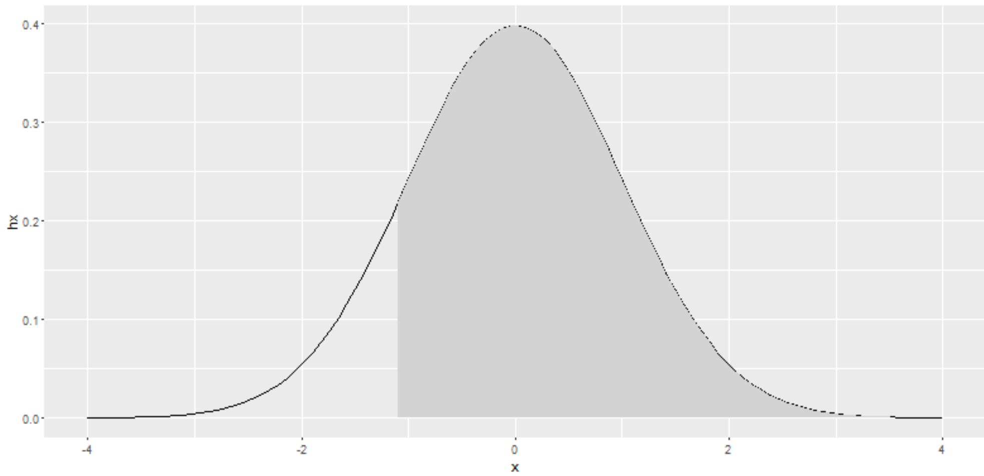
Chapter 3 – Distribution Homework

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

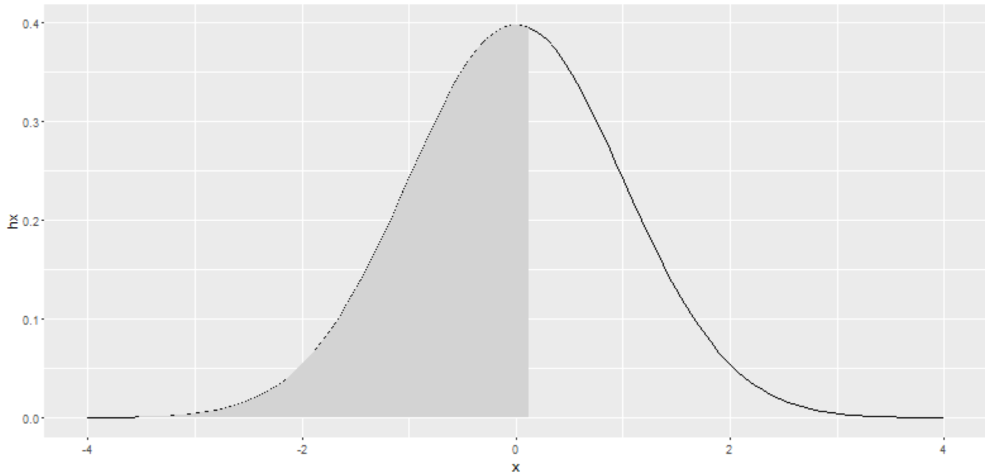
(a) $Z > -1.13$ (b) $Z < 0.18$ (c) $Z > 8$ (d) $|Z| < 0.5$

Ans:

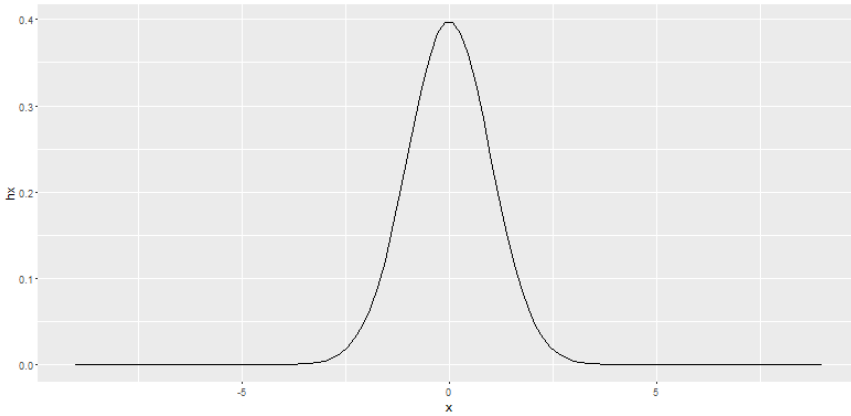
a) Z > -1.13 = (1 – 0.1292) = 0.8708 = 87.08%



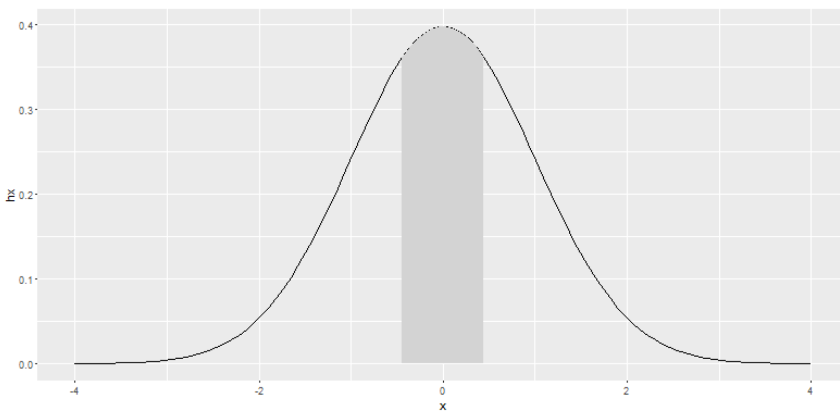
b) Z < 0.18 = 0.5714 = 57.14%



c) Z > 8 = 6.661338e-16



d) | Z | < 5 = -0.5 < Z < 0.5 = (Z < 0.5 – Z < -0.5) = (0.6915 – 0.3085) = 0.383 = 38.3%



3.4 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.

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

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

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

(e) What percent of the triathletes did Mary finish faster than 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.

Ans:

a) Men, Ages 30 – 34 : N(μ = 4313, σ = 583)
Women, Ages 25 – 29 : N(μ = 5261, σ = 807)

b) Leo's Z score = (4948 – 4313) / 583 = 1.09.
Mary's Z score = (5513 – 5261) / 807 = 0.31

These Z scores indicate that both performed better than the average performance in their respective groups.

c) Based on the Z scores it seems that Mary performed better as her performance is more nearer to the Women's mean and hence a lower value in her group compared to Leo's Z score.

d) Z < 1.09 = (1 – 0.8621) = 0.1379. Therefore Leo finished faster than 13.79% of the group.

e) Z < 0.31 = (1 – 0.6217) = 0.3783. Therefore Mary finished faster than 37.83% of her group.

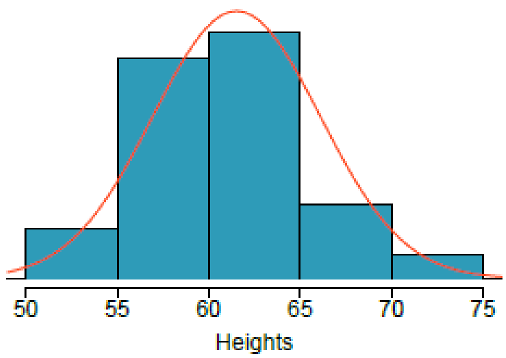
f) The above answers have been arrived at based on the assumption of Normal distribution. If the distribution was different than normal then we couldn't calculate the areas under the curve using the current Z score calculations

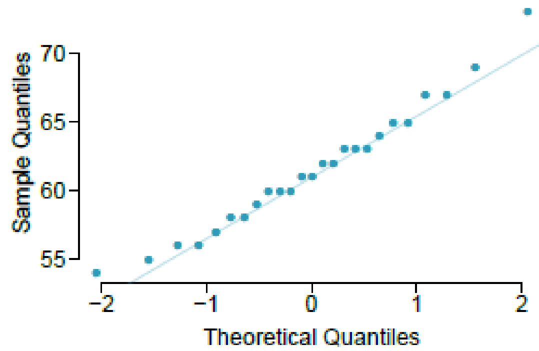
3.18 Heights of female college students. Below are heights of 25 female college students.

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.

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





Ans:

a) Let's work out the table of values as below.

SD	Lower	Upper	Count	Total	Percentage
SD-1	56.94	66.1	17	25	0.68
SD-2	52.36	70.68	24	25	0.96
SD-3	47.78	75.26	25	25	1

SD1 – SD3 are depicted with the Lower and Upper bounds of the standard deviations

Count – This is the students that are within the lower and upper bound for the respective SDs

Based on the above table, we can conclude that the distribution approximately follows the 68-95-99.7 % rule.

b) The left graph distribution curve visually looks to be normal. The right plot shows the values deviate a bit on the upper side, this indicates a slight skew.

3.22 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 10^{th} transistor produced is the first with a defect?

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

(c) On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?

(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) 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?

Ans:

a) DefectiveRate = 0.02
GoodRate = 1 – 0.02 = 0.98
n=10
probability = (goodrate)^(n-1) * DefectiveRate = 0.016675 = 1.67%

b) We can reword the question as follows: What is the probability that the 101th transistor produced is the first with the defect? To answer this question, we use the above formula:

DefectiveRate = 0.02
GoodRate = 1 – 0.02 = 0.98
n=101
probability = (goodrate)^(n-1) * DefectiveRate = 0.00265 = 0.265%

c) Expected = 1 / DefectiveRate = 50
SD = sqrt((1-DefectiveRate) / DefectiveRate^2)) = 49.49%

d) Expected = 1 / 0.05 = 20
SD = sqrt((1 - 0.05) / 0.05^2)) = 19.49%

e) Based on c) and d) it seems that increasing the probability of an event decreases the mean and SD of the wait time until success.

3.38 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.

(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.

(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).

Ans:

a)

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad (3.40)$$

n=3

k=2

p=0.51

Using the above values in the formulae we get

P(2 out of 3 kids are boys) = 0.382 = 38.2%

b)

	Scenario 1	**Scenario 2**	**Scenario 3**
Child 1	Boy	Boy	Girl
Child 2	Boy	Girl	Boy
Child 3	Girl	Boy	Boy
Probability	p(boy) * p(boy) * p(1-boy)	p(boy) * p(1-boy) * p(boy)	p(1-boy) * p(boy) * p(boy)
	0.127449	0.127449	0.127449

Summing up the probabilities, we get 38.2%. This is the same as we got from a).

c) To use the b) approach, we need to define all combinations of the 3 boys in 8 kids. Drawing up this table is quite tedious and subsequently we need to calculate the disjoint probabilities. The approach a) is more straight forward.

3.42 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 10^{th} try she will make her 3^{rd} successful serve?

(b) Suppose she has made two successful serves in nine attempts. What is the probability that her 10^{th} 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?

Ans:

a) p= 0.15
   n=10
   k=3

$$P(\text{the } k^{th} \text{ success on the } n^{th} \text{ trial}) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

P(3^{rd} success on 10^{th} try) = (9! / (2! * 7!)) * (0.15^3) * (1 – 0.15)^(10-3)
= 36 * 0.003375 * 0.320577
= 0.03895
= 3.895%

b) Since the serves are independent of each other, the probability of 10^{th} serve being successful is the same as others i.e 15%

c) The first case a) deals with the case that the 10^{th} serve will be the 3^{rd} successful serve given that there are already 2 successful serves from the earlier 9. So the probabilities of this serve are tied to the outcomes of the earlier 9 serves. In the second case b) we are only checking whether the 10^{th} serve will be successful without any regards to the earlier 9 serves.