Chapter 4 - Distributions of Random Variables

Samuel Iguku Kigamba

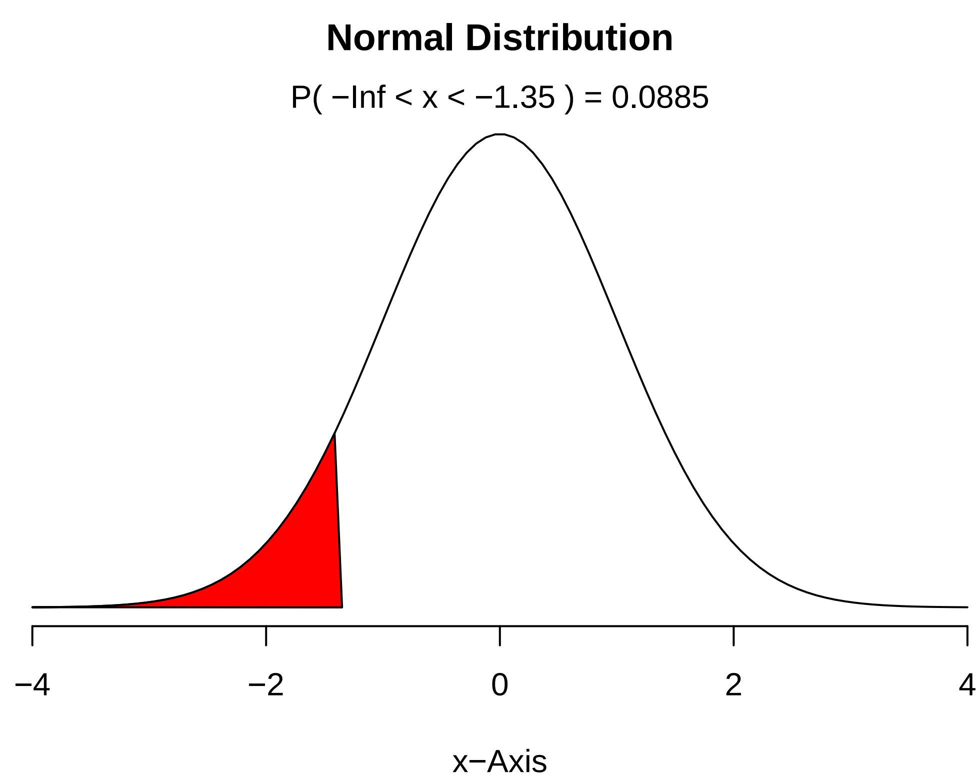
Area under the curve, Part I. (4.1, p. 142) What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

(a) $Z < -1.35$

```
# use the DATA606::normalPlot function
mu <- 0
sd <- 1
Z <- -1.35
#find x
x <- Z * sd + mu
pnorm(x, mean = 0, sd = 1)
```

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

```
normalPlot(mean = 0, sd = 1, bounds = c(-Inf, x), tails = FALSE)
```

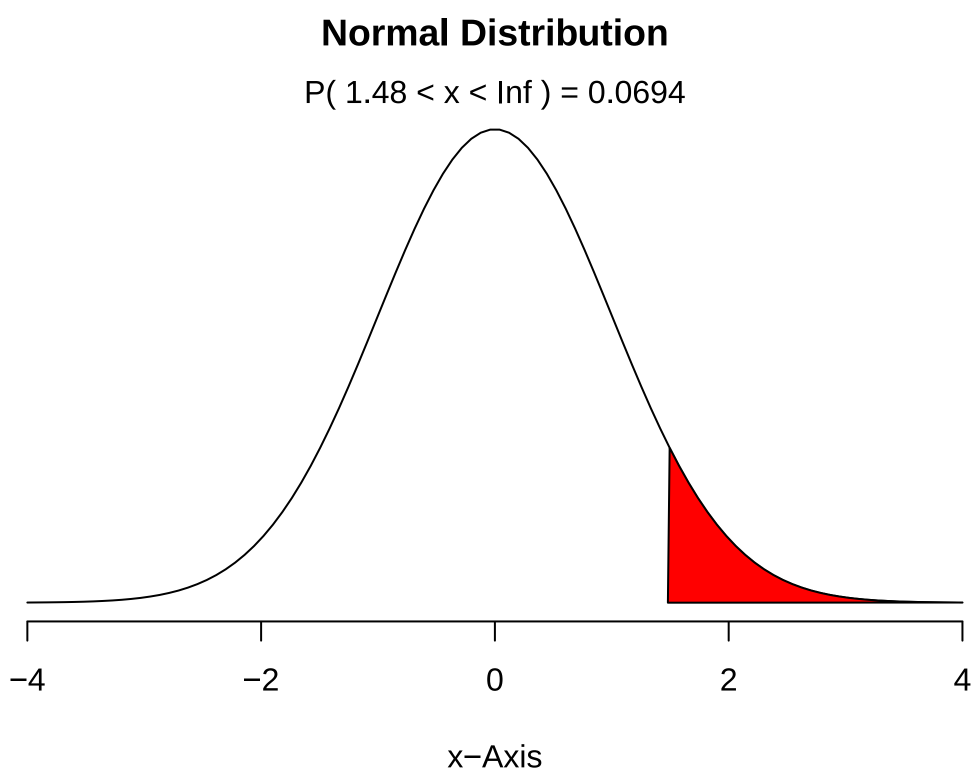


(b) $Z > 1.48$

```r
# use the DATA606::normalPlot function
mu <- 0
sd <- 1
Z <- 1.48
#find x
x <- Z * sd + mu
1 - pnorm(x, mean = 0, sd = 1)
```

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

```r
normalPlot(mean = 0, sd = 1, bounds = c(x, Inf), tails = FALSE)
```



(c) $-0.4 < Z < 1.5$

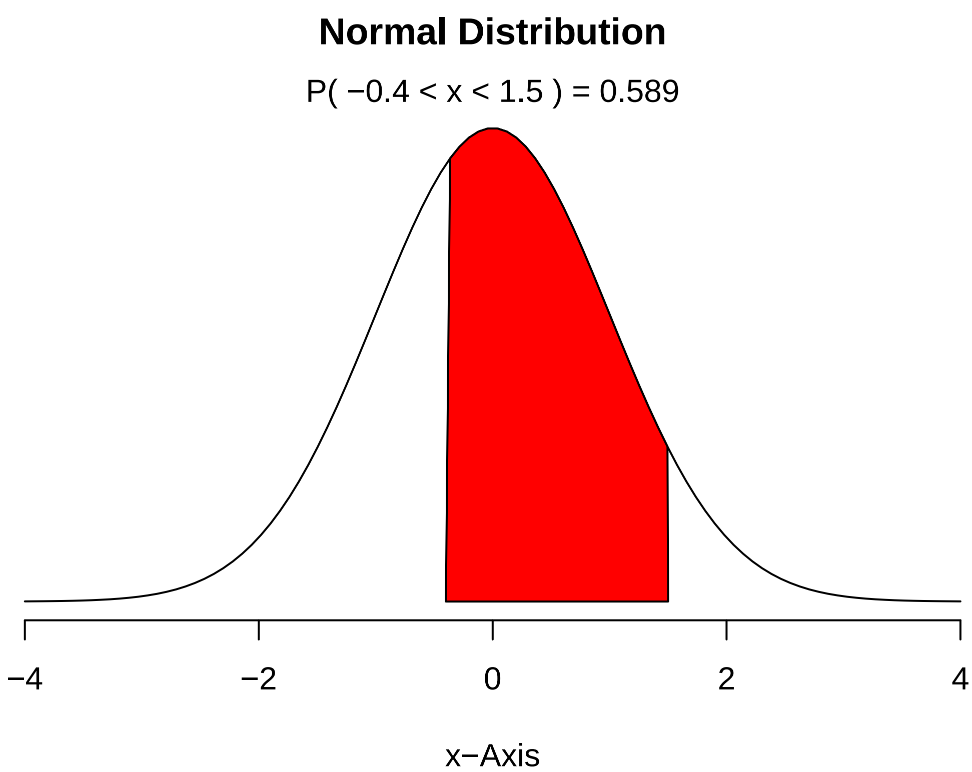
```r
# use the DATA606::normalPlot function
mu <- 0
sd <- 1
Z1 <- -0.4
Z2 <- 1.5
#find x1 and x2
x1 <- Z1 * sd + mu
x2 <- Z2 * sd + mu

p1 <- pnorm(x1, mean = 0, sd = 1)
```

```r
p2 <- pnorm(x2, mean = 0, sd = 1)
p2 - p1
```

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

```r
normalPlot(mean = 0, sd = 1, bounds = c(x1, x2), tails = FALSE)
```



(d) $|Z| > 2$

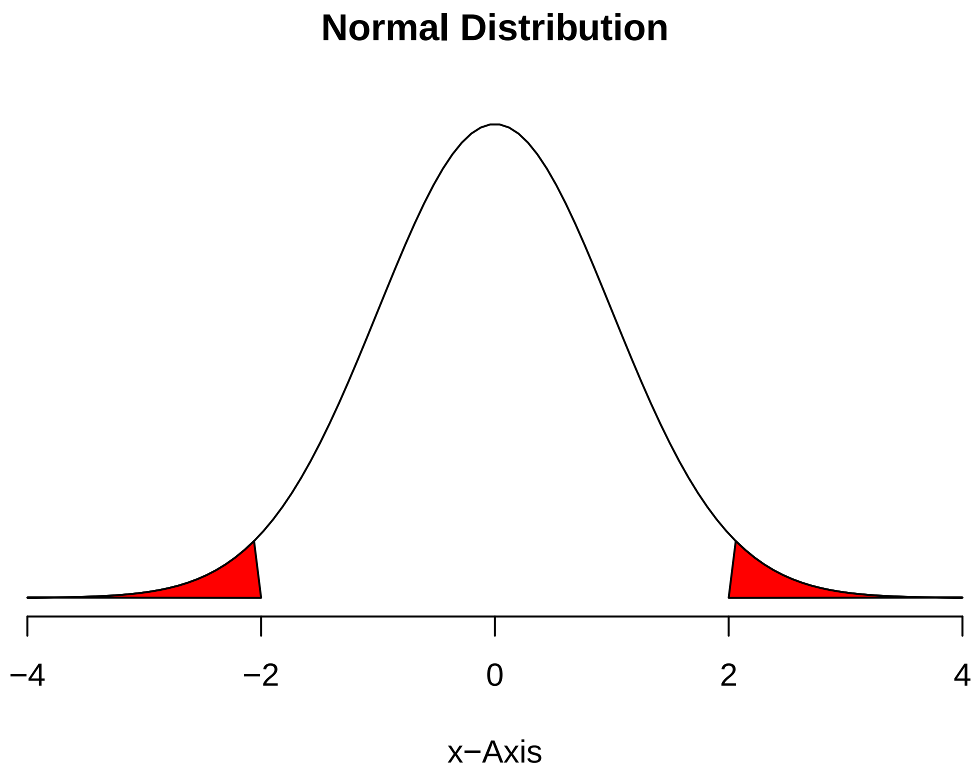
```r
mu <- 0
sd <- 1

pnorm(-2, mean = 0, sd = 1) + 1 - pnorm(2, mean = 0, sd = 1)
```

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

```r
#Distribution of 4.55%

normalPlot(mean = 0, sd = 1, bounds = c(-2, 2), tails = TRUE)
```


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

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_30_34: N(µ = 4313, ?? = 583) Women_25_29: N(µ = 5261, ?? = 807)

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

Z score is an indivator of how much an observation deviates from the mean (by how many standard deviations)

```
Z_Leo <- (4948-4313)/583
Z_Leo #Leo is 1.09 sds slower than the mens mean for his specific group.
```

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

```
Z_Mary <- (5513-5261)/807
Z_Mary #Mary is 0.31 sds slower than the womens mean for her specific group.
```

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

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

```
1 - pnorm(4948, mean = 4313, sd = 583) # Leo performed better than 13.8% of his group
```

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

```
1 - pnorm(5513, mean = 5261, sd = 807) # Mary performed better than 37.7% of her group.
```

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

Solution: Mary performed better than Leo, since she performed better than 37.7% of her group while Leo only outperformed 13.8%.

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

```
Leo performed better than 13.8% of his group. see workings above.
```

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

```
Mary performed better than 37.7% of her group. See workings above.
```

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

```
The findings under b would reamain the same,but all others would change. b relates to an individual
the two results for both Leo and Mary would not be comparable due to different distrubutions.
```

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.

`The heights of the 25 female college students approximately follow the 68-95-99.7% Rule.`

```r
pnorm(66.1, mean = 61.52, sd = 4.58) - pnorm(56.94, mean = 61.52, sd = 4.58) # 1 sd
```

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

```r
pnorm(70.68, mean = 61.52, sd = 4.58) - pnorm(52.36, mean = 61.52, sd = 4.58) # 2 sd
```

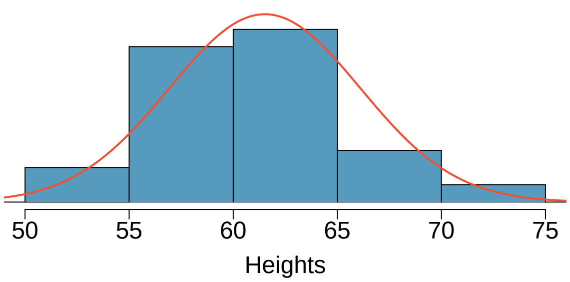
```
## [1] 0.9544997
```

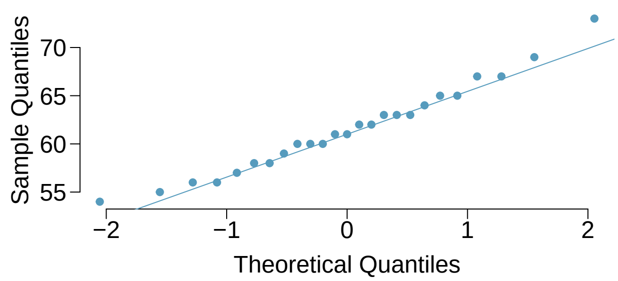
```r
pnorm(75.26, mean = 61.52, sd = 4.58) - pnorm(47.78, mean = 61.52, sd = 4.58) # 3 sd
```

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

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

```
The bell curve is a good approximation of the normal distribution and the qqplot is close to the li
distribution.
```






```r
# Use the DATA606::qqnormsim function
```

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?

```
p <- ((1-.02)^(10-1))*.02
p
```

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

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

```
p <- ((1-.02)^(100))
p
```

```
## [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?

```
e_v = 1/.02
e_v
```

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

```
sd <- sqrt((1 - .02)/(.02 * .02))
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?

```
e_v <- 1/.05
e_v
```

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

```
sd <- sqrt((1 - .05)/(.05 * .05))
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 probability of an event means that the event is more likely to happen and thus reduc
```

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
p <- dbinom(2,3, .51)
p
```

```
## [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 probabilities are both in case a and case b
```

```r
p <- (0.51 * 0.51 * 0.49) + (0.51 * 0.49 * 0.51) + (0.49 * 0.51 * 0.51)
p
```

```
## [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).

```
The method under section b would be too tedious since a total of 56 combinations would be required.
```

```r
p <- dbinom(3, 8, 0.51)
p
```

```
## [1] 0.2098355
```

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 teams 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_serv <- 0.15
n <- 10
k <- 3
p <- choose(n - 1, k - 1) * (1 - p_serv)^(n - k) * p_serv^k
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?

```
Probability of 10th serve will still be 15%. All her serves have a probability of 15% independent o
the case if we factor in fatigue but its not relevant in answering the question)
```

(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 instance is all about making a third successful serve in 10 attempts ( a case of joint pr
only looks at the probability of just the 10th serve which is independent of all other serves.
```