Section 1

Announcements & Recap

Announcements

Upcoming due dates:

- Homework 4 due by 11:59pm on Tuesday, August 20
- Homework 5 due by 11:59pm on Friday, August 23

Coming up:

- Exam 1 on Tuesday, August 27 during lecture time (8-9:20am) in Buchanan 1910
- Slide Sets 1-5
- Homework assignments 1-5
- Exam 1 information on Canvas

Section 2

Learning Objectives

R Learning Objectives

1. Learn how to visualize the normal distribution using `plot_norm()`
2. Learn how to use `pnorm()` to find probabilities under the normal curve
3. Learn how to use `qnorm()` to find values of a normally distributed variable with specified probability to the left or the right

Statistical Learning Objectives

1. Understand how area under the normal curve relates to probability
2. Understand how to move between probabilities and quantiles of the normal distribution

Functions covered in this lab

1. `plot_norm()`
2. `pnorm()`
3. `qnorm()`

Section 3

Lab Tutorial

Normal Distributions

Recall that a "distribution" refers to the possible values a random variable can take as well as the probability that it takes those values. It is commonly used to approximate all sorts of things in nature and life.

A normal distribution is completely described by just two numbers: the **mean** and the **standard deviation**. The mean and standard deviation are called *parameters* of the distribution.

We denote a normal distribution by N(μ, σ), where μ is the population mean and σ is the population standard deviation.

Normal distributions are all bell-shaped, unimodal, and symmetric about their means, regardless of the values of the mean (μ) and the standard deviation (σ).

The mean μ specifies the *center* of the distribution. The standard deviation σ specifies the *variability* of the distribution (meaning, how narrow or wide it is).

The Standard Normal Distribution

Early statisticians did not have the technology we have at our disposal to find probabilities for variables that have normal distributions. They found probabilities under the normal curve by standardizing their variables as follows.

If a random variable X has a N(μ, σ) distribution, then the random variable
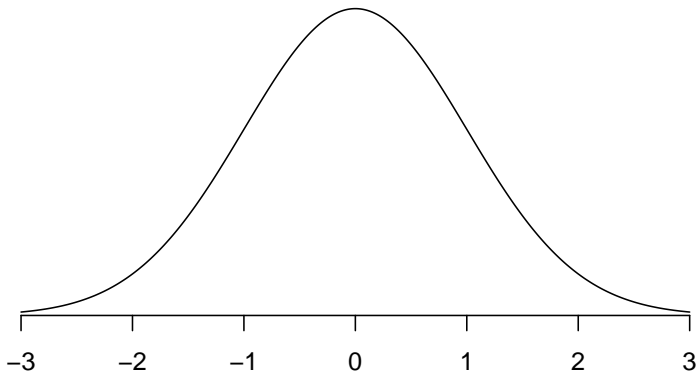
$$z = \frac{x - \mu}{\sigma}$$

has the **standard normal** N(0, 1) distribution, which is a normal distribution with mean 0 and standard deviation 1.

Then they used a standard normal table to estimate probabilities.

The Standard Normal Distribution

Here's a look at the standard normal distribution.

N(0, 1) Distribution

Visualizing a Normal Distribution with `plot_norm()`

The `stats250sbi` package that we are using includes the `plot_norm()` function to help you create a graphical display of a normal distribution. You will need to send the function the following arguments:

- `mean`: the mean of the normal distribution you'd like to draw (μ)
- `sd`: the standard deviation or standard error of the normal distribution you'd like to draw (σ or $\sqrt{\frac{p_0(1-p_0)}{n}}$, respectively)
- `shadeValues` (optional): either a number or a vector of two numbers (using `c()`) that are the boundaries of the region you'd like to shade.
- `direction`: where to shade (`"less"`, `"greater"`, `"between"`, or `"beyond"`)
- `col.shade`: the color to use when shading
- any other graphical parameters you want to use to control the appearance of the plot (like `main`, etc.)

Example: AP Statistics Scores

Suppose that test scores of AP Statistics students can be described by a normal distribution with mean 2.85 and standard deviation 0.43.

Example: AP Statistics Scores

What proportion of AP Statistics test scores are less than 2?

What does this look like? Let's use `plot_norm()` to find out:

```
plot_norm(mean = 2.85,
          sd = 0.43,
          shadeValues = 2,
          direction = "less",
          col.shade = "cornflowerblue")
```

Example: AP Statistics Scores

What proportion of AP Statistics test scores are less than 2?

To answer this question, we will use the `pnorm()` function.

Finding Probabilities with the `pnorm()` Function

The `pnorm()` function gives us a way to compute probabilities when a variable has a normal distribution. The arguments you need to send to `pnorm()` are:

- q: the quantile (value on the x-axis) for the normal distribution
- `mean`: the mean of the normal distribution (μ)
- `sd`: the standard deviation of the normal distribution (σ)
- `lower.tail`: set to 'TRUE' as the default, signifying that R will compute the probability **to the LEFT** of q; if you would like R to compute the probability *to the right* of q, set `lower.tail` to FALSE

Example: AP Statistics Scores

What proportion of AP Statistics test scores are less than 2?

```
pnorm(q = 2,
      mean = 2.85,
      sd = 0.43,
      lower.tail = TRUE)
```
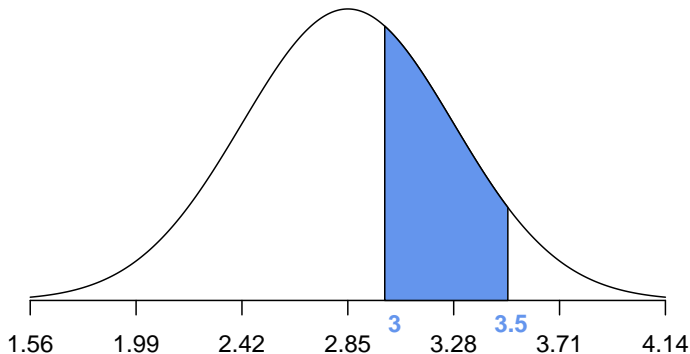
[1] 0.02403528

The R output tells us that about 2.4% of AP Statistics test scores are less than 2.

Example: AP Statistics Scores

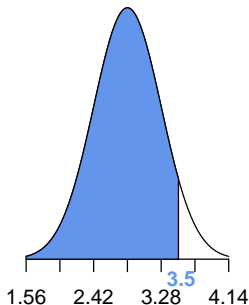Calculate the percentage of AP Statistics test scores that are between 3 and 3.5.
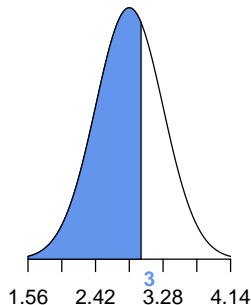
N(2.85, 0.43) Distribution

Example: AP Statistics Scores

We can find the area between 3 and 3.5 by taking the area to the left of 3.5 and subtracting the area to the left of 3:

N(2.85, 0.43) Distribution

N(2.85, 0.43) Distribution

Example: AP Statistics Scores

We can use the pnorm() function to find the area to the left (or right) of a specified value, so to find the area between two values, we need to do a little subtraction.
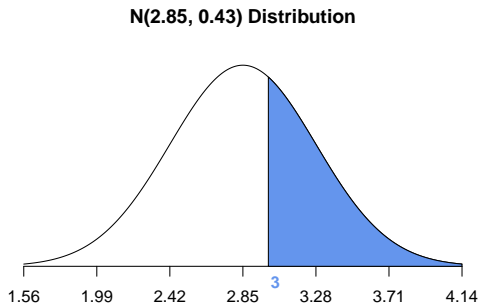
```
pnorm(3.5, mean = 2.85, sd = 0.43, lower.tail = TRUE) -
  pnorm(3, mean = 2.85, sd = 0.43, lower.tail = TRUE)
```

```
## [1] 0.2982915
```

Approximately 29.8% of AP Statistics test scores are between 3 and 3.5.

Example: AP Statistics Scores

What percentage of AP Statistics text scores are higher than 3?

N(2.85, 0.43) Distribution


```
pnorm(q = 3, mean = 2.85, sd = 0.43, lower.tail = FALSE)
```

[1] 0.3636058

Example: AP Statistics Scores

What score does an AP Statistics student need to be in the top 5%?

A student who has a score in the top 5% is at the 95th percentile.

Um, wait. We can't do this with `pnorm()`!

Finding Values of the Variable with the `qnorm()` Function

The `qnorm()` function gives us a way to find the values of a normally distributed variable when you are given a probability. The arguments you need to send to `qnorm()`:

- `p`: the probability or area under the curve you want to find an x-axis value for
- `mean`: the mean of the normal distribution, defaults to 0
- `sd`: the standard deviation of the normal distribution, defaults to 1
- `lower.tail`: determines whether `qnorm()` finds the value of the variable with area p to its left or right. If `lower.tail` is set to 'TRUE' (the default), the area p is to the LEFT. If `lower.tail` is set to 'FALSE', the area p is to the RIGHT.

Example: AP Statistics Scores

We can find the score the student needs in one of two ways:

1. p = 0.95 and lower.tail = TRUE tells R that we want the score that has area 0.95 to the left

```
qnorm(p = 0.95,
      mean = 2.85,
      sd = 0.43,
      lower.tail = TRUE)
```

2. p = 0.05 and lower.tail = FALSE tells R that we want the score that has area 0.05 to the right

```
qnorm(p = 0.05,
      mean = 2.85,
      sd = 0.43,
      lower.tail = FALSE)
```

Example: AP Statistics Scores

Hopefully it doesn't surprise you that the probabilities are the same:

➊

```r
qnorm(p = 0.95,
      mean = 2.85,
      sd = 0.43,
      lower.tail = TRUE)
```

```
## [1] 3.557287
```

➋

```r
qnorm(p = 0.05,
      mean = 2.85,
      sd = 0.43,
      lower.tail = FALSE)
```

```
## [1] 3.557287
```