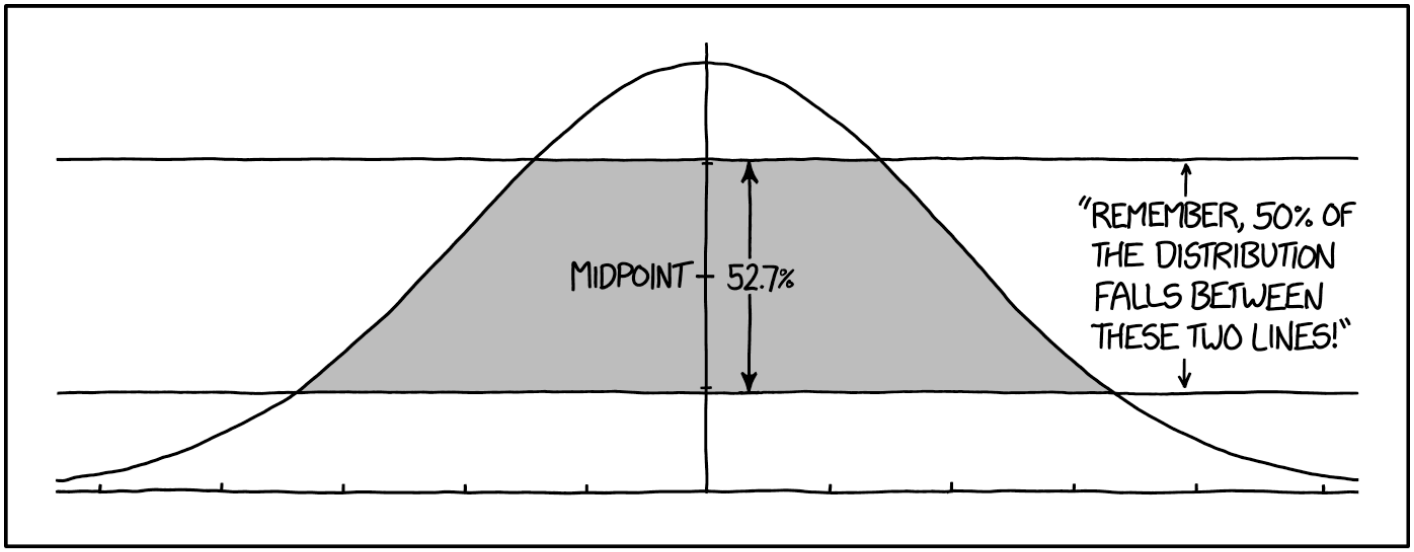
Day 11 - Normal random variables

Learning outcomes and terms to know

In this section, we focus on understanding the normal distribution, which is a probability distribution that is ubiqutous throughout statistics. By the end of this activity, you should understand:

- the normal distribution
- `pnorm`
- `qnorm`
- z-scores
- the standard normal distribution



HOW TO ANNOY A STATISTICIAN

xkcd comic

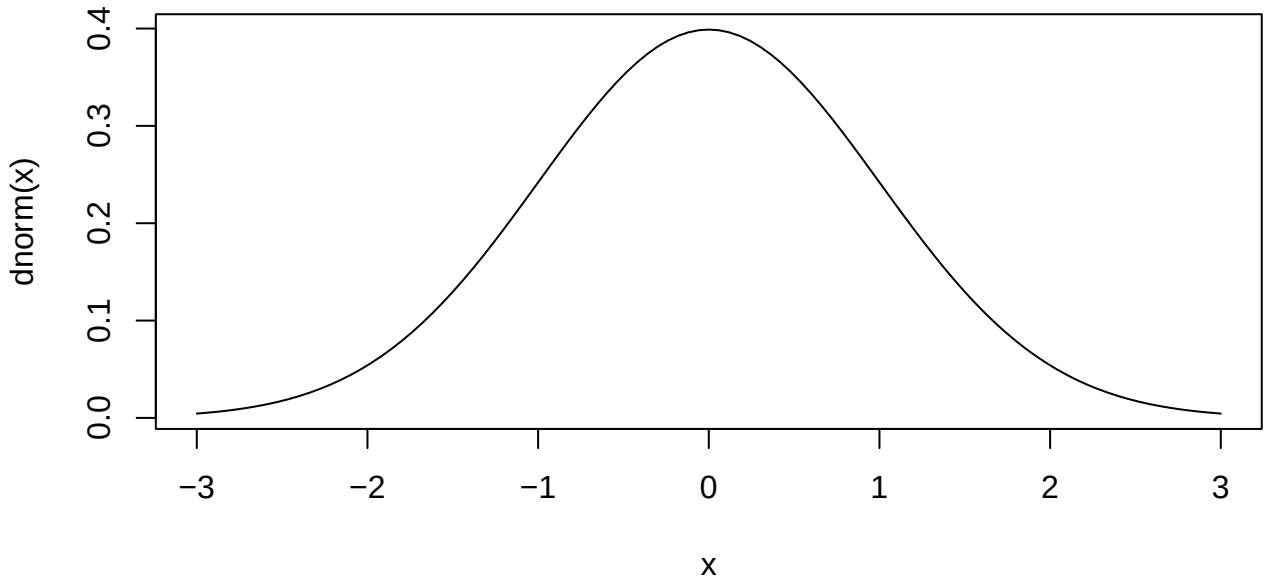
The Normal distribution

Note

The _____ is a probability distribution that is used to model a variety of phenomena in statistics. It is worth noting that many variables are **approximately** normally distributed, but none are truly normally distributed.

You won't be tested on this, but the equation that describes the normal curve, called a probability density function, is given below:

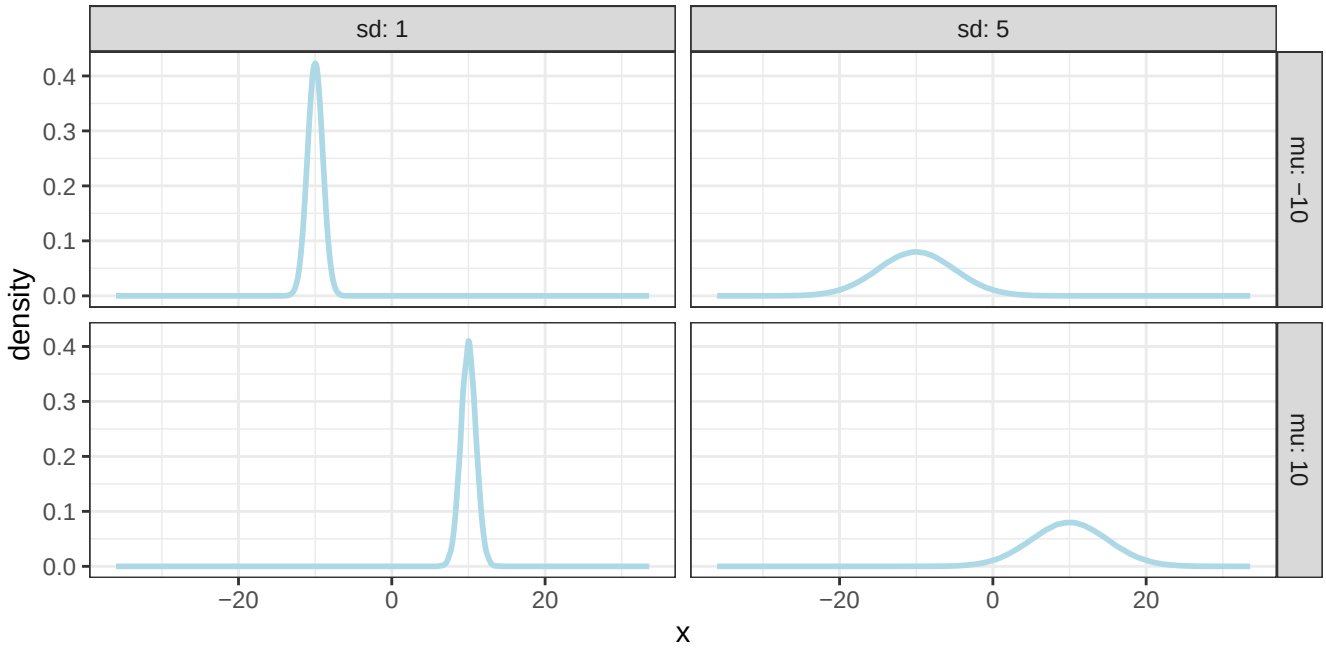
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



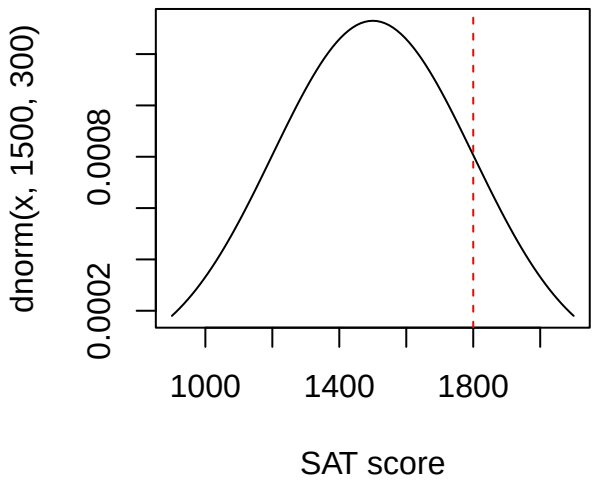
- **Shape**:

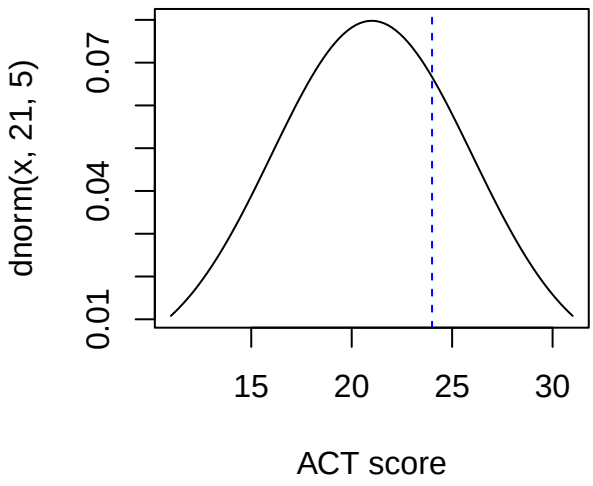
- **Notation:**

Understanding the parameters



Example: SAT scores are distributed approximately normally with mean 1500 and standard deviation 300. ACT scores are distributed approximately normally with mean 21 and standard deviation 5. A college admissions officer wants to determine which of the two applicants scored better on their standardized test with respect to the other test takers: Pam, who earned an 1800 on her SAT, or Jim, who scored a 24 on his ACT?





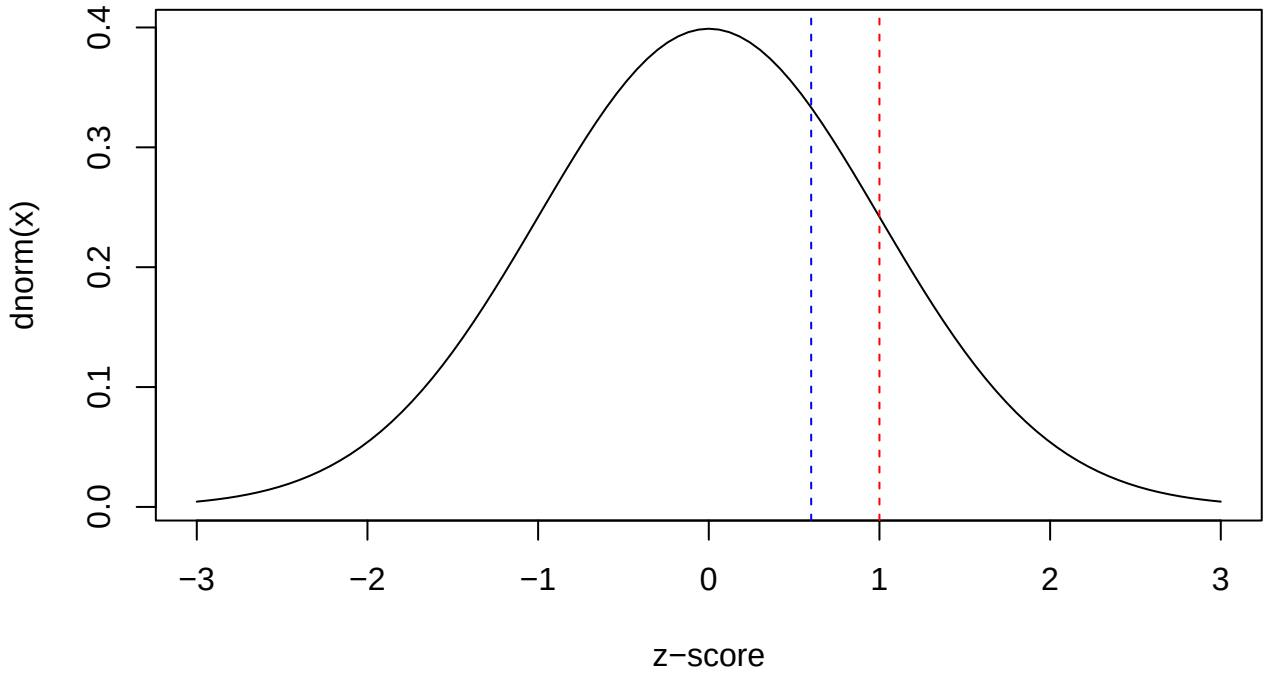
Z-scores and the standard normal distribution

Note

A _____ describes the number of standard deviations an observation is from the mean and is calculated independent of the unit of measurement. Therefore, it allows for direct comparison between normal random variables on different scales. If the population mean and standard deviation are known, the z-score is calculated as:

$$Z = \frac{X - \mu}{\sigma}$$

It can be shown that Z-scores are distributed according to a **standard normal** distribution, or a normal distribution with mean 0 and standard deviation 1.



Percentiles revisited

Review

The p^{th} percentile denotes the value (which we call a quantile q) such that $p\%$ of observations are less than it.

$$P(X \leq q) = p$$

We can use the `qnorm` and `pnorm` functions in R to obtain quantiles and the probability of observing a value less than or equal to a quantile. For example, `pnorm(99, 100, 15)` would tell us the probability of observing a value less than or equal to 99 from a Normal distribution with mean 100 and standard deviation 15. Conversely, `qnorm(.15, 100, 15)` would tell us the value from a Normal distribution with mean 100 and standard deviation 15 such that 15% are less than it.

Example: Body temperatures in adults is assumed to be approximately normally distributed with a mean of 98.2 degrees Fahrenheit (36.8 degrees Celsius), with a standard deviation of 0.72 degrees Fahrenheit (0.4 degrees Celsius).[1]

- What is the temperature such that 80% of people are below it?
- What is the temperature such that 15% of people are above it?
- What proportion of people have a body temperature of *at most* 99 degrees Fahrenheit?
- What proportion of people have a body temperature of *at least* 97.3 degrees Fahrenheit?
- What proportion of people have a body temperature of *between* 97.5 and 99 degrees?

[1] Source: Wikipedia

The Empirical rule (68-95-99.7 rule)

> **Note**
>
> Since the normal distribution is so ubiquitous, it is helpful to memorize what proportion of outcomes falls within one, two, and three standard deviations of the mean. These proportions define the _____.

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

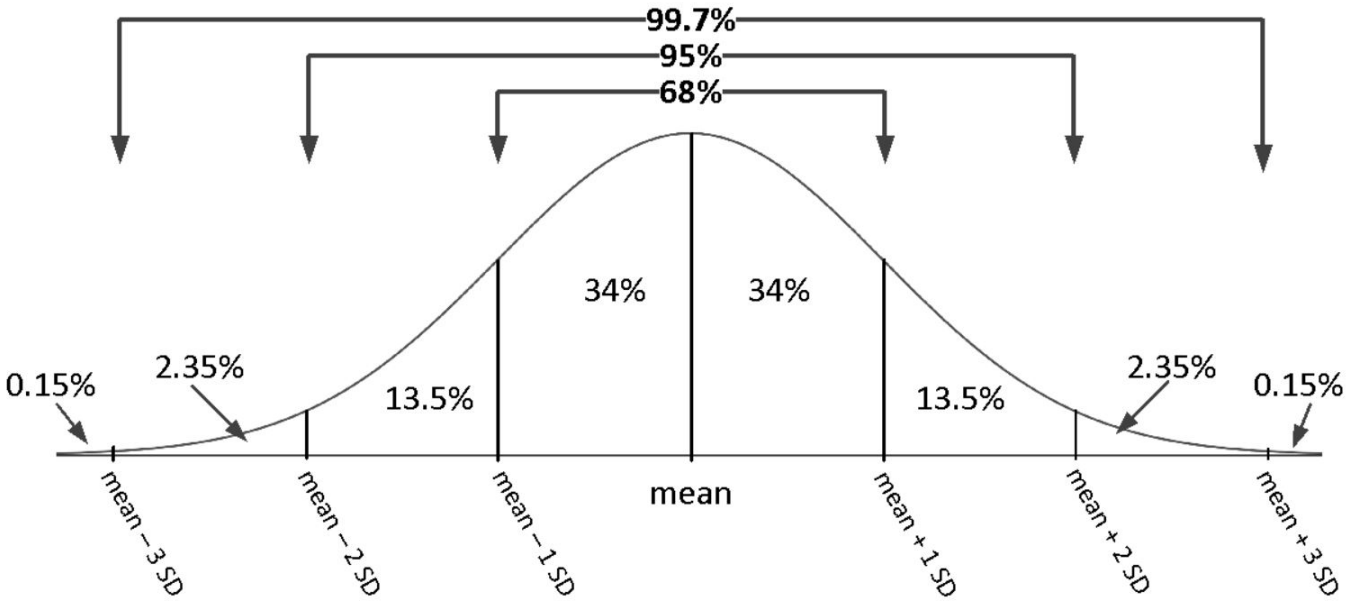
```
[1] 0.6826895
```

```
pnorm(2) - pnorm(-2)
```

```
[1] 0.9544997
```

```
pnorm(3) - pnorm(-3)
```

```
[1] 0.9973002
```



Source: lumen

Practice

1. When the original Star Wars movie came out (1977), there was much excitement about the movie. Here are some classic problems that were considered soon after.

   a) On the average, it takes Han Solo 45 seconds to check the coordinates and make the jump into hyperspace. The standard deviation on this important task is 5 seconds. When Han, Chewbacca, and their passengers are leaving for Alderaan, they make the jump in 33 seconds or less. What is the probability of such an accomplishment? You may assume that Han's jump times are approximately normally distributed.

   ```
   # pnorm or qnorm?
   ```

   b) Jawas, the short little orange-eyed scrappers on Tatooine, live in the desert and travel by sandcrawler. Their height is approximately normally distributed with a mean of four feet and a standard deviation of three inches. Suppose the Jawas wish to redesign their sandcrawler such that only 5% of them have to duck when exiting the craft. How tall should the escape exit be?

   ```
   # pnorm or qnorm?
   ```

2. *Tyrannosaurus* is a genus of large theropod dinosaurs, including *Tyrannosaurus Rex* (T-Rex) as one of the best represented theropods (fun fact: "Rex" means "king" in Latin). While there is no clear answer, paleontologists generally agree that T-Rex length measurements are approximately normally distributed, with a mean around 35 feet and a standard deviation of 2.5 feet.[2]

   a) Find the length such that 50% of T-Rex are longer.

   ```
   # do we need to use code here?
   ```

   b) Find the length such that 63% of T-Rex are shorter.

   c) Find the length such that 84% of T-Rex are shorter. What is the z-score associated with this value? You may use the empirical rule to solve this problem.

   d) What proportion of (known) T-Rex measured between 31 and 38 feet in length?

   e) Find the length such that 5% of T-Rex are longer.

   f) Between what two lengths are the middle 95% of all (known) T-Rex lengths? You may use the empirical rule for this question.

   g) "Big Mike" is a complete T-Rex fossil on display at the Museum of the Rockies in Bozeman, MT. Big Mike measures 38 feet long. Do you believe that Big Mike represents an exceptionally long specimen? Why or why not?

[2] Source: Wikipedia



Big Mike himself