02
no_defects<-(1-defects)
trials<-10
# nine events of no defects and on event with a defect
(no_defects^(trials-1))*defects
```

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

(b)

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

```
trials<-100
no_defects^trials
```

```
## [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?
Expected value is 1/p where p=P(defects)

```
1/defects
```

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

```
# sd = sqrt((1-p)/p^2
sd<-sqrt((1-defects)/defects^2)
sd
```

```
## [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?

```
another_defect<-0.05
#expected value
1/another_defect
```

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

```
another_sd<-sqrt((1-another_defect)/another_defect^2)
another_sd
```

```
## [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?
Increasing the likelihood of an event happening decreases the standard deviation and expected value. This means the first occurrence of failure is more likely to happen which makes sense

---

Question 5

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.

```
boys<-.51
girls<-(1-boys)
n_children<-3
n_boys<-2
#??dbinom
dbinom(n_boys,n_children,boys)
```

```
## [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.
GBB BGB BBG

```
(girls*boys*boys)+(boys*girls*boys)+(boys*boys*girls)
```

```
## [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).
Part (b)'s approach would be tedious in that it would require the 56 possible combinations available for the 8 child 3 boy scenario. Typing out each combination is less favorable than using the dbinom() function.

Question 6

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?

```
# using p for probability of success q probability of failures
# n for number of trys n_p for number of successes
p<- .15
q<- (1-p)
n<- 10
n_p<-3
dnbinom(n-n_p,n_p,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?
Independent events mean the probabilities of failure and success are not impacted by a previous success or failure. Therefore the probability will be 0.15 that the 10th serve will be successful

(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?
Basically the individual event's probability will remain the same, however a certain sequence or combination of events will be calculated taking in to account the likelihood of previous and future events within that sequence.