Normal Probabilities

The benefit of computers

- Prior to the invention of computers, scientists employed regular people to sit and do their calculations for them.
- Could you imagine? Sitting around working out hard math problems all day?
- With computers, calculations that used to take a *really* long time became quick and easy.
- Calculations with normal curves for instance.
- In the previous labs, you saw how that data could be modeled or simulated with the normal curve.
- In this lab, we'll show you show you how to compute the most useful calculations of *normal models*.

Let's begin

- Start by loading the `cdc` data.
- Create a histogram of `heights`
- Split the plot into two, one for each `gender`.
- Finally, `fit` a `normal` curve to the histograms.
- Since each gender's height looks normally distributed, we'll use a *normal model* to compute the probability of people being different heights.
- **Estimate the proportion of *males* you think are short than 1.5 meters tall. What proportion of *males* do you think are taller then 1.6 meters tall?**

Why use a model?

- *"If we already have a sample of data, why not just count the number of people who are taller or shorter than some height?"*.
- This would certainly give us an estimate of the probability, but what would happen if you obtained a second sample?
- Our second sample of data would have people with different heights. Which means our probability estimate might end up being strikingly different.
- Estimating probabilities with models lead to estimates that are stable (i.e. Don't change).
- They also allow us to generalize samples to populations of people.
- **Who is the population of people our `cdc` data represents?**

Calculating probabilities with models!

- To estimate probabilities with our normal model we need to compute the `mean` and `stdev` of our observations.
- Calculate the `mean` and `stdev` of the males in our `cdc` data.

```
## Warning: The data contains 546 missing values
## The data contains 420 missing values
```

- To find the probability randomly choosing a teenage male shorter than 1.5`m` tall we use the `pnorm` function.

- `pnorm` is short for *probability using a normal model*

```
pnorm(1.5, mean=1.757365, sd=0.08377103)
```

```
## [1] 0.001062
```

- About 0.1% of teenage males are shorter than 1.5`m`.

Using the pnorm funciton

- By default, the `pnorm` function computes the probability of everything smaller than the value of interest.
- In our example, the value of interest was `1.5`.
- So `pnorm` gave us the probability that an American teenage male would be shorter than 1.5`m` tall.
- To flip this, and find the probability of finding someone taller than 1.5`m` tall, we write:

```
pnorm(1.5, mean=1.757365, sd=0.08377103,
      lower.tail = FALSE)
```

- **Compute the probability of randomly selecting a male that is taller than 1.6 meters tall. How did your estimate compare?**

Going the other way

- Now you know how to take a value and compute a probability.
- Next we'll flip this, and use a probability to compute a value.
- We'll start by calculating the tallest height, in meters, a male could be and still be in the shortest 10% of males
- We'll do this using the `qnorm` function (which stands for *quantile of a normal probability*)
- To find out how tall the shortest 10% of males are:

```
qnorm(0.10, mean=1.757365, sd=0.08377103)
```

```
## [1] 1.65
```

- 10% of men are shorter than 1.65`m`.

Using the qnorm function

- Just like the `pnorm` function, `qnorm` by default will always default to computing values starting from the left.
- Which is why 1.65 `m` was the height in which 10% of our male teenage population is shorter than.
- Also just like `pnorm` we can find the height which 10% of the teenage males are taller than by writing:

```
qnorm(0.1, mean=1.757365, sd=0.08377103,
      lower.tail = FALSE)
```

- **Find this value on the histogram. What does it mean for 10% of teenage boys to be taller than this height?**

On your own

- Compute the following using a *normal model* for the `heights` of females:
- **What's the probability that a randomly chosen female will be shorter than 1.5`m`?**
- **What percentage of females shorter are than 1.78`m` and taller than 1.48`m`?**
- **What's proportion of teenage females are shorter than the mean height of teenage males?**