Week 3: Testing Statistical Hypotheses

Univariate Statistics and Methodology using R

Department of Psychology
The University of Edinburgh

Today's Key Topics

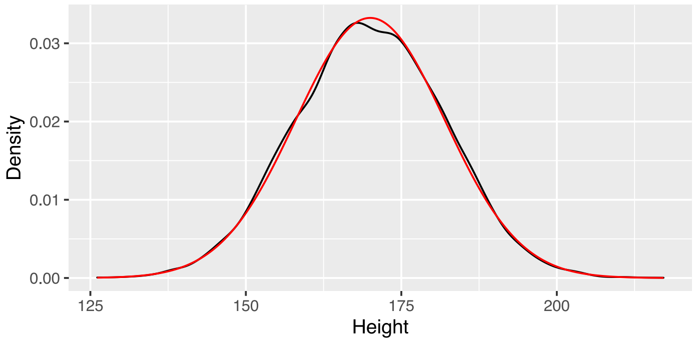
- One-tailed vs Two-tailed Hypotheses
- Null vs Alternative Hypotheses
- The Null Distribution
- z-Scores
- t-tests
- α

Part 1

More about Height

- Last time we simulated the heights of a population of 10,000 people
  - $\bar{x}$ = 170 cm
  - σ = 12 cm

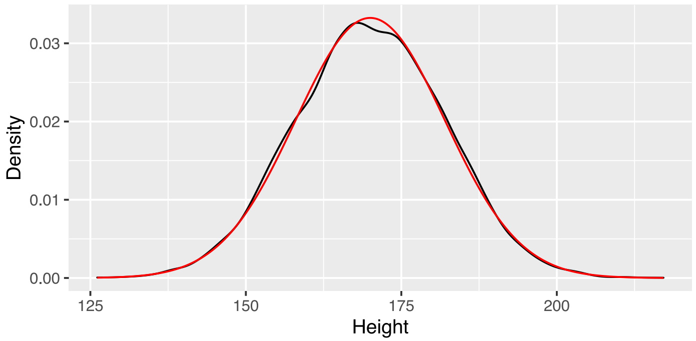
```
##   height
## 1  183.4
## 2  168.6
## 3  179.1
## 4  170.1
```



More about Height

- Last time we simulated the heights of a population of 10,000 people
  - $\bar{x}$ = 170 cm
  - σ = 12 cm

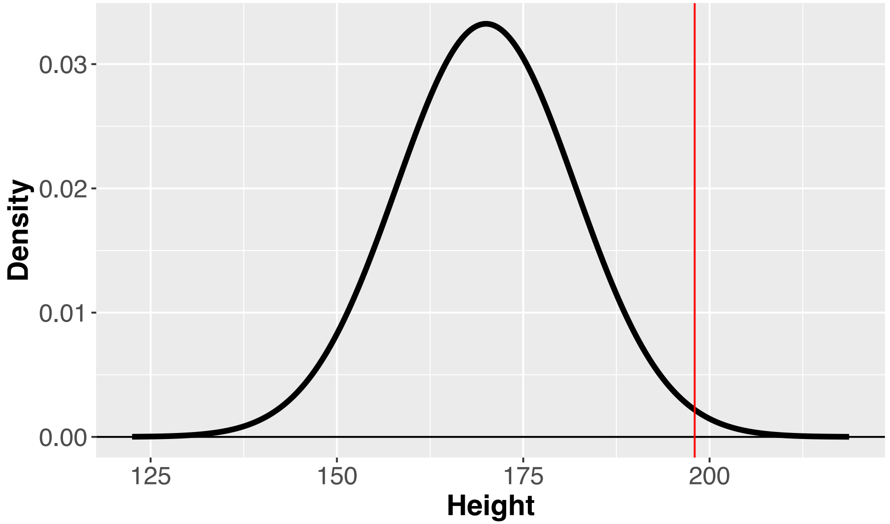
```
##   height
## 1  183.4
## 2  168.6
## 3  179.1
## 4  170.1
```



- This time, you'll learn how to compute the probability of randomly observing a specific value within the normal distribution

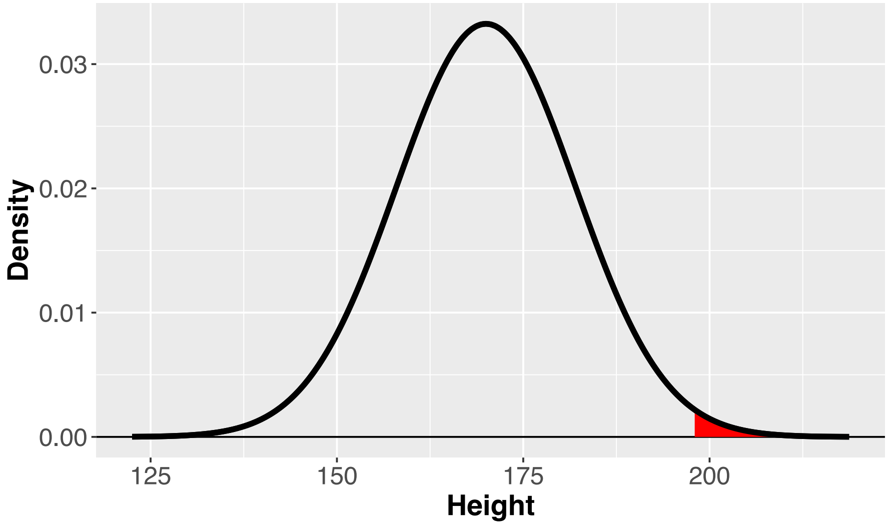
How Unusual is Casper?

- In his socks, Casper is 198 cm tall
- How likely would we be to find someone Casper's height in our population?



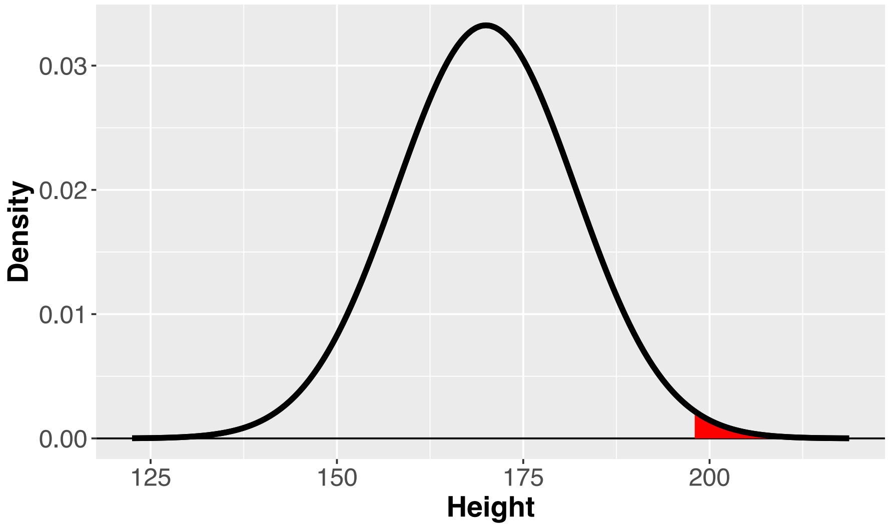
How Unusual is Casper (Take 2)?

- In his socks, Casper is 198 cm tall
- How likely would we be to find someone Casper's height or taller in our population?



How Unusual is Casper (Take 2)?

- In his socks, Casper is 198 cm tall
- How likely would we be to find someone Casper's height **or taller** in our population?
- The area is 0.0098
- So the probability of finding someone in the population of Casper's height or greater is 0.0098 (or, $p = 0.0098$)

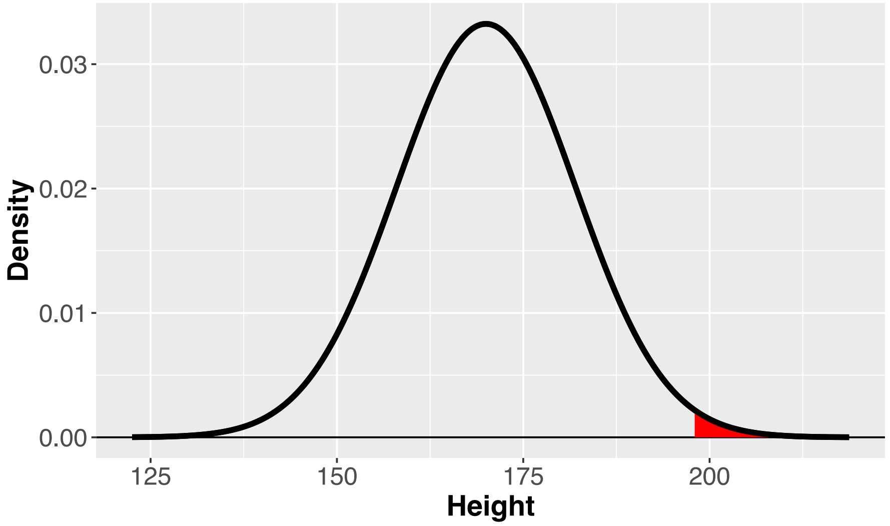


Area under the Curve

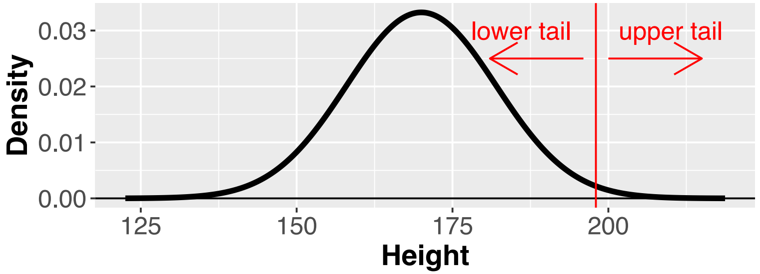
- So now we know that the area under the curve can be used to quantify **probability**
- But how do we calculate area under the curve?
- Luckily, R has us covered, using (in this case) the `pnorm()` function

```
pnorm(198, mean = 170, sd=12,
      lower.tail = FALSE)
```

```
## [1] 0.009815
```



Area under the Curve



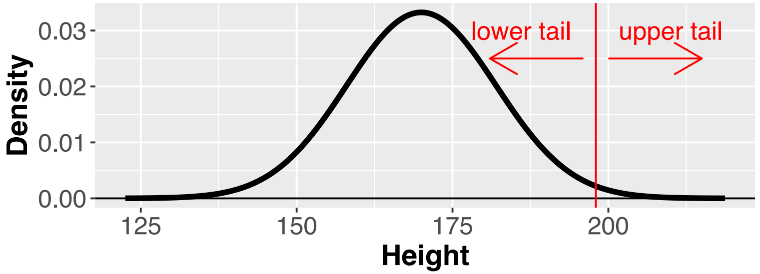

```
pnorm(198, mean = 170, sd=12,
      lower.tail = TRUE)
```

```
## [1] 0.9902
```

```
pnorm(198, mean = 170, sd=12,
      lower.tail = FALSE)
```

```
## [1] 0.009815
```

Area under the Curve




```
pnorm(198, mean = 170, sd=12,
      lower.tail = TRUE)
```

```
## [1] 0.9902
```

```
pnorm(198, mean = 170, sd=12,
      lower.tail = FALSE)
```

```
## [1] 0.009815
```

```
pnorm(198, mean = 170, sd=12, lower.tail = TRUE) +
  pnorm(198, mean = 170, sd=12,lower.tail = FALSE)
```

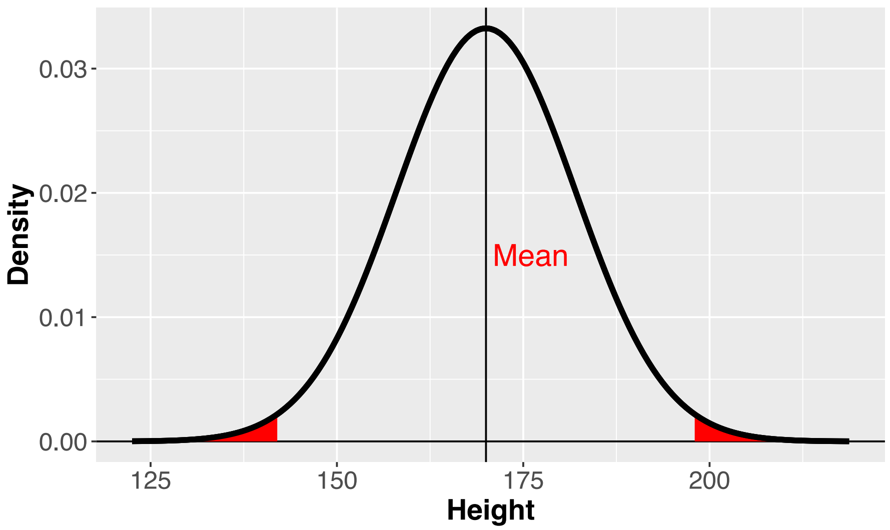
```
## [1] 1
```

Tailedness

- In this example, we kind of knew that Casper was *tall*
    - It made sense to ask what the likelihood of finding someone 198 cm *or greater* was
    - This is called a **one-tailed hypothesis** (we're not expecting Casper to be well *below* average height!)

Taildedness

- In this example, we kind of knew that Casper was *tall*
    - It made sense to ask what the likelihood of finding someone 198 cm *or greater* was
    - This is called a **one-tailed hypothesis**
- Often our hypothesis might be vaguer
    - We expect Casper to be *"different"*, but we're not sure how
    - In this case, we would make a non-directional, or **two-tailed hypothesis**

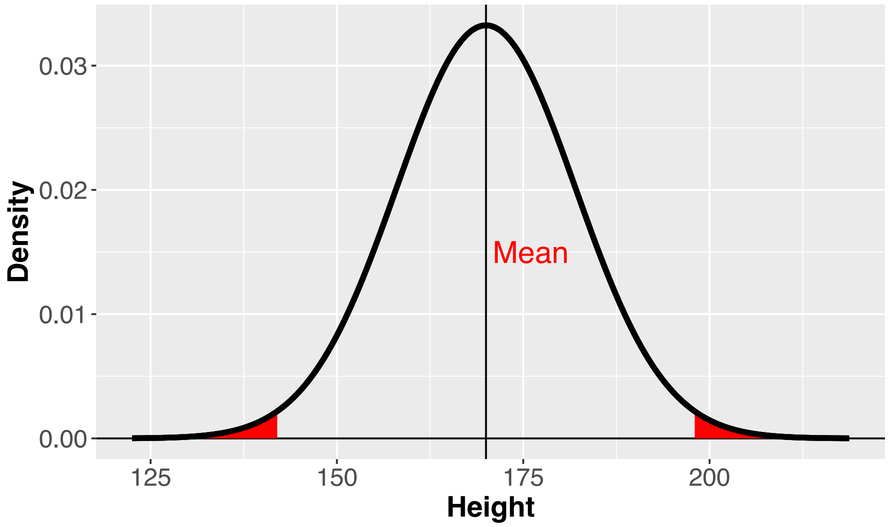


Tailedness

- For a two-tailed hypothesis we need to sum the relevant upper and lower areas:

```
2 * pnorm(198, 170, 12, lower.tail = FALSE)
```

```
## [1] 0.01963
```



So: Is Casper Special?

- How surprised should we be that Casper is 198 cm tall?
- Given the population he's in, the probability that he's 28cm or more taller than the mean of 170 is 0.0098
  - (Keep in mind, this is according to a *one-tailed hypothesis*)

So: Is Casper Special?

- How surprised should we be that Casper is 198 cm tall?
- Given the population he's in, the probability that he's 28cm or more taller than the mean of 170 is 0.0098
  - (Keep in mind, this is according to a *one-tailed hypothesis*)
- A more accurate way of saying this is that 0.0098 is the probability of selecting him (or someone even taller than him) from the population at random
  - There is about a 1% chance of selecting someone Casper's size or taller from the population.

A Judgement Call

We have to decide:

A Judgement Call

We have to decide:

If a 1% probability is *small enough*

A Judgement Call

We have to decide:

If a 1% probability is *small enough*

If a 1% chance doesn't impress us much

A Judgement Call

We have to decide:

If a 1% probability is *small enough*

If a 1% chance doesn't impress us much

Note that, in either case, we have nothing (mathematical) to say about the *reasons* for Casper's height

End of Part 1

Part Two

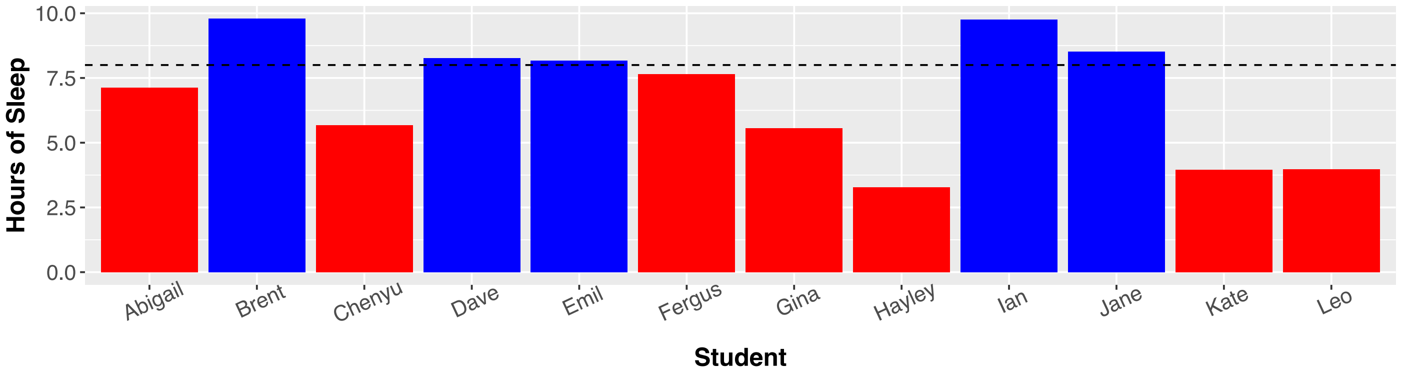
Group Means

Sleeping Guidelines

The USMR instructors are concerned that university students are not following the recommended sleep guidelines of 8 hours per night, and worry this could affect their academic performance. Is this idea worth further investigation?

Sleeping Guidelines

The USMR instructors are concerned that university students are not following the recommended sleep guidelines of 8 hours per night, and worry this could affect their academic performance. Is this idea worth further investigation?



Information About our Study

- Null Hypothesis (H_0)
  - Students are getting the recommended amount of sleep
- Alternative Hypothesis (H_1)
  - Students are getting less than the recommended amount of sleep
- There are 12 students

```
summary(m)
```

```
##      sleep          names
##  Min.   :3.28   Abigail:1
##  1st Qu.:5.16   Brent  :1
##  Median :7.38   Chenyu :1
##  Mean   :6.81   Dave   :1
##  3rd Qu.:8.33   Emil   :1
##  Max.   :9.79   Fergus :1
##                 (Other):6
```

```
sd(m$sleep)
```

```
## [1] 2.273
```

Information About our Study

- Null Hypothesis (H_0)
  - Students are getting the recommended amount of sleep
- Alternative Hypothesis (H_1)
  - Students are getting less than the recommended amount of sleep
- There are 12 students

```
summary(m)
```

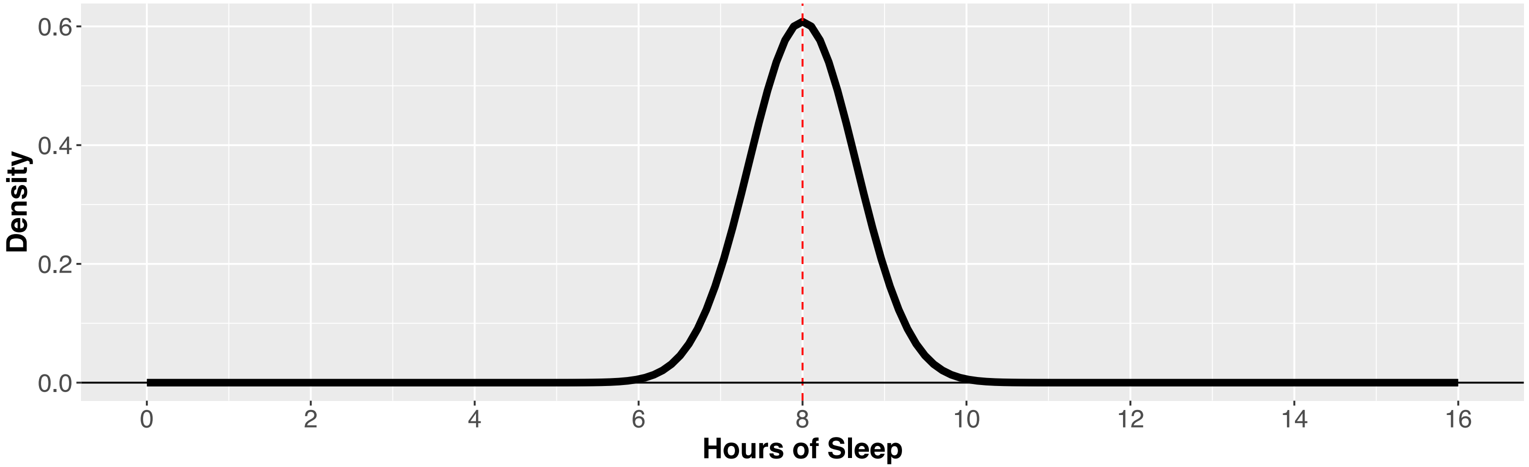
```
##       sleep            names   
##  Min.   :3.28   Abigail:1  
##  1st Qu.:5.16   Brent  :1  
##  Median :7.38   Chenyu :1  
##  Mean   :6.81   Dave   :1  
##  3rd Qu.:8.33   Emil   :1  
##  Max.   :9.79   Fergus :1  
##                 (Other):6
```

```
sd(m$sleep)
```

```
## [1] 2.273
```

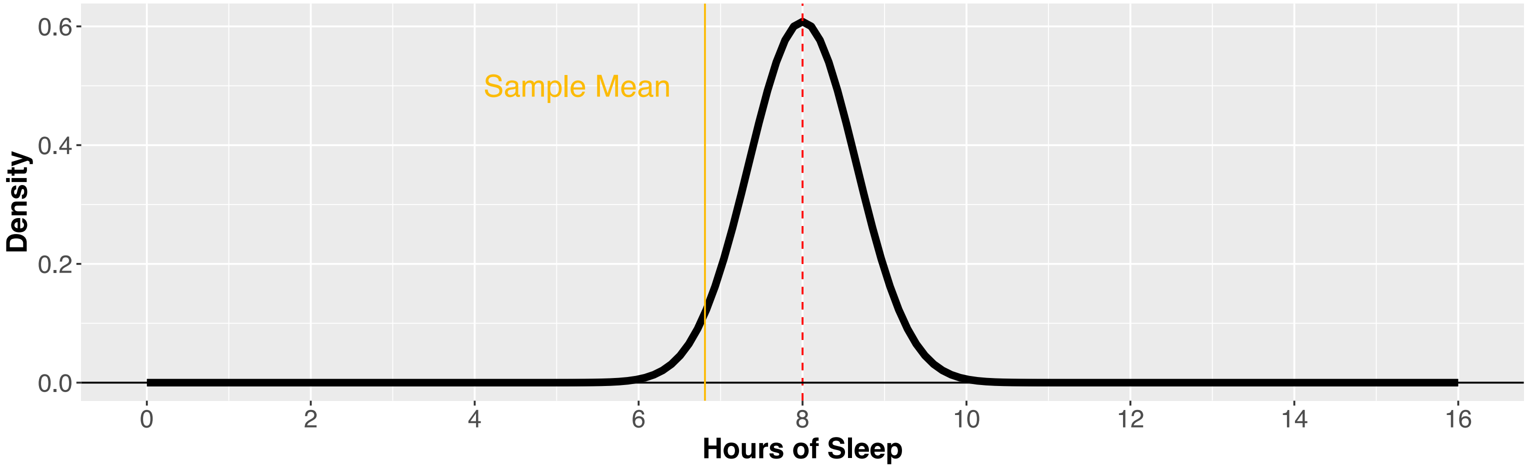
What is the probability of a group of 12 people getting a mean of 6.81 hours of sleep, given that most adults need around 8? (assuming they came from the same population)

Back to the Normal Distribution



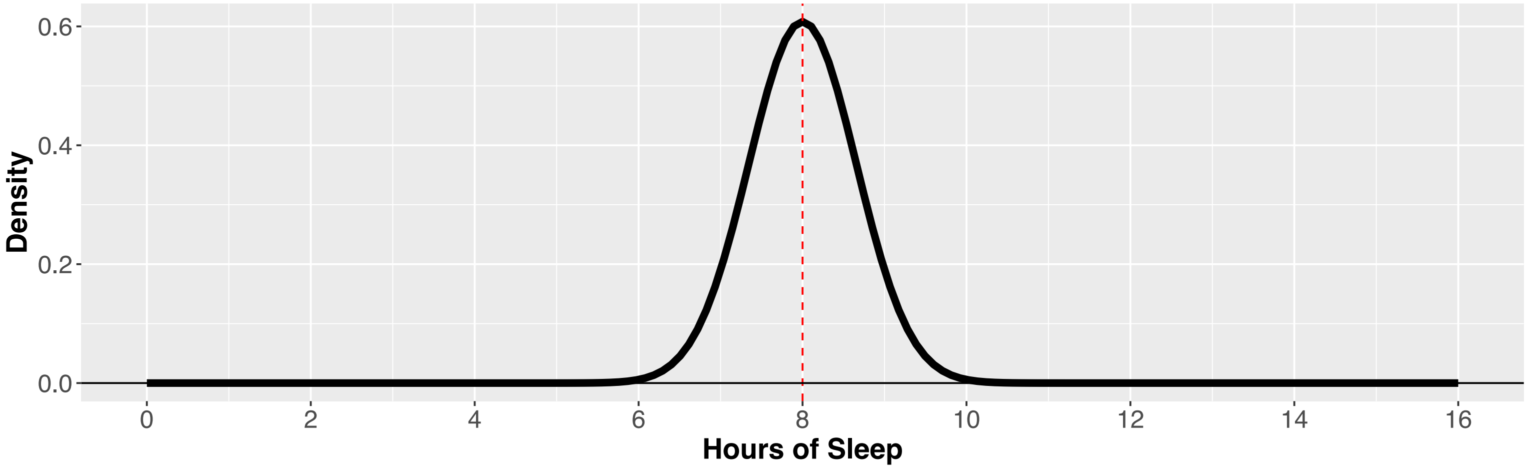
What is the probability of a group of 12 people getting a mean of 6.81 hours of sleep, given that most adults need around 8? (assuming they came from the same population)

Back to the Normal Distribution



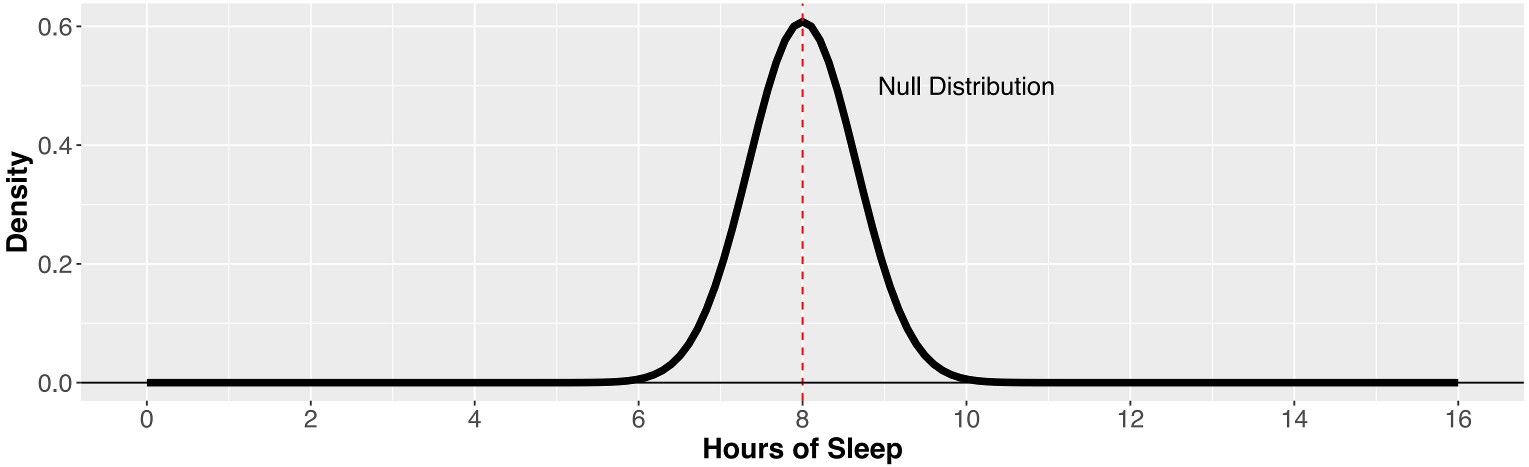
What is the probability of a group of 12 people getting a mean of 6.81 hours of sleep, given that most adults need around 8? (assuming they came from the same population)

The Null Distribution



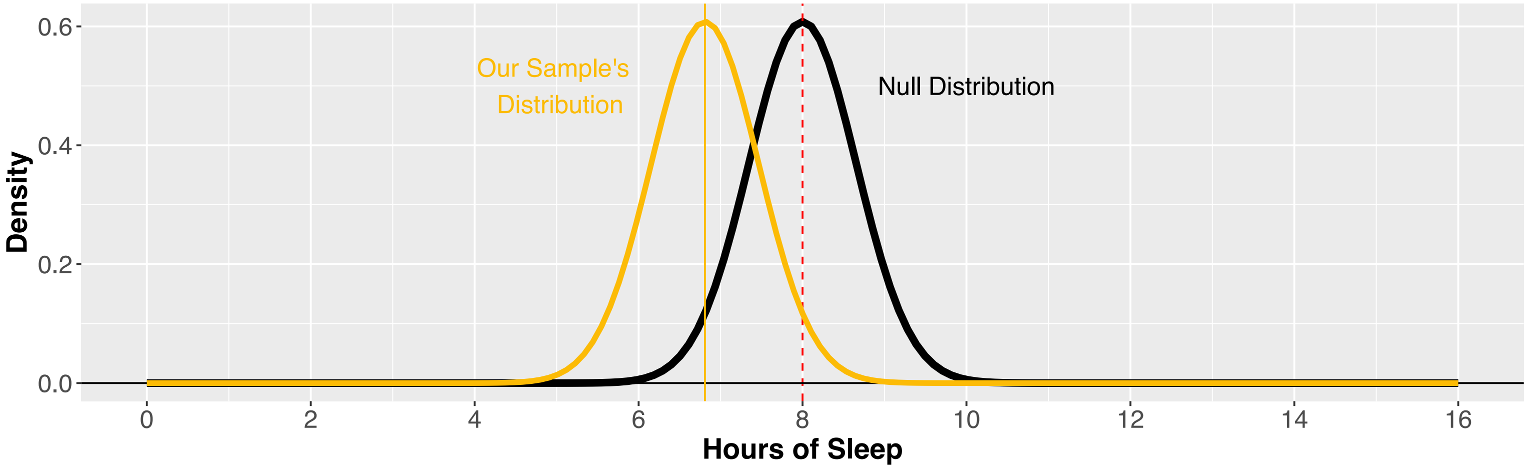
- H_0: Students are getting the recommended amount of sleep
- If H_0 reflects the **ground truth** (and *sleep* is a normally distributed variable), we would expect a frequency distribution of student measurements to look pretty similar to this

The Null Distribution

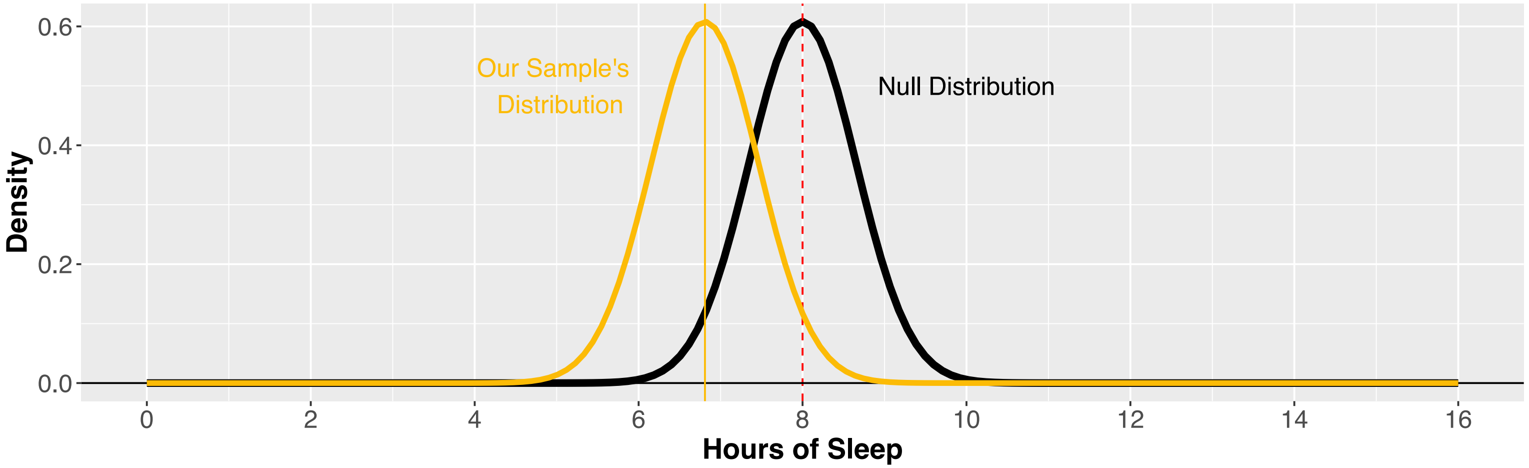


- H_0: Students are getting the recommended amount of sleep
- If H_0 reflects the **ground truth** (and *sleep* is a normally distributed variable), we would expect a frequency distribution of student measurements to look pretty similar to this

The Null Distribution

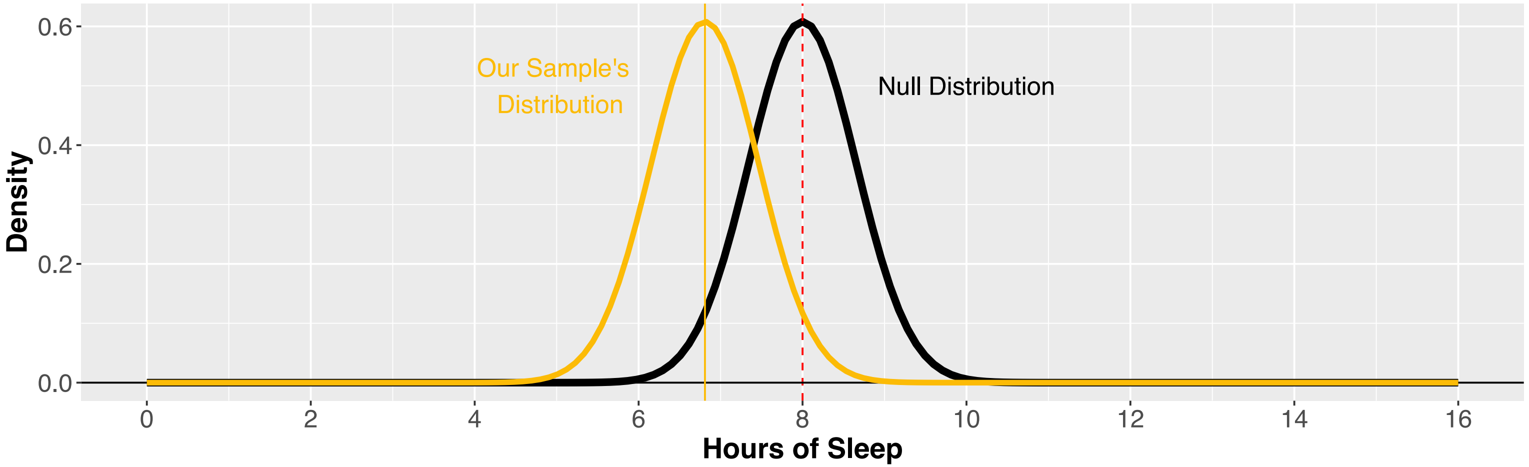


The Null Distribution



- From this, it *looks* as though our students may be getting less sleep than they should.

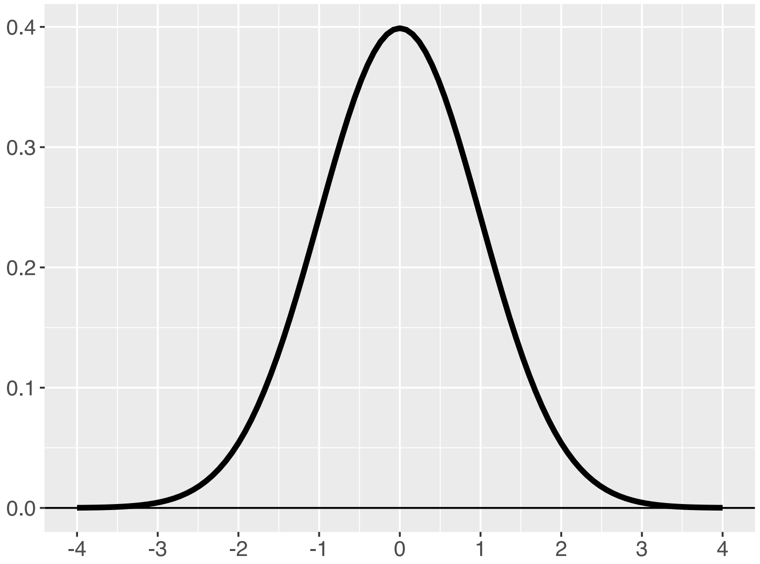
The Null Distribution



- From this, it *looks* as though our students may be getting less sleep than they should.
- However, simply seeing a shift in the curves isn't enough evidence to make that claim with certainty.

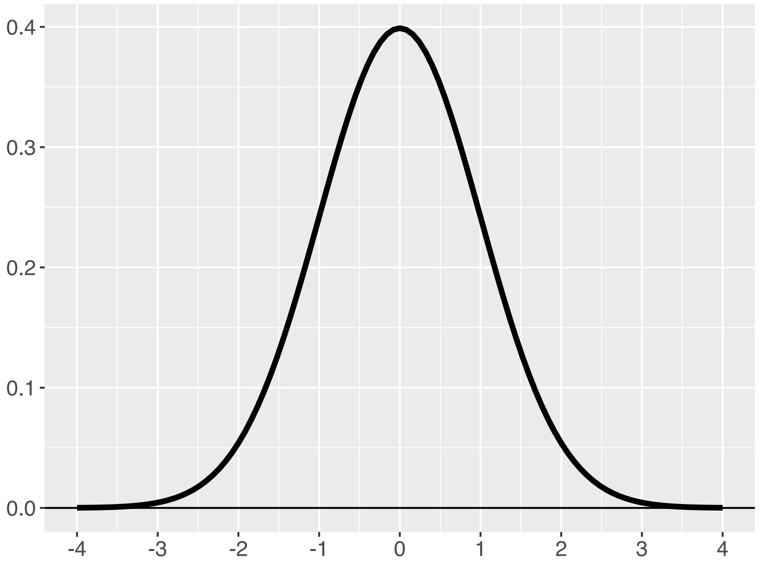
Back to the Normal Distribution...Again

We can compute the probability of a score's occurrence if it is part of a standardized normal distribution ($\mu = 0, \sigma = 1$)



Back to the Normal Distribution...Again

We can compute the probability of a score's occurrence if it is part of a standardized normal distribution ($\mu = 0, \sigma = 1$)

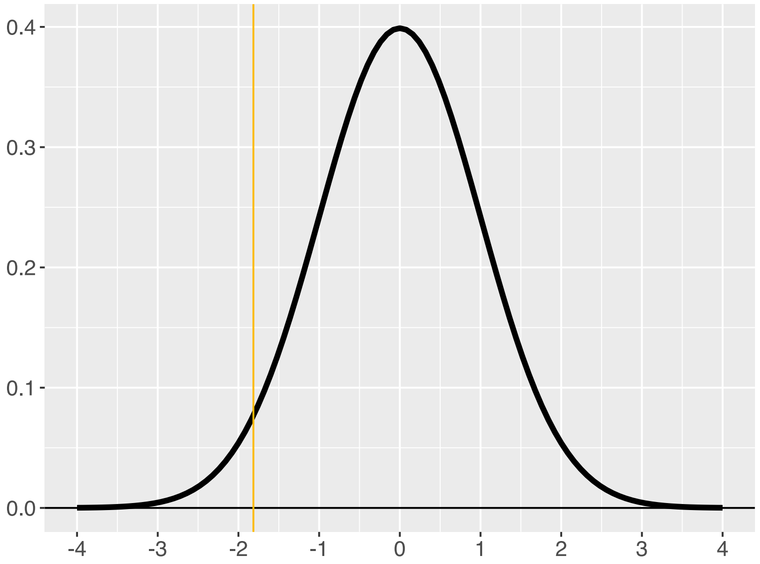


If our data are normally distributed, we can standardize the mean by converting it to a **z-score**

$$z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$

Back to the Normal Distribution...Again

We can compute the probability of a score's occurrence if it is part of a standardized normal distribution ($\mu = 0, \sigma = 1$)



If our data are normally distributed, we can standardize the mean by converting it to a **z-score**

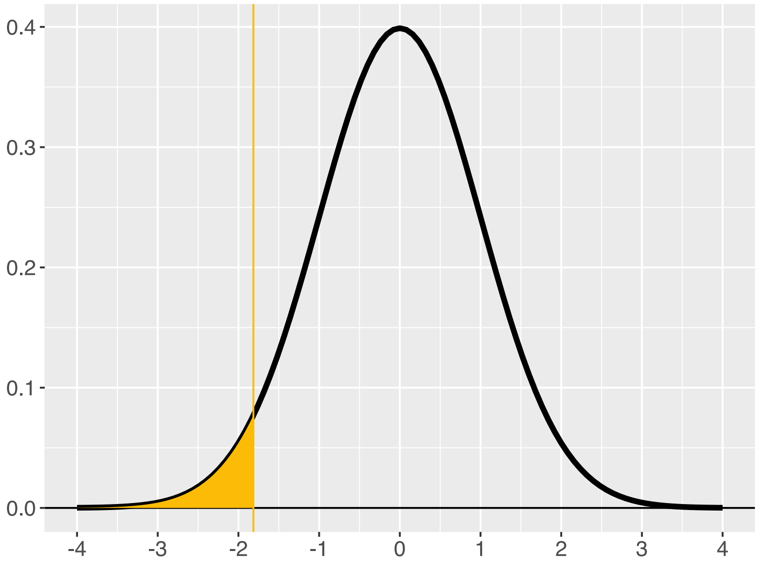
$$z = \frac{\bar{x}-\mu}{(\sigma/\sqrt{n})}$$

```
(mean(m$sleep) - 8)/(sd(m$sleep)/sqrt(12))
```

```
## [1] -1.814
```

Back to the Normal Distribution...Again

We can compute the probability of a score's occurrence if it is part of a standardized normal distribution ($\mu = 0, \sigma = 1$)



If our data are normally distributed, we can standardize the mean by converting it to a **z-score**

$$z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$

```
(mean(m$sleep) - 8)/(sd(m$sleep)/sqrt(12))
```

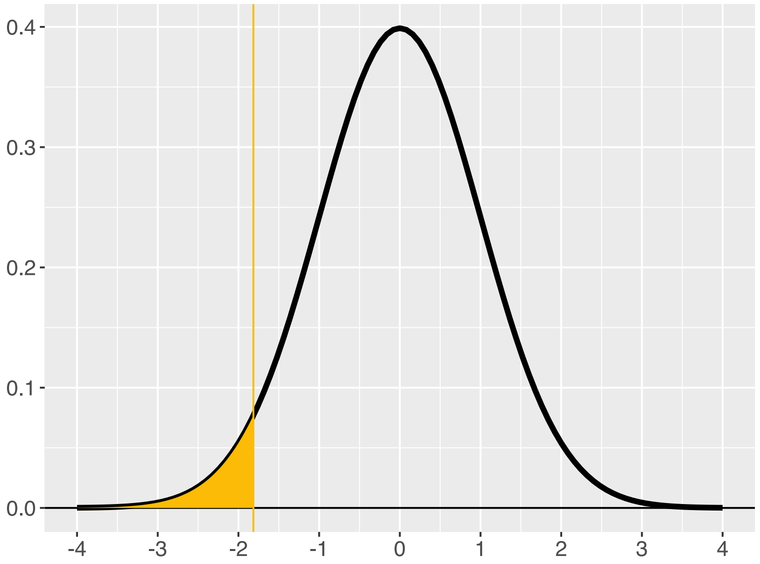
```
## [1] -1.814
```

```
pnorm(-1.814, mean = 0, sd = 1)
```

```
## [1] 0.03484
```

Back to the Normal Distribution...Again

We can compute the probability of a score's occurrence if it is part of a standardized normal distribution ($\mu = 0, \sigma = 1$)



If you picked 12 people at random from a population of people who get the recommended number of hours of sleep, there would be a 3% chance that their average sleep would be 6.81 hours or less

End of Part 2

Part 3

The t-test

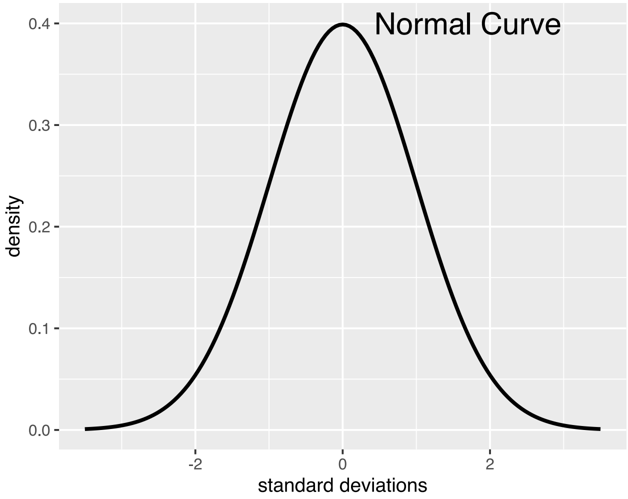
A Small Confession

Part Two wasn't entirely true

- All of the principles are correct, but for smaller n the normal curve isn't the best estimate
- For that we use the t distribution

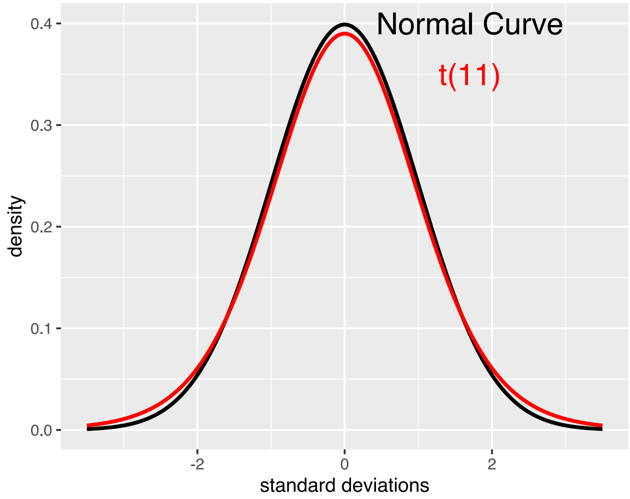
The t Distribution

"A. Student", or William Sealy Gossett



The t Distribution

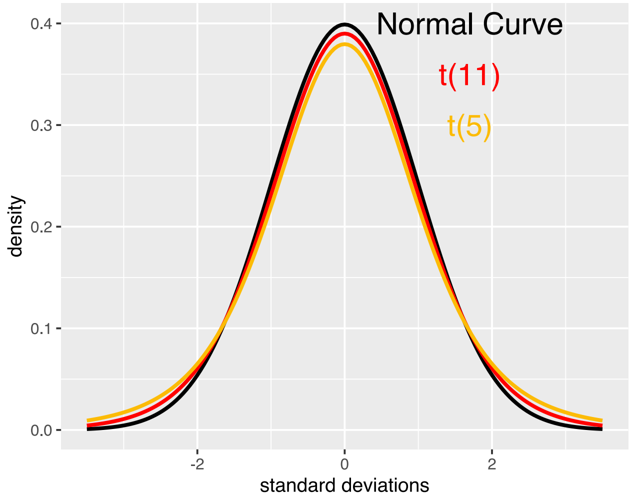
"A. Student", or William Sealy Gossett



Note that the shape changes according to *degrees of freedom*

The t Distribution

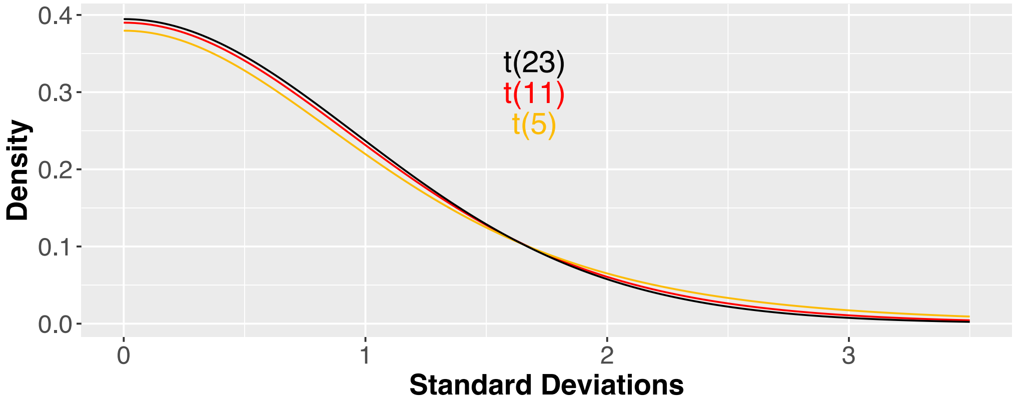
"A. Student", or William Sealy Gossett



Note that the shape changes according to *degrees of freedom*

The t Distribution

- Conceptually, the t distribution increases uncertainty when the sample is small
  - The probability of more extreme values is slightly higher
- Exact shape of distribution depends on sample size



Using the t Distribution

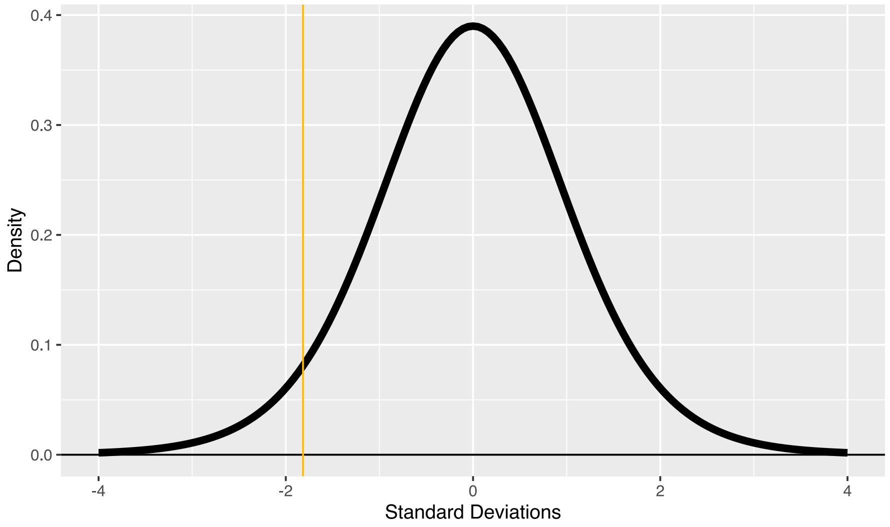
- In part 2, we calculated the mean hours of sleep for the group as 6.81
- We used the formula $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ to calculate z, and the standard normal curve to calculate probability

Using the t Distribution

- In part 2, we calculated the mean hours of sleep for the group as 6.81
- We used the formula $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ to calculate z, and the standard normal curve to calculate probability
- **The formula for a one-sample t-test is the same as the formula for z**
    - What differs is the *distribution we are using to calculate probability*
    - We need to know the degrees of freedom (to get the right t-curve)
- so $t(\mathrm{df}) = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

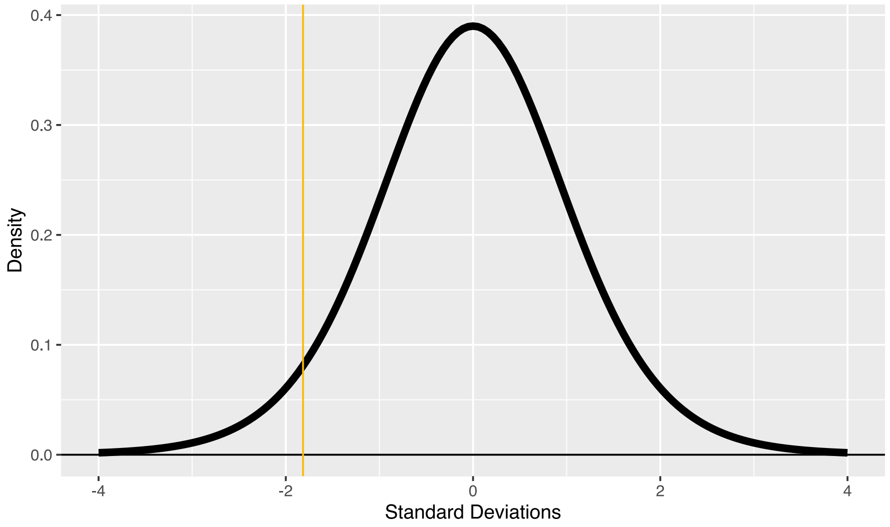
Probability According to t

- for 12 people who got a mean 6.81 hours of sleep with a sd of 2.2732
- $t(11) = \frac{6.81-8}{2.27/\sqrt{12}} = -1.816$



Probability According to t

- for 12 people who got a mean 6.81 hours of sleep with a sd of 2.2732
- $t(11) = \frac{6.81 - 8}{2.27/\sqrt{12}} = -1.816$
- instead of `pnorm()` we use `pt()` for the t distribution

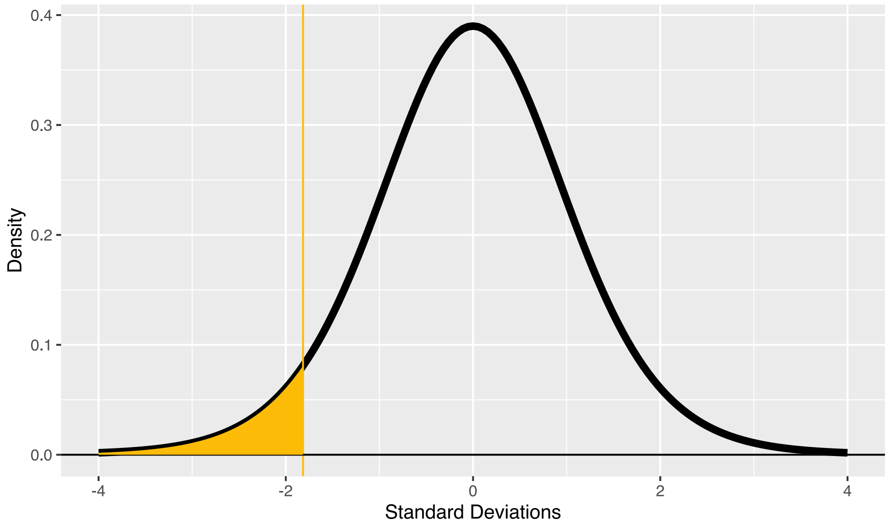


Probability According to t

- for 12 people who got a mean 6.81 hours of sleep with a sd of 2.2732
- $t(11) = \frac{6.81-8}{2.27/\sqrt{12}} = -1.816$
- instead of `pnorm()` we use `pt()` for the t distribution
- `pt()` requires the degrees of freedom:

```
pt(-1.814, df=11, lower.tail = TRUE)
```

```
## [1] 0.04851
```



Did We Have to Do All That Work?

Did We Have to Do All That Work?

No.

Did We Have to Do All That Work?

No.

```
head(m$sleep)
```

```
## [1] 7.122 9.791 5.674 8.264 8.176 7.647
```

```
t.test(m$sleep, mu=8, alternative = "less")
```

```
##
##     One Sample t-test
##
## data:  m$sleep
## t = -1.8, df = 11, p-value = 0.05
## alternative hypothesis: true mean is less than 8
## 95 percent confidence interval:
##   -Inf 7.988
## sample estimates:
## mean of x
##     6.809
```

- **One-sample** t-test
- Compares a single sample against a hypothetical mean (mu)

Types of Hypothesis

```
t.test(m$sleep, mu=0, alternative = "less")
```

- Note the use of `alternative="less"`
- This refers to the direction of our **alternative hypothesis, H_1**
  - H_1 is that our students would be getting *less* sleep than the average person.
- Can also have `alternative="greater"`...

Types of Hypothesis

```
t.test(m$sleep, mu=0, alternative = "less")
```

- Note the use of `alternative="less"`
- This refers to the direction of our **alternative hypothesis, H_1**
  - H_1 is that our students would be getting *less* sleep than the average person.
- Can also have `alternative="greater"`...
  - Our students are getting *more* sleep than the average person

Types of Hypothesis

```
t.test(m$sleep, mu=0, alternative = "less")
```

- Note the use of `alternative="less"`
- This refers to the direction of our **alternative hypothesis, H_1**
  - H_1 is that our students would be getting *less* sleep than the average person.
- Can also have `alternative="greater"`...
  - Our students are getting *more* sleep than the average person
- ...And `alternative="two.sided"`

Types of Hypothesis

```
t.test(m$sleep, mu=0, alternative = "less")
```

- Note the use of `alternative="less"`
- This refers to the direction of our **alternative hypothesis, H_1**
    - H_1 is that our students would be getting *less* sleep than the average person.
- Can also have `alternative="greater"`...
    - Our students are getting *more* sleep than the average person
- ...And `alternative="two.sided"`
    - Our students are getting *different* amounts of sleep than the average person

Putting it Together

For $t(11) = -1.816, p = 0.0483$:

If you picked 12 people at random from a population of people who get the recommended number of hours of sleep, there would be a 5% chance that their average sleep would be 6.81 hours or less

Putting it Together

For $t(11) = -1.816, p = 0.0483$:

If you picked 12 people at random from a population of people who get the recommended number of hours of sleep, there would be a 5% chance that their average sleep would be 6.81 hours or less

- Is 5% low enough for you to believe that the mean sleep probably wasn't due to chance?
- Perhaps we'd better face up to this question!

Making a Decision

- To make this decision, we use a cut-off value for p called α

Making a Decision

- To make this decision, we use a cut-off value for p called α
- α is the probability of rejecting H_0 when it actually reflects the ground truth

Making a Decision

- To make this decision, we use a cut-off value for p called α
- α is the probability of rejecting H_0 when it actually reflects the ground truth
  - If p is less than α, we can decide to *reject* H_0 and *accept* H_1
  - If p is greater than α, we *fail to reject* H_0

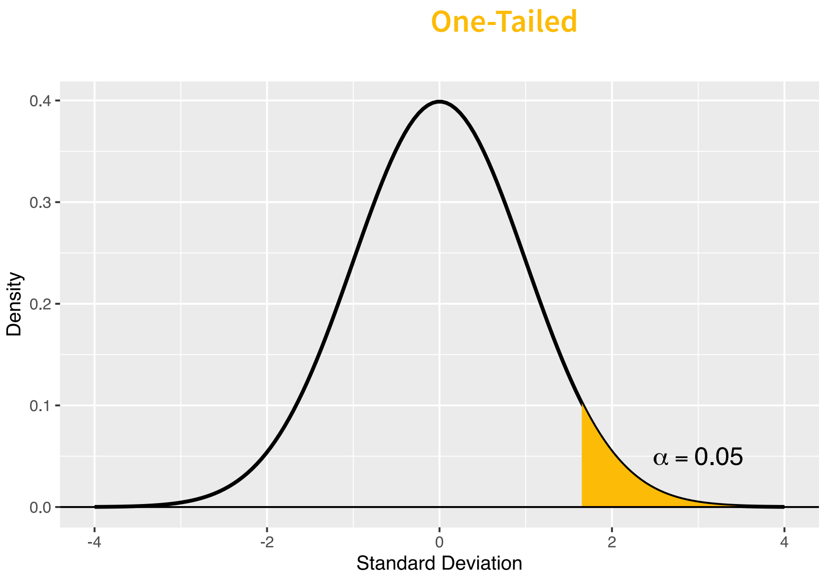
Making a Decision

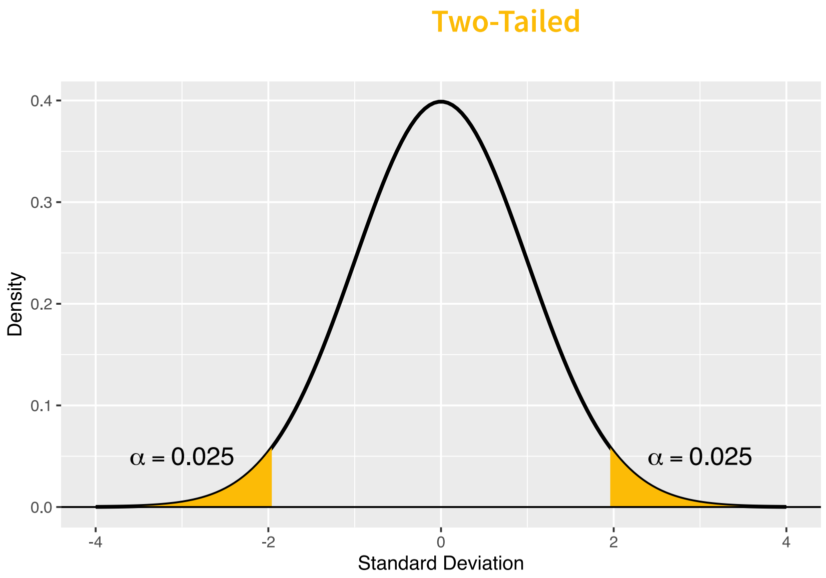
- To make this decision, we use a cut-off value for p called α
- α is the probability of rejecting H_0 when it actually reflects the ground truth
    - If p is less than α, we can decide to *reject* H_0 and *accept* H_1
    - If p is greater than α, we *fail to reject* H_0
- Typically, in Psychology, **α is set to .05**
    - We're willing to take a 5% risk of incorrectly rejecting the null hypothesis.

Making a Decision

- To make this decision, we use a cut-off value for p called α
- α is the probability of rejecting H_0 when it actually reflects the ground truth
  - If p is less than α, we can decide to *reject* H_0 and *accept* H_1
  - If p is greater than α, we *fail to reject* H_0
- Typically, in Psychology, **α is set to .05**
  - We're willing to take a 5% risk of incorrectly rejecting the null hypothesis.
- It's important to set α *before* any statistical analysis

Making a Decision





# $p < .05$

- The p-value is the probability of finding our results under H_0, the null hypothesis
- H_0 is essentially "💩 happens"
- α is the maximum level of p at which we are prepared to conclude that H_0 is false (and argue for H_1)

# $p < .05$

- The p-value is the probability of finding our results under H_0, the null hypothesis
- H_0 is essentially "💩 happens"
- α is the maximum level of p at which we are prepared to conclude that H_0 is false (and argue for H_1)

there is a 5% probability of falsely rejecting H_0

- Wrongly rejecting H_0 (false positive) is a **type 1 error**
- Wrongly failing to reject H_0 (false negative) is a **type 2 error**

Types of t-tests

- All t-tests compare two means, but with different group constraints

Types of t-tests

- All t-tests compare two means, but with different group constraints
- **One-sample t-test**

Types of t-tests

- All t-tests compare two means, but with different group constraints
- **One-sample t-test**
  - Compares the mean from a range of scores to a specific value

Types of t-tests

- All t-tests compare two means, but with different group constraints
- **One-sample t-test**
    - Compares the mean from a range of scores to a specific value
    - Lets you examine whether the mean of your data is significantly different from a set value

Types of t-tests

- All t-tests compare two means, but with different group constraints
- One-sample t-test
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- Independent-samples t-test

Types of t-tests

- All t-tests compare two means, but with different group constraints
- **One-sample t-test**
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- **Independent-samples t-test**
  - Compares the means of two independent groups

Types of t-tests

- All t-tests compare two means, but with different group constraints
- **One-sample t-test**
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- **Independent-samples t-test**
  - Compares the means of two independent groups
  - Lets you examine whether two groups differ significantly from each other on the variable of interest

Types of t-tests

- All t-tests compare two means, but with different group constraints
- One-sample t-test
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- Independent-samples t-test
  - Compares the means of two independent groups
  - Lets you examine whether two groups differ significantly from each other on the variable of interest
- Paired-samples t-test

Types of t-tests

- All t-tests compare two means, but with different group constraints
- One-sample t-test
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- Independent-samples t-test
  - Compares the means of two independent groups
  - Lets you examine whether two groups differ significantly from each other on the variable of interest
- Paired-samples t-test
  - Compares means that are paired in some way

Types of t-tests

- All t-tests compare two means, but with different group constraints
- One-sample t-test
  - Compares the mean from a range of scores to a specific value
  - Lets you examine whether the mean of your data is significantly different from a set value
- Independent-samples t-test
  - Compares the means of two independent groups
  - Lets you examine whether two groups differ significantly from each other on the variable of interest
- Paired-samples t-test
  - Compares means that are paired in some way
  - Allows you to compare measures that come from the same individual, e.g.

End