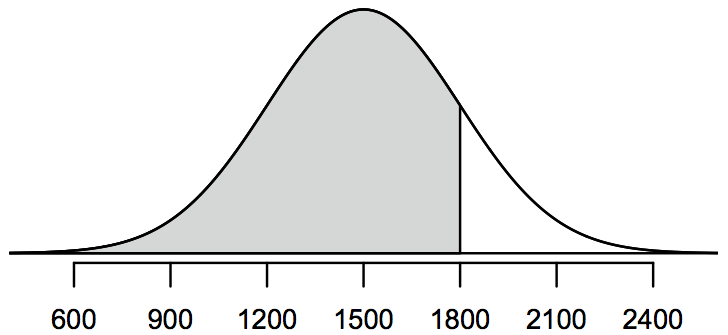
**The Normal Distribution in R**

If Marcus scored an 1800, what percentile was he in? In other words, what percentage of people did he score better than?

## Graphically, percentile is the area below the probability distribution curve to the left of that observation.



DRAW A PICTURE

## There are many ways to compute percentiles/areas under the curve.

## 

## Applet

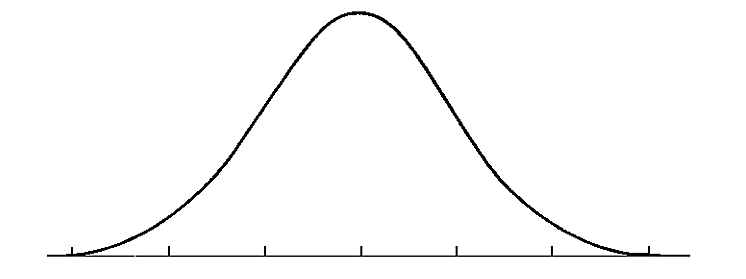
2) R

## <https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html>

## 

*At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection.*

What percent of bottles have less than 35.8 ounces of ketchup?

pnorm(\_\_\_\_\_, mean = \_\_\_\_, sd = \_\_\_\_)

Find again by first computing Z-score:

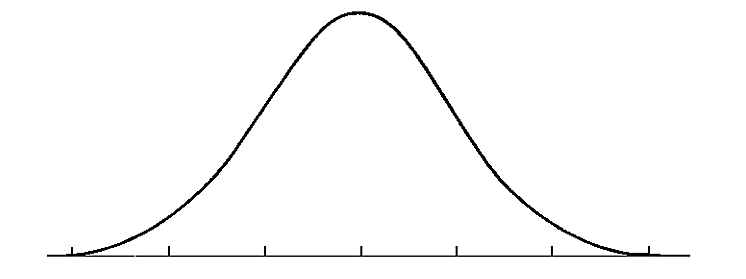
pnorm(\_\_\_\_\_, mean = \_\_\_\_, sd = \_\_\_\_)

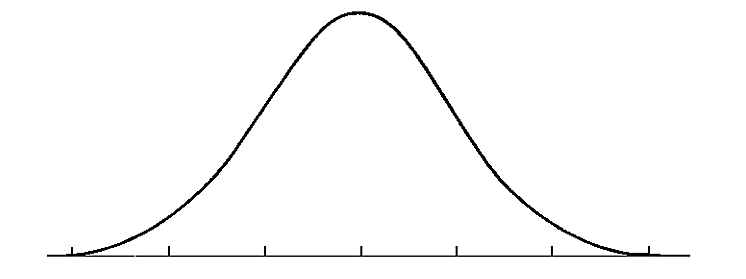
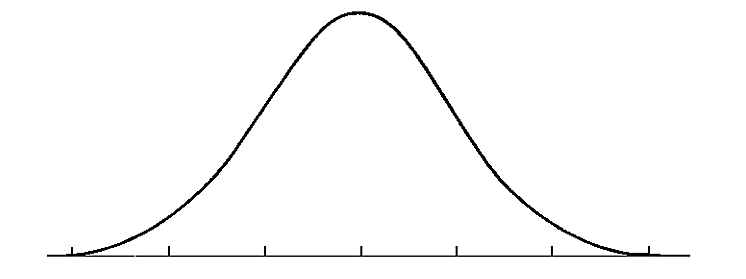
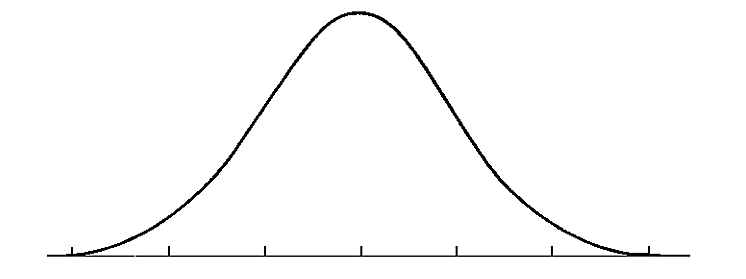
By default, R assumes mean = 0 and sd = 1, so when dealing with Z-scores, you can leave those arguments out.

A number and a number

Description automatically generated with medium confidence

What percent of bottles have MORE than 36.2 ounces of ketchup?





1 - pnorm(\_\_\_\_\_, mean = \_\_\_\_, sd = \_\_\_\_)

Find again by first computing Z-score:

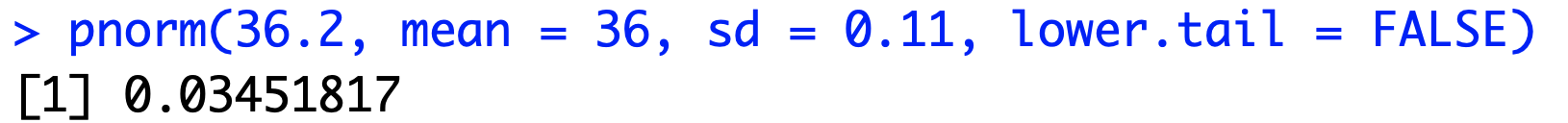
1 - pnorm(\_\_\_\_\_)

Z-scores make it a little easier to recognize the SYMMETRY of the normal distribution

P(Z < -1.82) = P(Z > 1.82)

By default, R calculates the probability that falls BELOW the cutoff value (i.e. the LOWER tail).

To calculate the probability that falls ABOVE a cutoff value, you should add lower.tail = FALSE



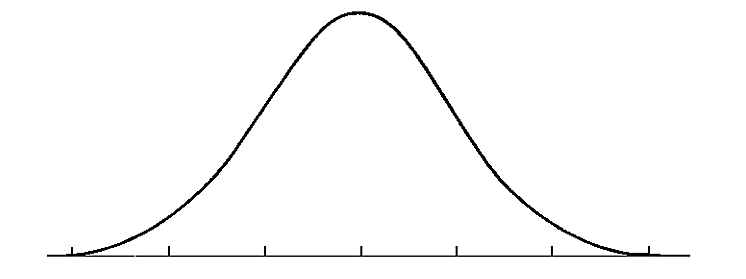
So, what percent of bottles pass the quality control inspection?

***What if you have a probability/percentile, and you need the corresponding cutoff?***

Use qnorm() instead of pnorm()

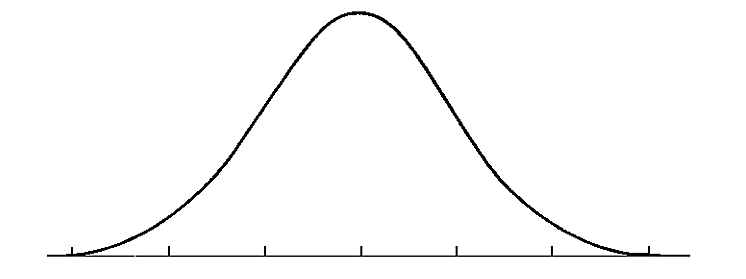
Body temperatures of healthy humans are distributed nearly normally with mean 98.2 (degrees Fahrenheit) and standard deviation 0.73.

What is the cutoff for the lowest 3% of human body temperatures?



qnorm(.03, mean = \_\_\_\_\_\_\_\_\_, sd = \_\_\_\_\_\_\_\_\_\_\_)

What is the cutoff for the highest 10% of human body temperatures?



qnorm(\_\_\_\_\_, mean = \_\_\_\_\_\_\_\_\_, sd = \_\_\_\_\_\_\_\_\_\_\_)

Careful: the “1 minus” trick doesn’t work in quite the same way with qnorm()

**Problem 1**

Suppose that birthweights of newborn babies in the United States follow a normal distribution with mean 3300 grams and standard deviation 500 grams. Babies who weigh less than 2500 grams at birth are classified as low birthweight. Use R code to answer each of the following:

1. What percentage of newborn babies weigh less than 2500 grams?
2. What is the probability that a randomly selected newborn baby weighs more than 10 pounds?
3. What percentage of newborn babies weigh between 3000 and 4000 grams?
4. How little must a baby weight to be among the lightest 2.5% of all newborns? Convert your answer to pounds. Google to figure out the conversion!
5. How much must a baby weigh to be among the heaviest 10%? Convert your answer to pounds.

**Problem 2**

Recall that Z-scores are normally distributed with mean 0 and standard deviation 1. Find each of the following probabilities using R. code. Note, you should draw a picture to help you visualize what you are trying to find for each question

1. P(Z < 1.25) (
2. P(Z > 1.25) (
3. P(Z < -1.25)
4. P(Z > -1.25)
5. P(-1 < Z < 1)
6. P(-2 < Z < 2)
7. P(-3 < Z < 3)
8. P(-1.645 < Z < 1.645)
9. P(-1.96 < Z < 1.96)
10. P(-2.576 < Z < 2.576)

**Problem 3**

Use the qnorm() function to find:

1. the 95th percentile of Z-scores
2. the 5th percentile of Z-scores
3. the 2.5th percentile of Z-scores
4. the 97.5th percentile of Z-scores
5. the 99.5th percentile of Z-score