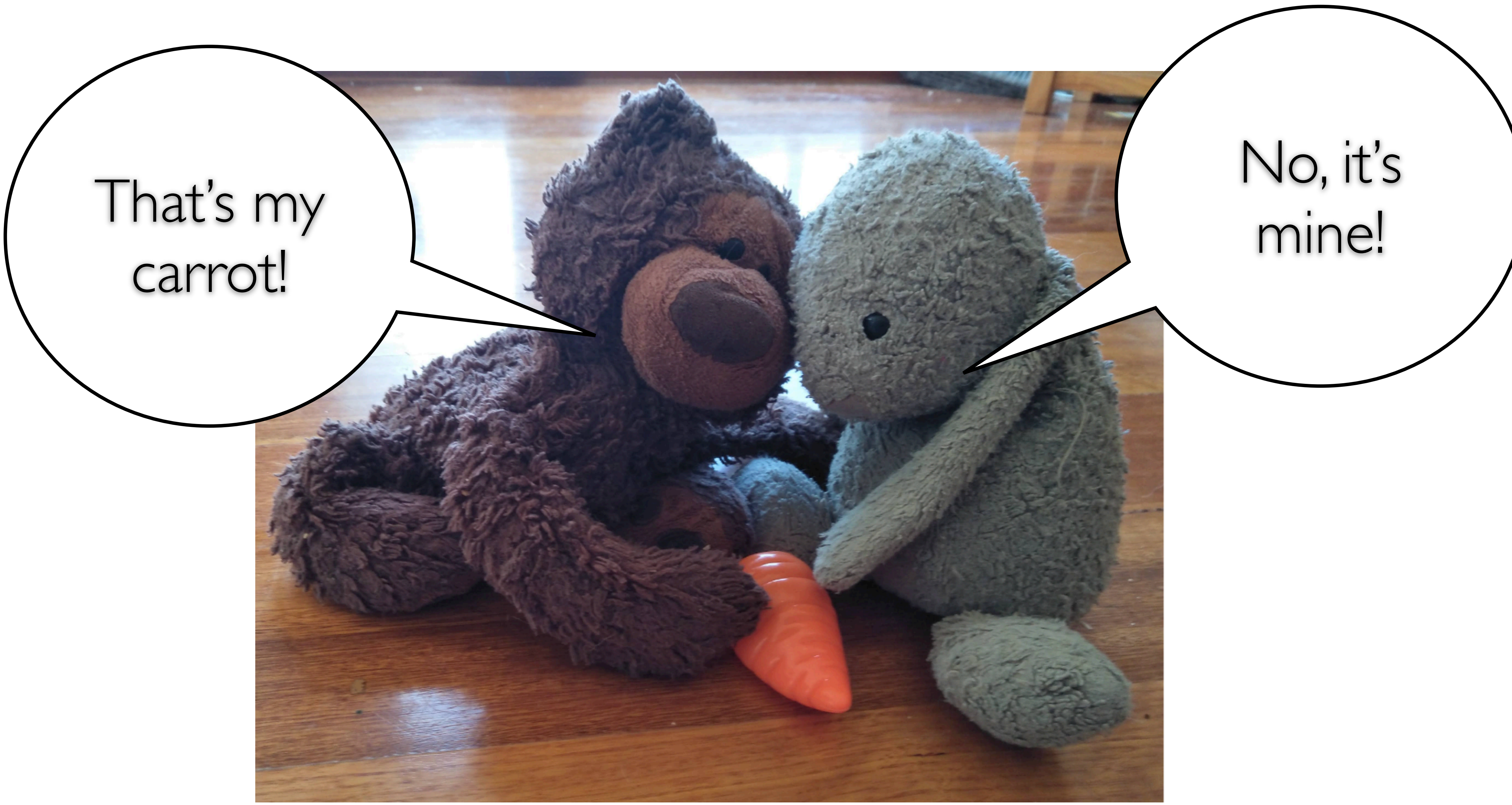
T-tests:
One sample theory

Research Methods for Human Inquiry

Andrew Perfors

Today's story…

Everyone in Bunnyland is going hungry, and they are starting to fight about it



Today's story…

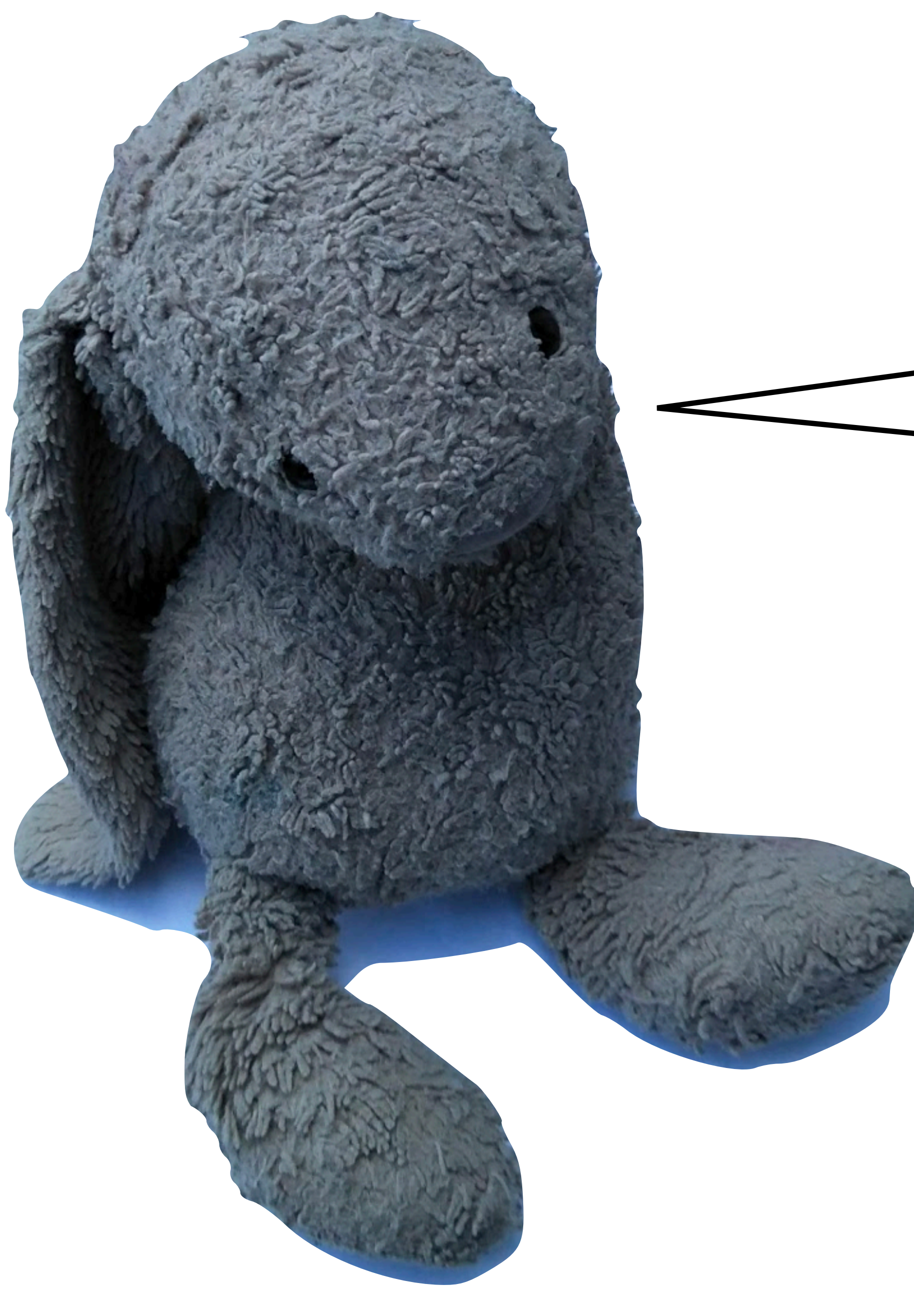
Most people think they need to take some sort of aggressive action against the Others

Today's story…

But some advise caution - there's no evidence the Others are responsible

Today's story…

Today's story…



I'm scared all right,
but I'm *more* scared of
us running out of food or
fighting with each other
when that's not even
the real problem

Today's story…

And *you* are the one
that's acting irrationally:
you're blinded by anger

Today's story…

Today's story…

Hey… I have an idea.
Let's each take the CR test
to make sure we're not
acting emotionally

Shadow’s CR test

Measures Calmness and Rationality. If everyone takes it, we can figure out who is acting irrationally and being emotion driven!

But how do you compare things like this?

Bunny’s CR score: 70

??

Sometimes you want to know how single scores compare in a somewhat objective way

But how do you compare things like this?

Bunny's CR score: 70

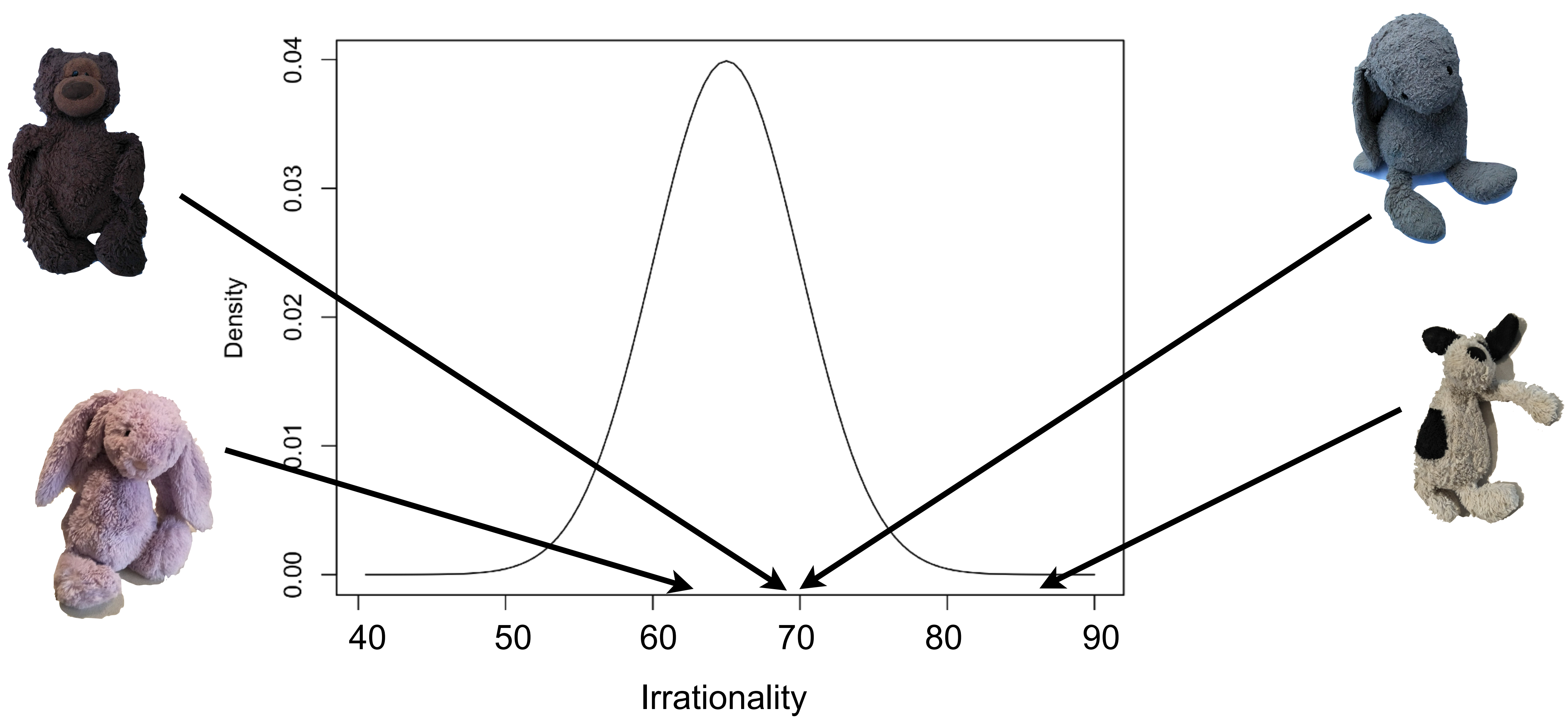
??

- Even knowing the scale doesn't help much…
  - Bunny answered 70 of 100 questions in an "irrational" way
  - Does that answer the question satisfactorily?
- For <u>her</u> CR score to make sense, we need to know what <u>everyone else's</u> CR scores were!

Sometimes you want to know how single scores compare in a somewhat objective way

This helps a lot

- Bunny's score: 70
- Population mean: μ = 64
- Population std dev: σ = 12



This helps a lot

- Bunny's score: 70
- Population mean: $\mu = 64$
- Population std dev: $\sigma = 12$

- Much more informative, right?
  - Bunny is 0.5 standard deviations above the mean CR score in the population
  - Or, she has a <u>standardised CR score of *z = 0.5*</u>

Standard scores

my standard score

my raw score

the population mean

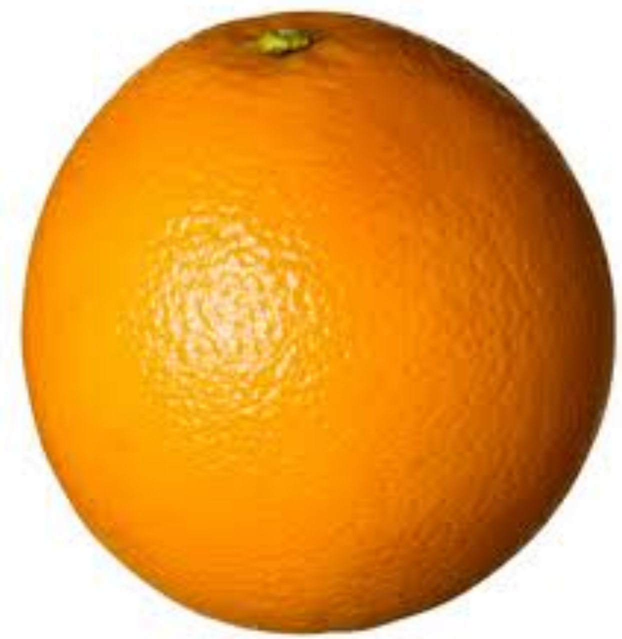
$$z = \frac{X - \mu}{\sigma}$$

the population standard deviation

All z-scores have mean 0 and standard deviation 1, by definition

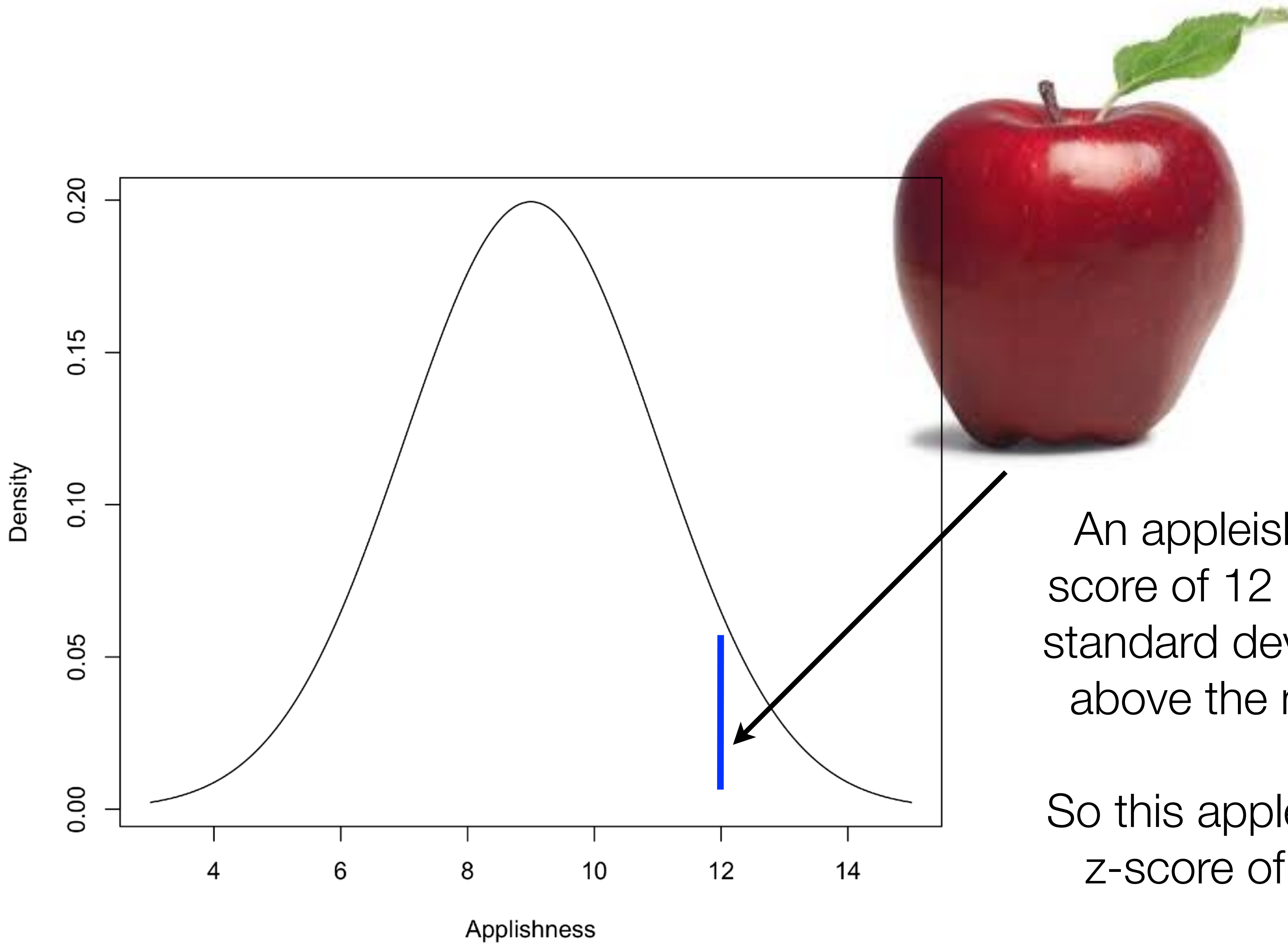
We can use this to compare two things too…

this apple has an apple-ness of 12



this orange has an orange-ness of 205

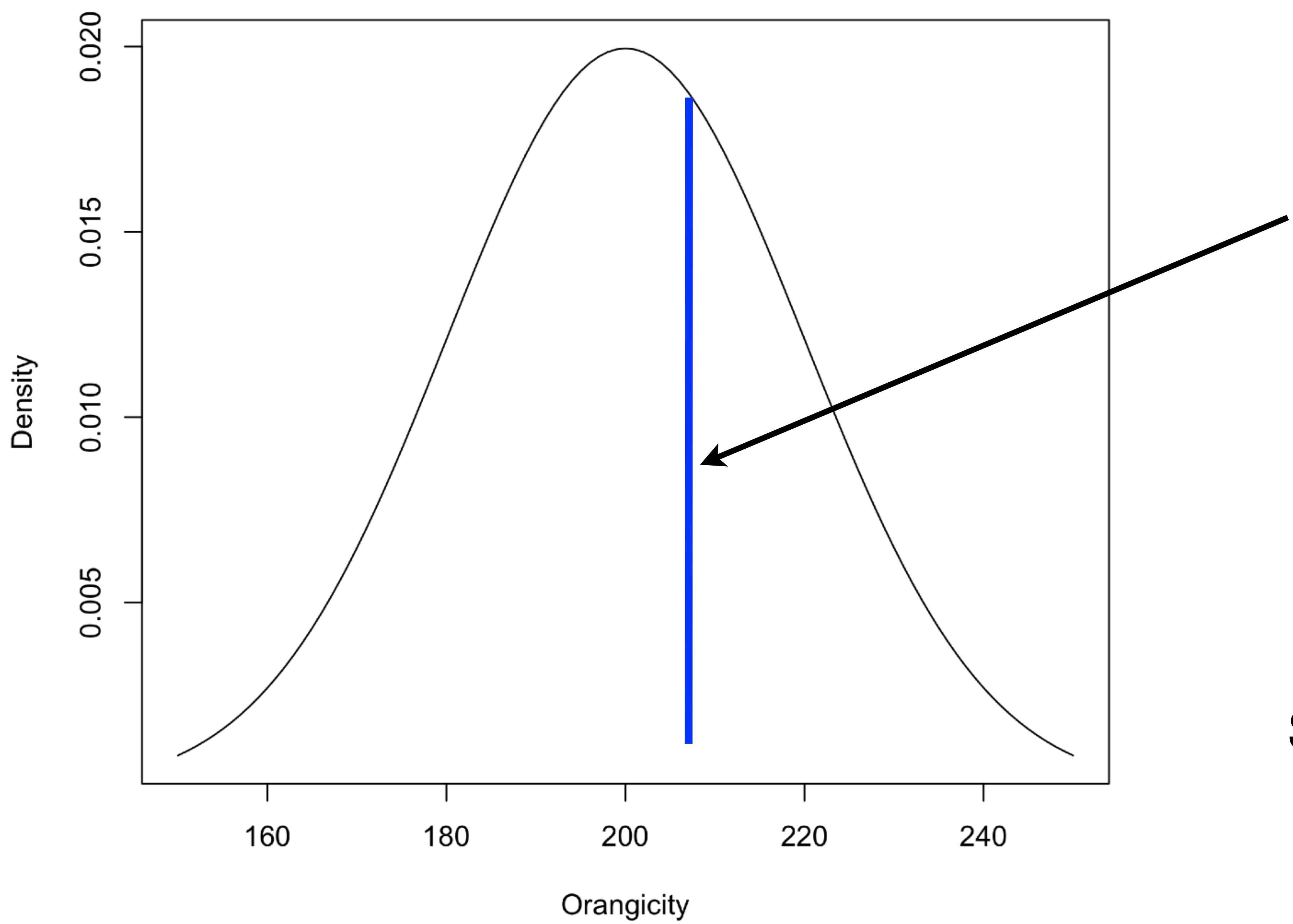
So… we begin by comparing our apple to other apples…



An appleishness score of 12 lies 1.5 standard deviations above the mean.

So this apple has a z-score of +1.5

And then we compare our orange to other oranges



An organgicity of 205 is only 0.25 standard deviations above the mean

So this orange has a standard score of $z = +0.25$

Comparing apples to oranges using standard scores actually makes sense!

The apple is more applish than the orange is orangish

this apple has a z-score of +2.5

this orange has an z-score of +0.25

This is great in psychology…

- Psychology is filled with situations when we want to make apples-to-oranges comparisons

- Examples:
  - I'm less introverted ($z = -1.0$) than nerdy ($z = +2.5$)
  - Voldemort is more evil ($z = +8$) than ugly ($z = +4.5$)

(This is also what adjusted residuals reveal)

- In Week 6 Day 2 we talked about adjusted residuals for chi-squared tests — they are conceptually equivalent to z-scores (different calculation though since based on E and O, not μ)

```
> ctResult$stdres
             box1       box2
doggie -0.6654653  0.6654653
gladly  0.4077840 -0.4077840
shadow  0.4968698 -0.4968698
```

Returns the adjusted residuals

These are the raw residuals divided by the standard error $\sqrt{E(1 - O_i/N)(1 - O_j/N)}$

Essentially what this does is allow us to interpret these as *kind of* indicating how many "standard deviations" away individual cells are.

A (weak) rule of thumb some use is that if you have an overall significant test and any adjusted residuals are more extreme than +/- 1.96 or so, those individual items are significant as well

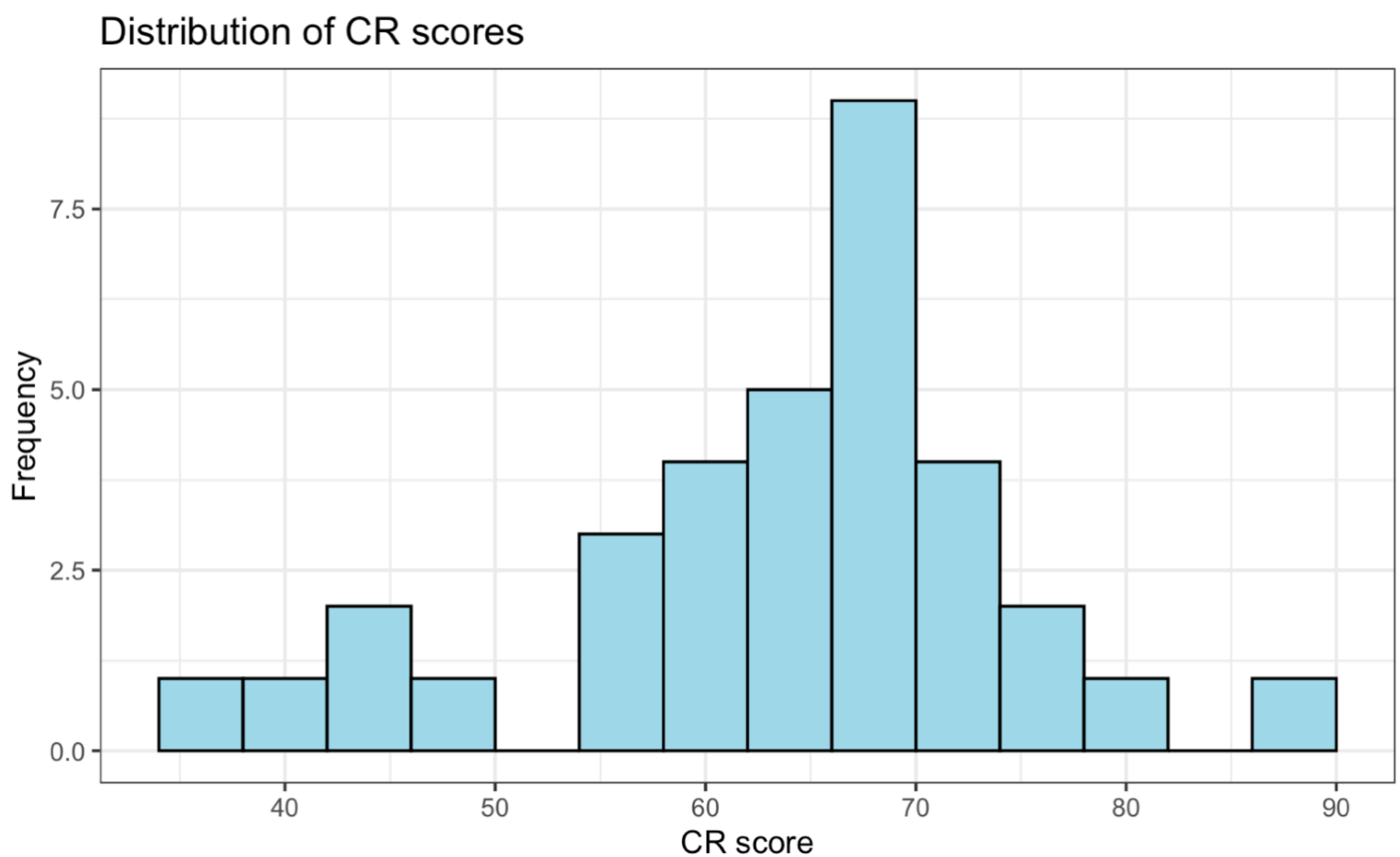
Comparing your data against a theoretically predicted mean using a <u>one sample t-test</u>

Some data

```
> dcr <- read_csv(file=here("crscores.csv"))
> dcr
# A tibble: 34 x 2
   name            cr
   <chr>        <dbl>
 1 bunny           70
 2 gladly          69
 3 flopsy          63
 4 doggie          87
 5 nosey           67
 6 cuddly paws     56
 7 shadow          56
 8 pink bunny      59
 9 purple bunny    68
10 blue bunny      55
# … with 24 more rows
```

Always have a look at it first!!

The CR test is designed to have a mean of 50 and sd of 10…
Does this describe Bunnyland, or are they more generally irrational as a population right now?



Are these two means the same???

	sample	theoretical population
our mean	63.6	50
our standard deviation	11.6	10

Let’s construct a test….

1) A diagnostic test statistic, *T*
2) Sampling distribution of *T* if the null is true
3) The observed *T* in your data
4) A rule that maps every value of *T* onto a decision (accept or reject H0)

A diagnostic test statistic

the two obvious quantities
of interest…

$$\bar{X} \qquad \mu$$

the sample mean we
actually observed

true population mean if the
null hypothesis is correct

A diagnostic test statistic

the difference between them should be pretty close to zero if the null hypothesis is actually true

$$\frac{\bar{X} - \mu}{\sigma}$$

ideally we could just convert this into a standard z score, by dividing by the true population standard deviation if the null hypothesis is correct…

but we don't know what that is!

A diagnostic test statistic

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

instead we convert
it to something *like*
a standard score,
by dividing it by the
SEM (which we
can calculate)

this is called *t*, and it is
our test statistic

this has mean of about
zero and standard
deviation of about one

Let's construct a test….

t ✔ 1) A diagnostic test statistic, *T*
2) Sampling distribution of *T* if the null is true
6.86 ✔ 3) The observed *T* in your data
4) A rule that maps every value of *T* onto a decision (accept or reject H0)

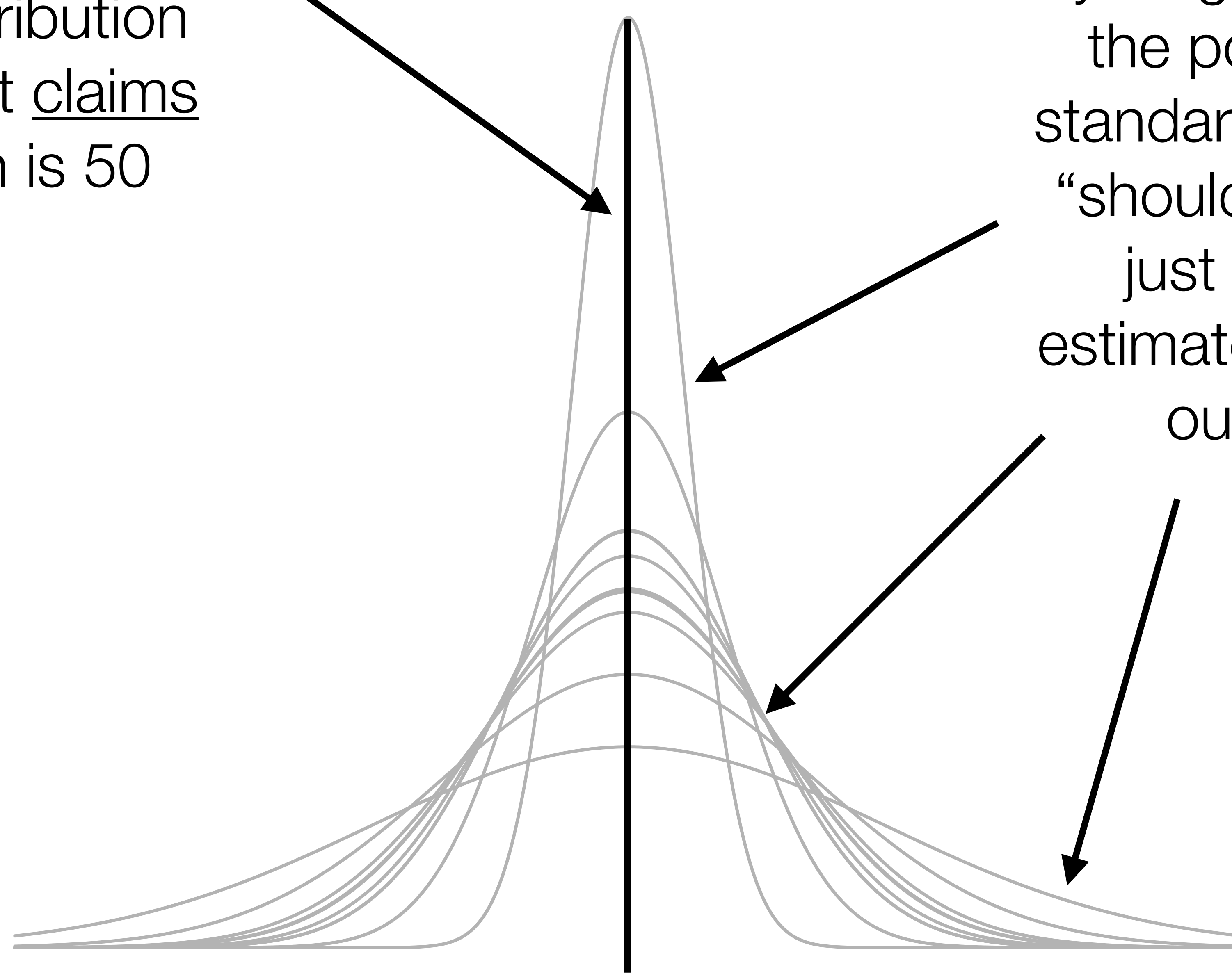
$$t = \frac{\bar{X} - \mu}{\hat{\sigma} / \sqrt{N}}$$

$$= \frac{63.6 - 50}{11.6/\text{sqrt}(34)} = 6.86$$

```
(mean(dcr$cr) - 50) / (sd(dcr$cr)/sqrt(nrow(dcr)))
```

Sampling distribution of t

What does our null hypothesis actually say? It assumes that the population distribution is normal, and it claims that the mean is 50

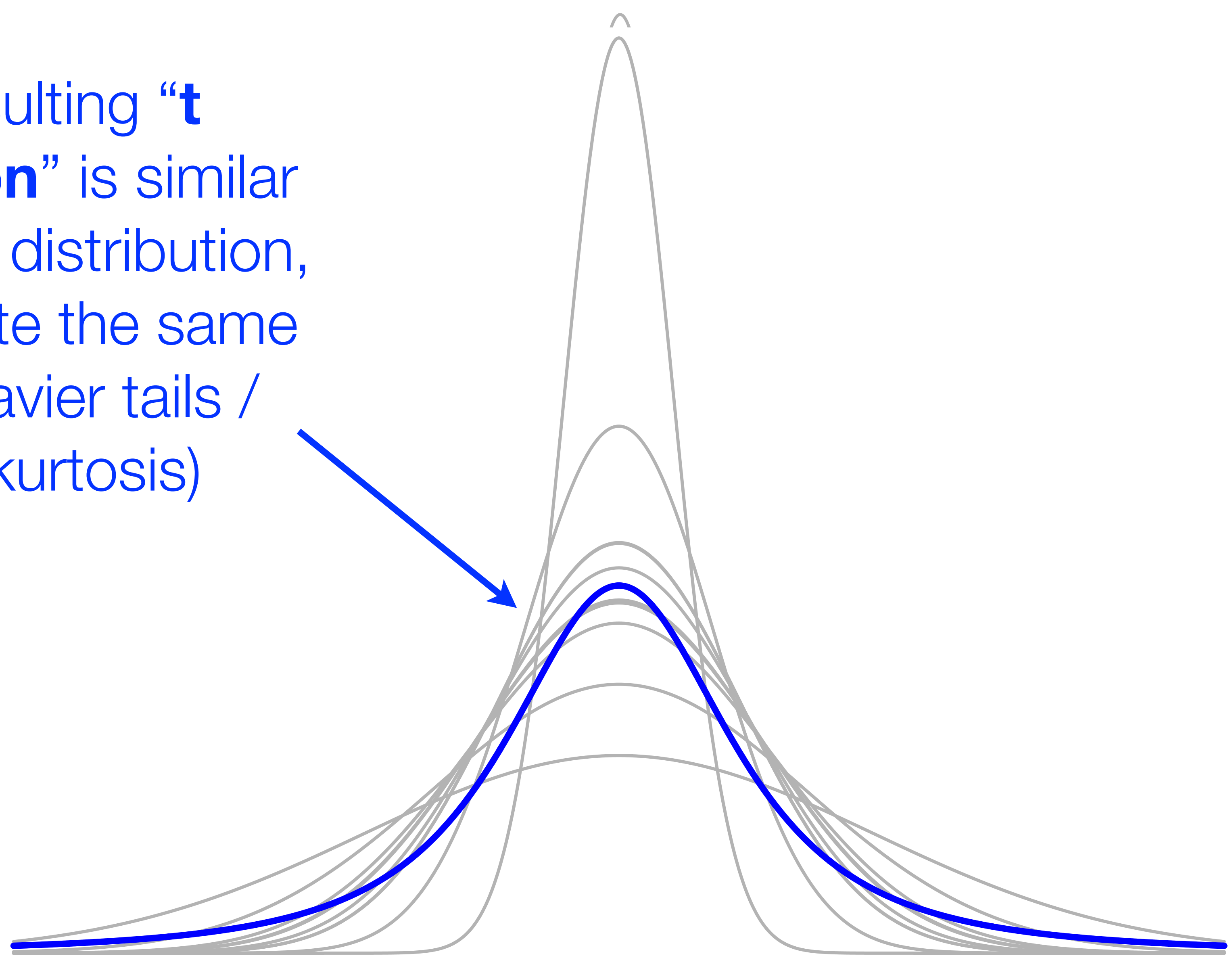


It doesn’t say anything about what the population standard deviation “should” be. We just have to estimate that from our data

Sampling distribution of t

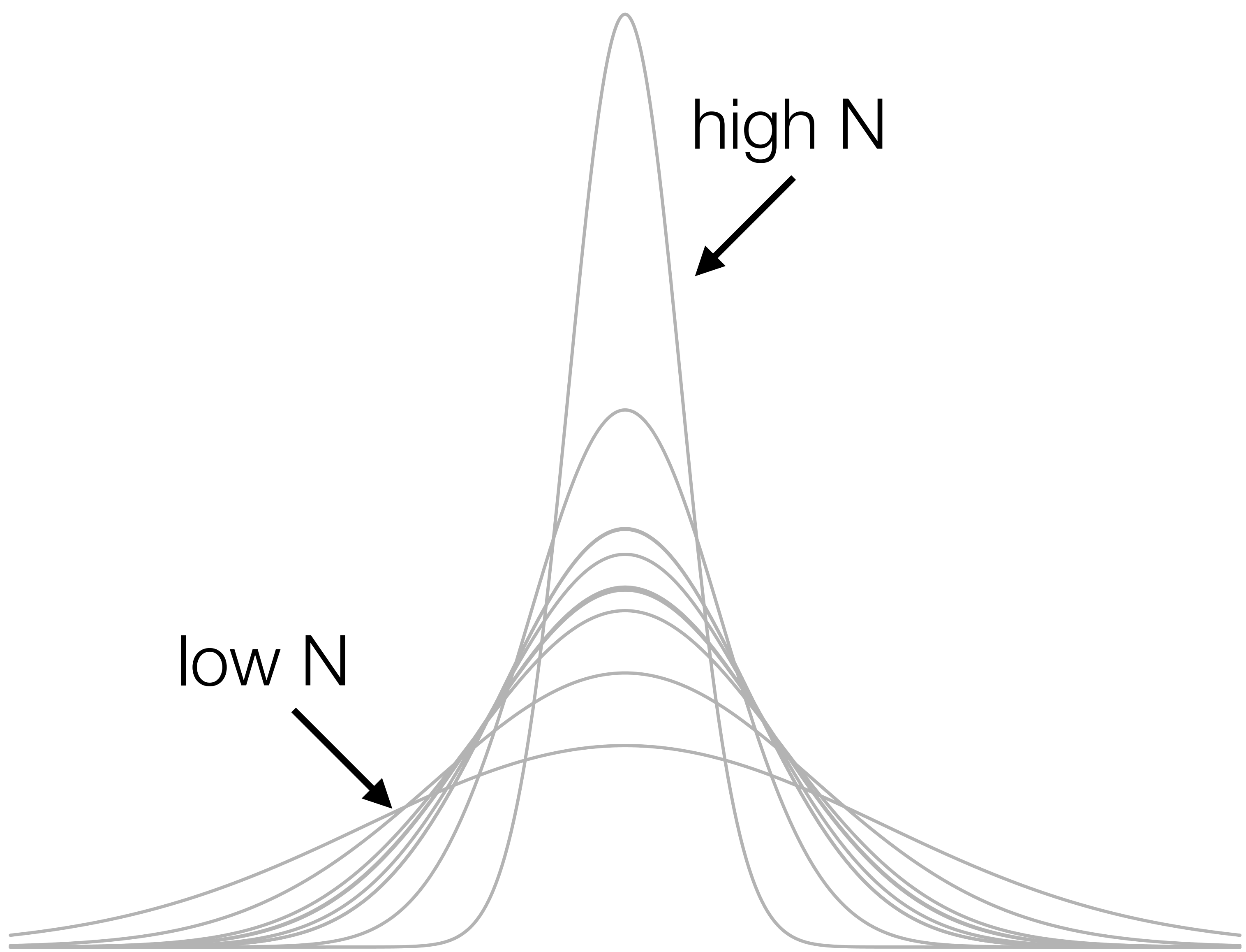
Since we're unsure about what the true standard deviation is, the overall sampling distribution for the t statistic is obtained by *averaging over lots of possible choices* for the population standard deviation

The resulting "**t distribution**" is similar to a normal distribution, but not quite the same (it has heavier tails / higher kurtosis)



Sampling distribution of t

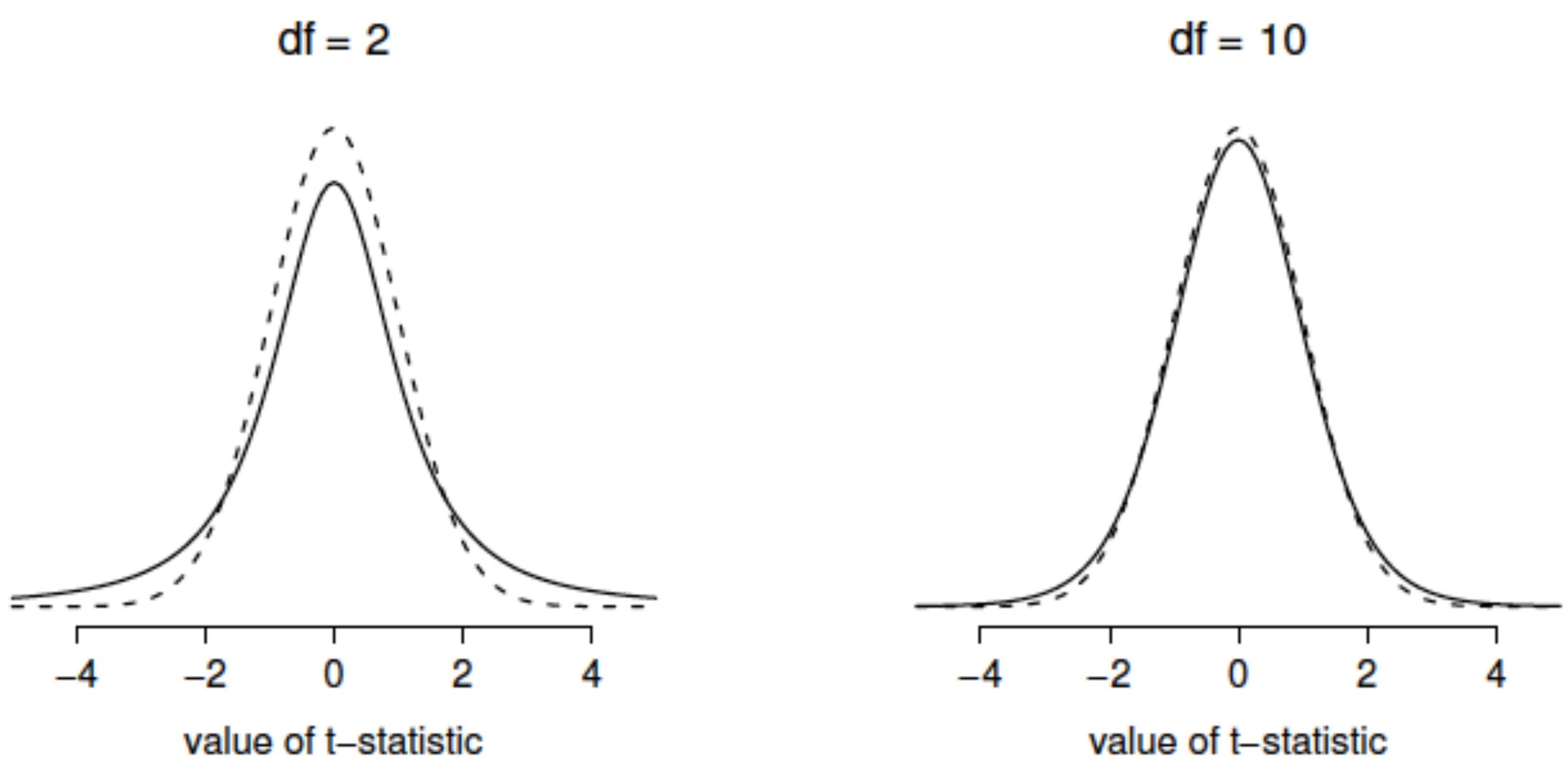
However, as sample size (N) gets larger, our estimate of the standard deviation is more precise, so the t-distribution grows more similar to normal



Sampling distribution of t

However, as sample size (N) gets larger, our estimate of the standard deviation is more precise, so the t-distribution grows more similar to normal

This is reflected in the *degrees of freedom*, which for a t distribution is N-1.
(N = # of data points; 1 = # of constraints, i.e., the mean)



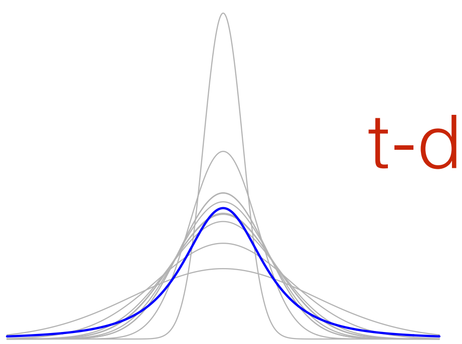
higher degrees of freedom / larger sample means the t-distribution looks more normal

Let's construct a test….

✔ 1) A diagnostic test statistic, *T*

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

✔ 2) Sampling distribution of *T* if the null is true



t-distribution with N-1 degrees of freedom

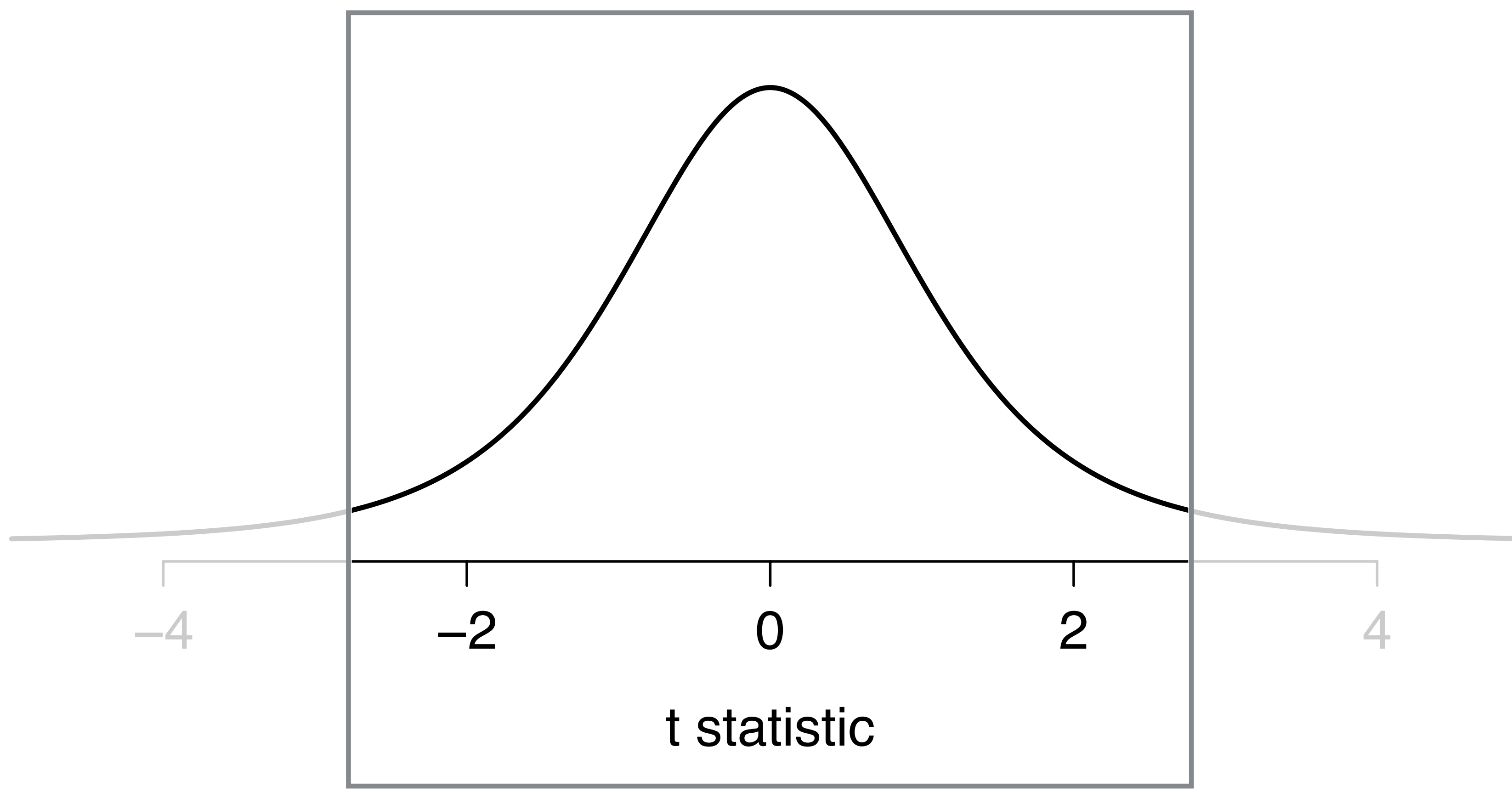
✔ 3) The observed *T* in your data

6.86

4) A rule that maps every value of *T* onto a decision (accept or reject H0)

A rule for mapping values of t onto a decision

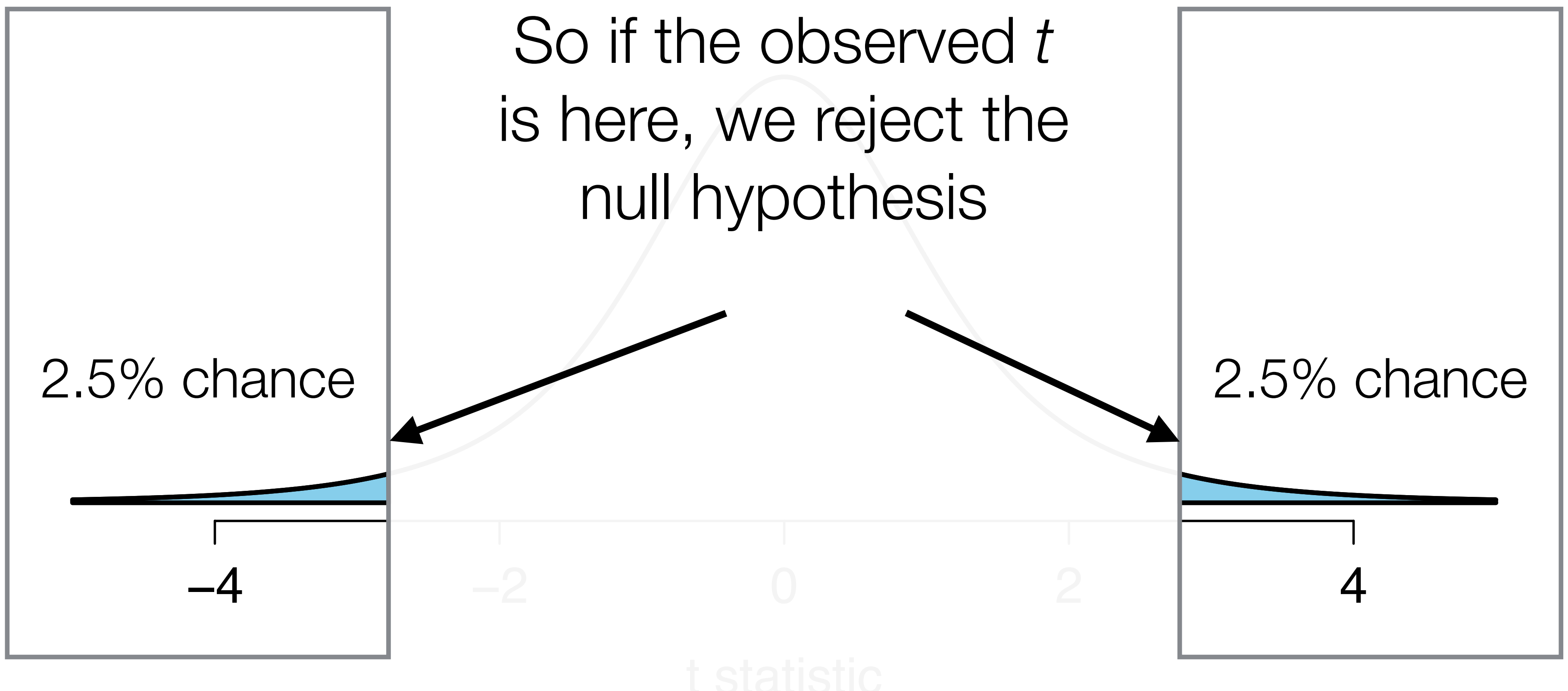
If the null hypothesis is true, there’s a 95% chance that the t-statistic falls in this range



If the null hypothesis is true, we expect the t-statistic to have come from this sampling distribution…

A rule for mapping values of t onto a decision

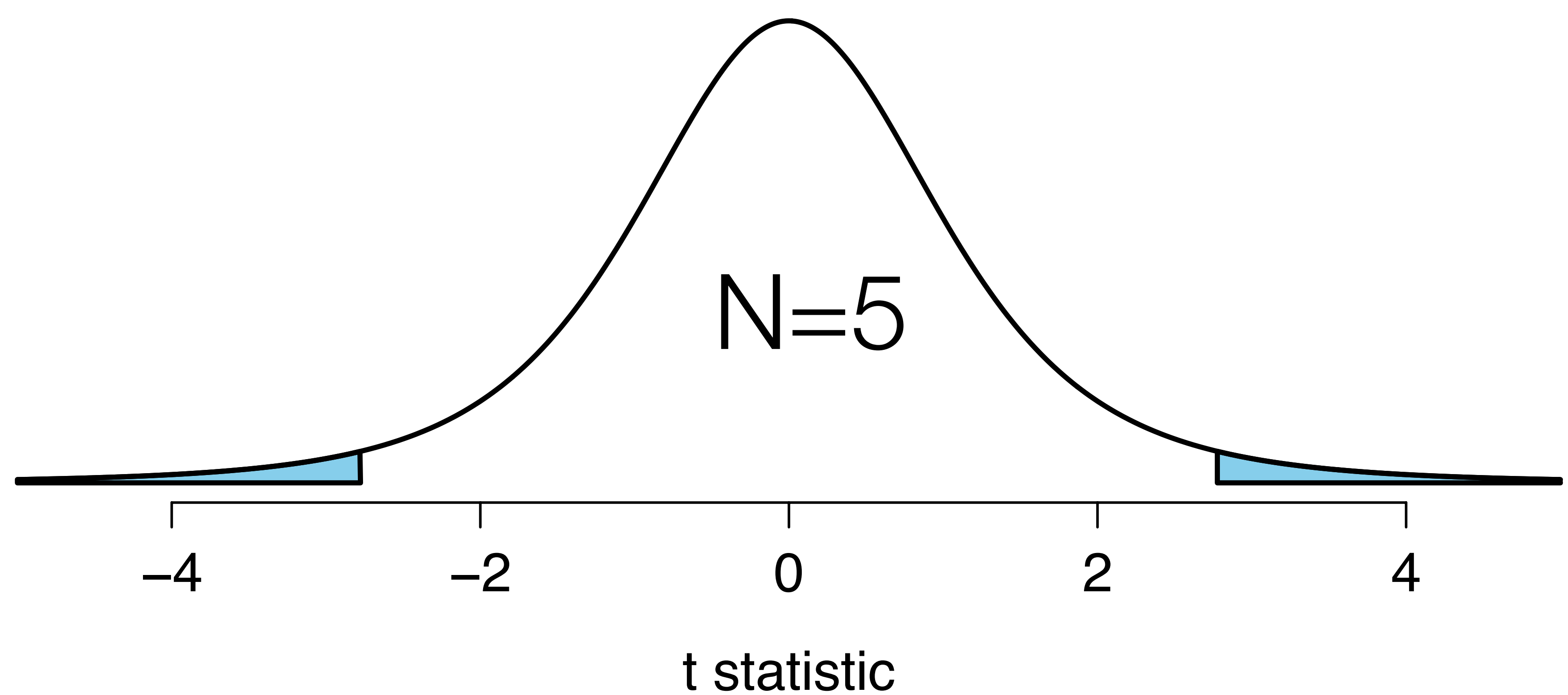
If the null hypothesis is true, there's a 95% chance that the t-statistic falls in this range



If the null hypothesis is true, we expect the t-statistic to have come from this sampling distribution…

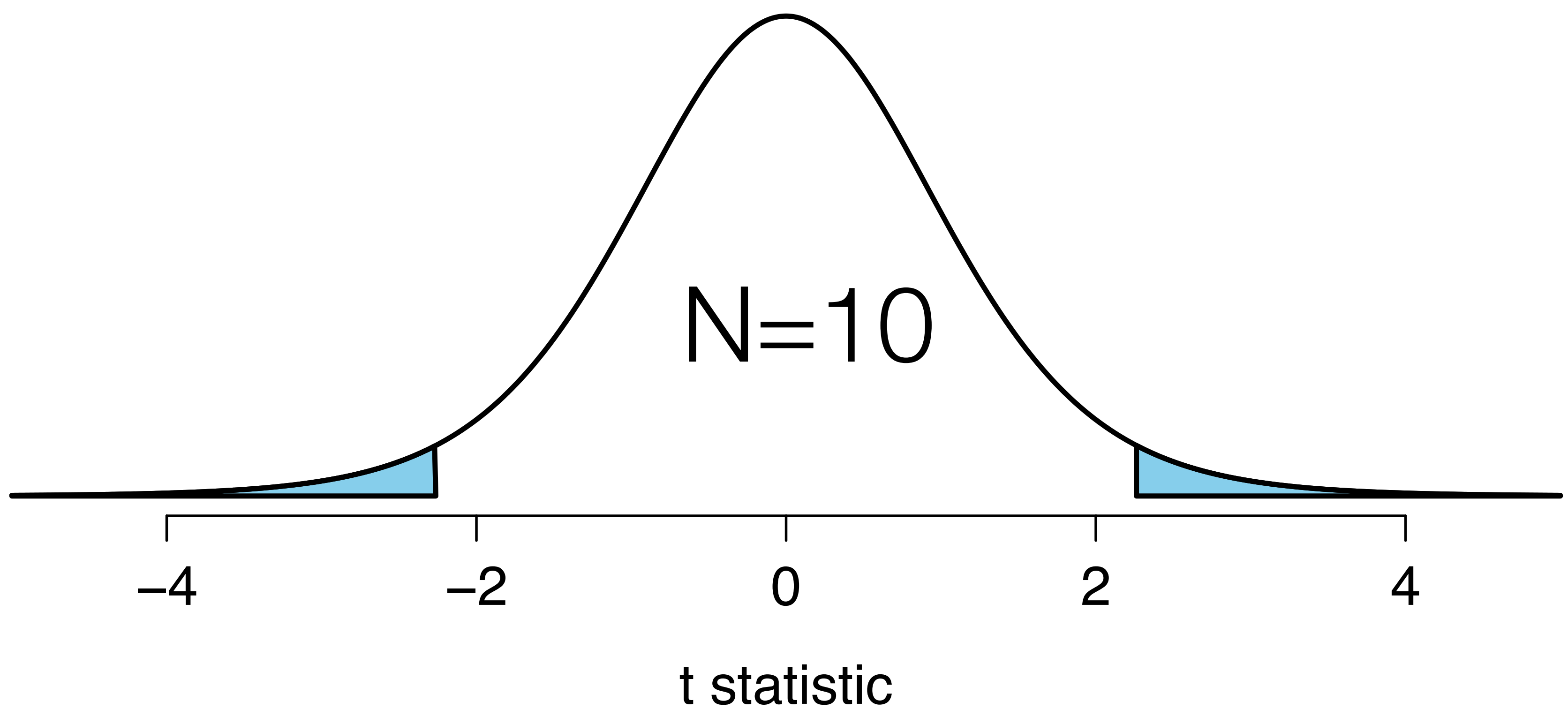
A rule for mapping values of t onto a decision

Because the width of the t distribution depends on sample size, the size of the rejection region changes as N increases…



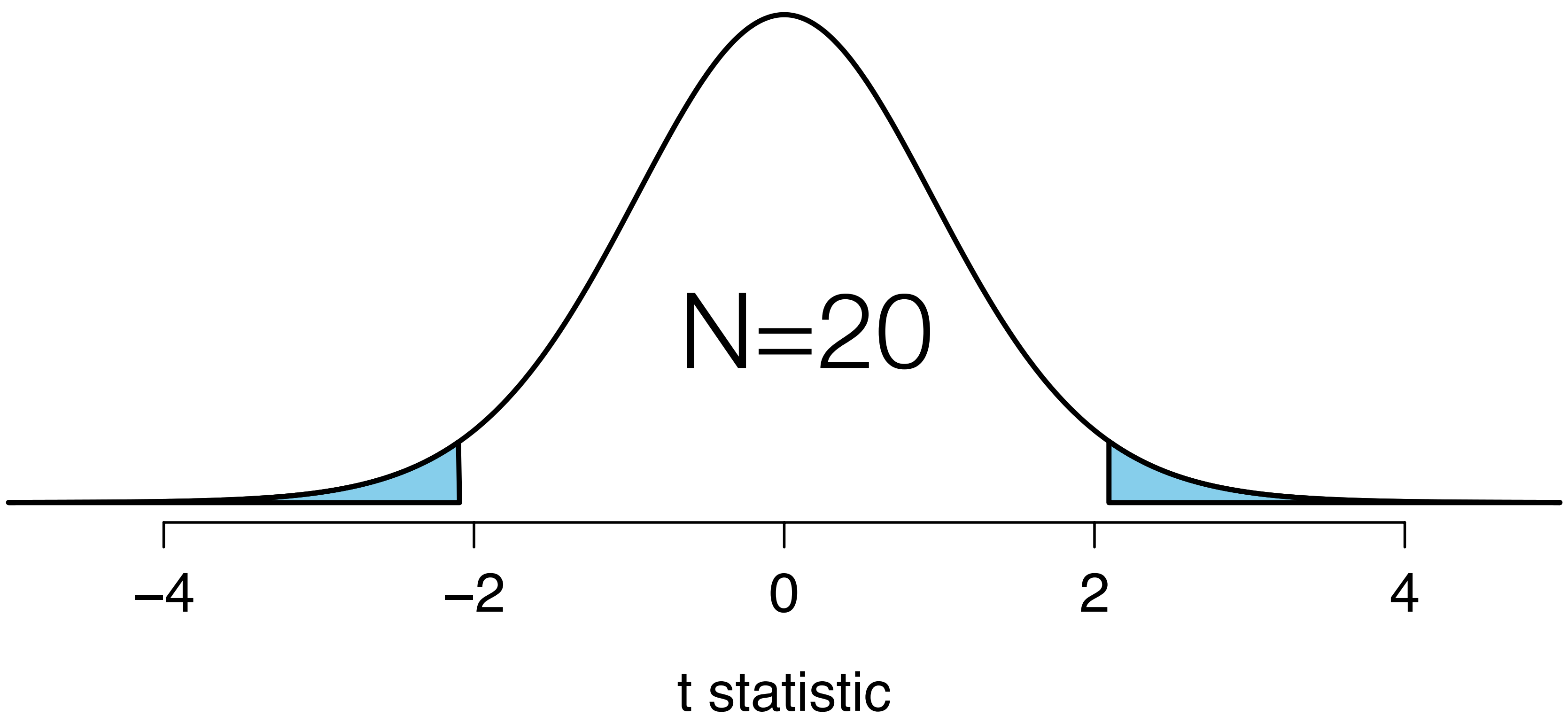
A rule for mapping values of t onto a decision

Because the width of the t distribution depends on sample size, the size of the rejection region changes as N increases…



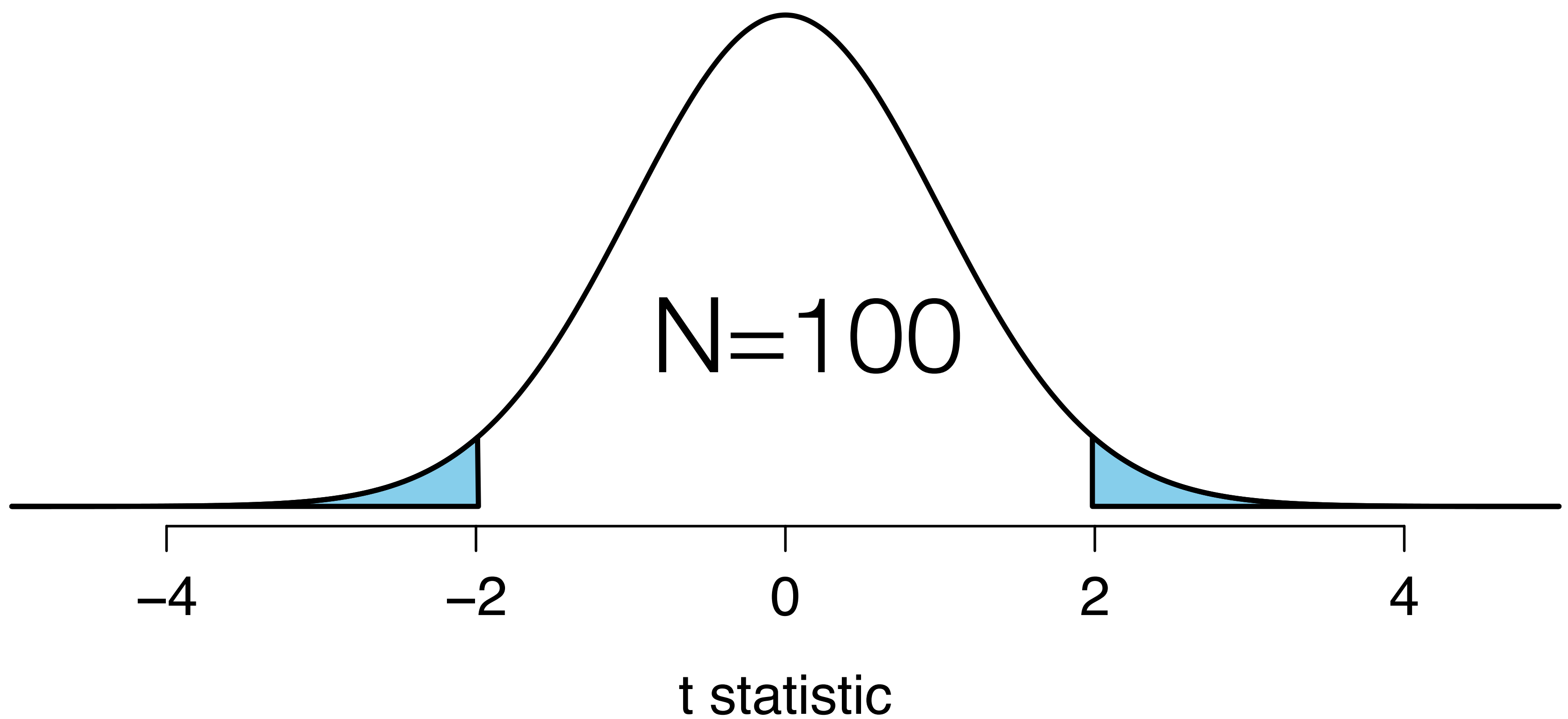
A rule for mapping values of t onto a decision

Because the width of the t distribution depends on sample size, the size of the rejection region changes as N increases…



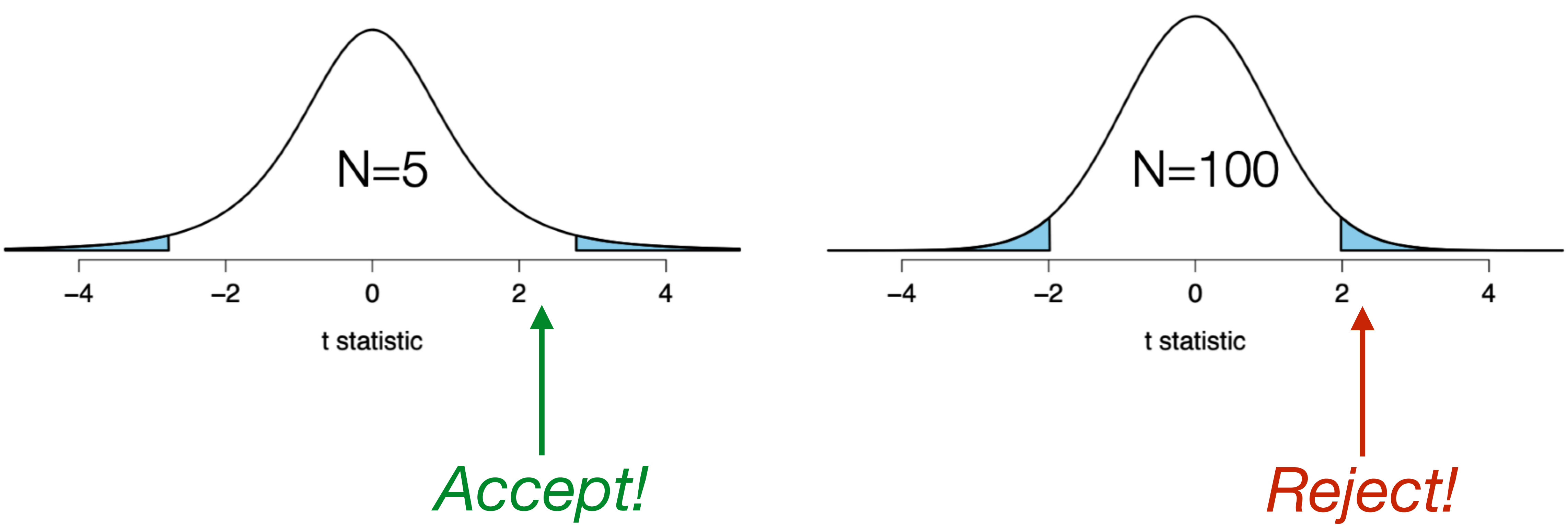
A rule for mapping values of t onto a decision

Because the width of the t distribution depends on sample size, the size of the rejection region changes as N increases…



A rule for mapping values of t onto a decision

Because the width of the t distribution depends on sample size, the size of the rejection region changes as N increases…



This means that whether a given t-statistic is significant depends on sample size!

Exercises are in w7day1exercises.Rmd