Poli-301: The Political Science Discipline

TODAY'S AGENDA

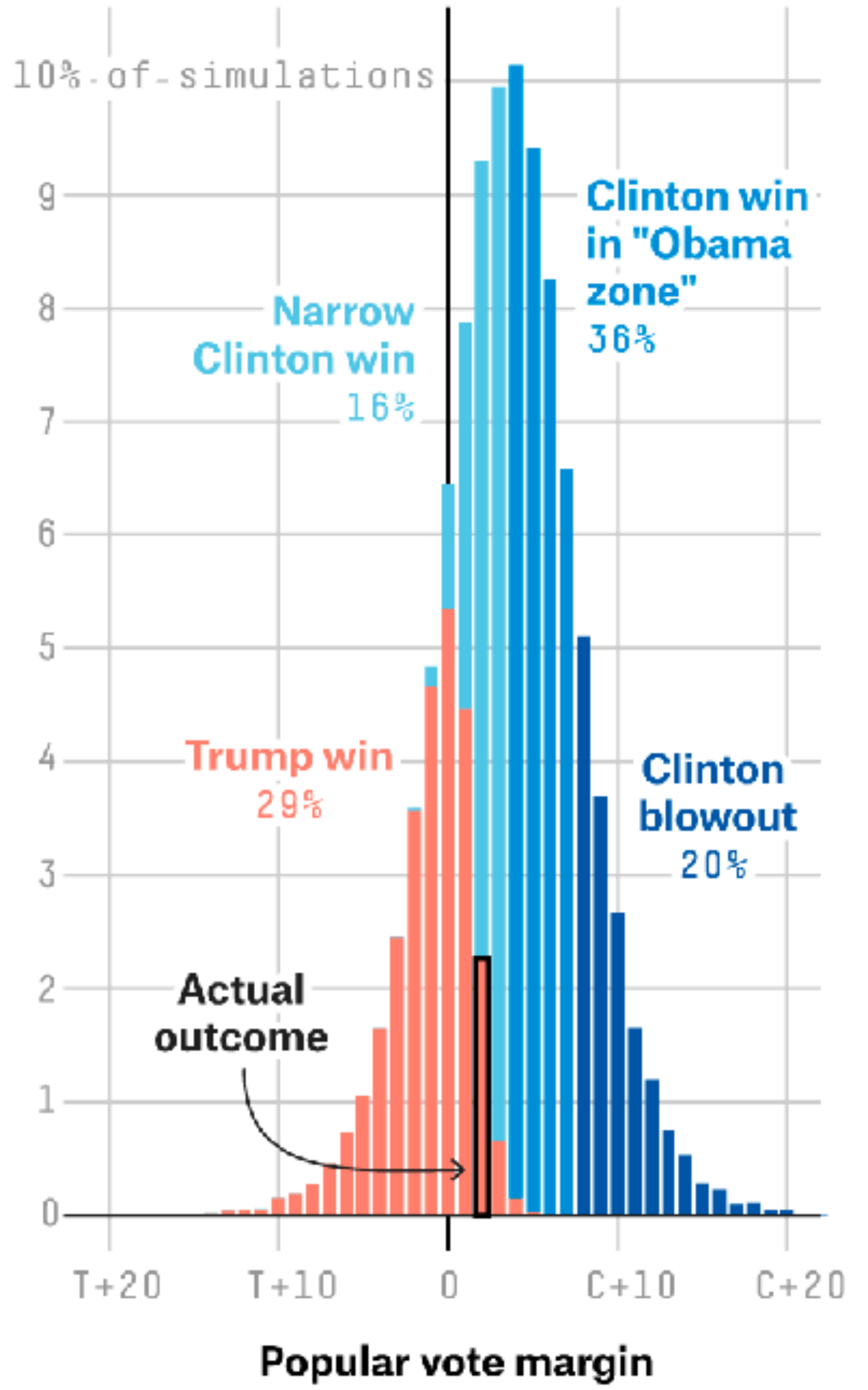
1. Uncertainty
2. Sampling
3. Bootstrapping and confidence intervals

Last week a poll came out showing Buttigieg leading at 25% in Iowa; **how certain are we of this?**

We also found that the Mariel Boatlift increased wages in Miami (relative to similar cities); **How certain are we of this?**

FiveThirtyEight's final forecast for the 2016 election

Likelihood of popular vote outcomes according to FiveThirtyEight's polls-only model at 9:35 a.m. on Election Day 2016. Based on 20,000 simulations.



FiveThirtyEight

Why are we **uncertain**?

We only have a **sample**,
but we're interested in a **population**

Polls often ask a couple thousand people (if that!), and try to **infer** something bigger

Each sample is going to give us different results!

Even when we have all the data, we can still be **uncertain**

Do taller/shorter Presidents tend to be more popular?

Collect heights of all Presidents, collect avg. approval rating during tenure

Problem: we might have all the data, but few Presidents + approval ratings are samples!

Solution

Turns out that if a sample is random, representative, and large…

… the sample mean will be pretty close to the population mean

Key terms

Population

All of the instances of the thing we care about

Sample

A subset of the population

Population paramater

The true percent of men and women who voted for Obama

Point Estimate/ sample statistic

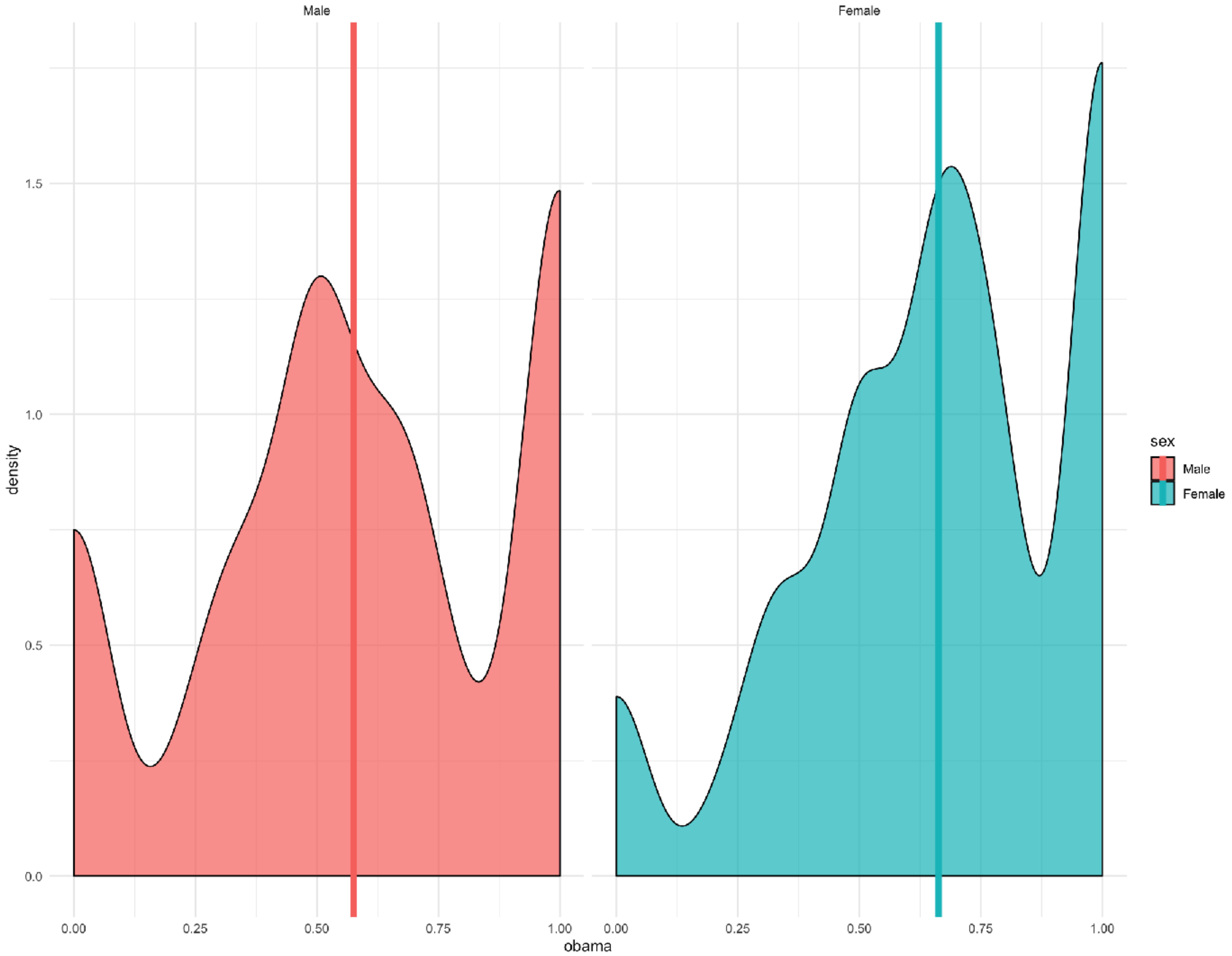
The average in each of those 10 random draws

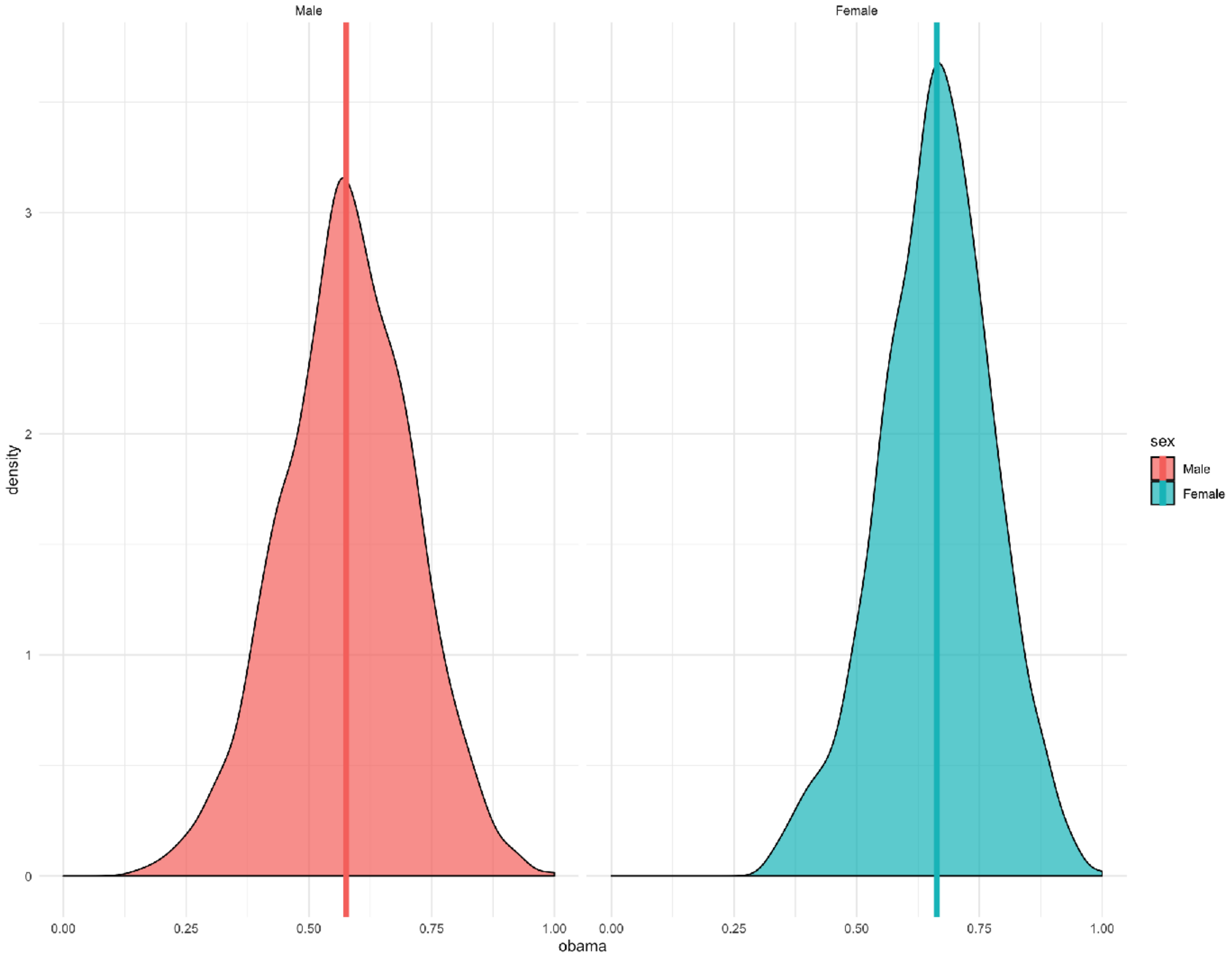
Example

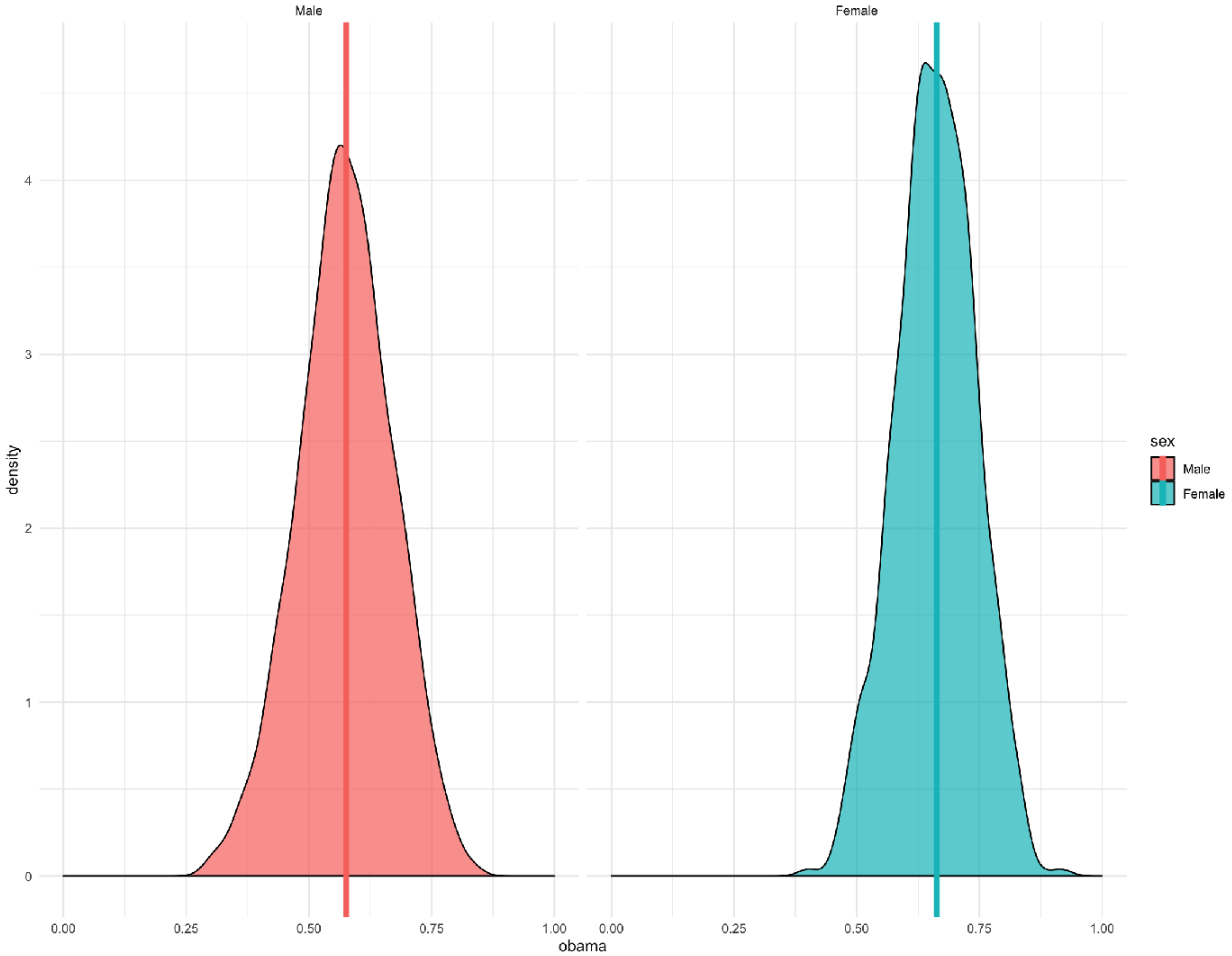
Let's imagine that there are only 2,867 people in the US and they were all surveyed in **gss_sm**

What's the **true** percent of men and women who voted for Obama?

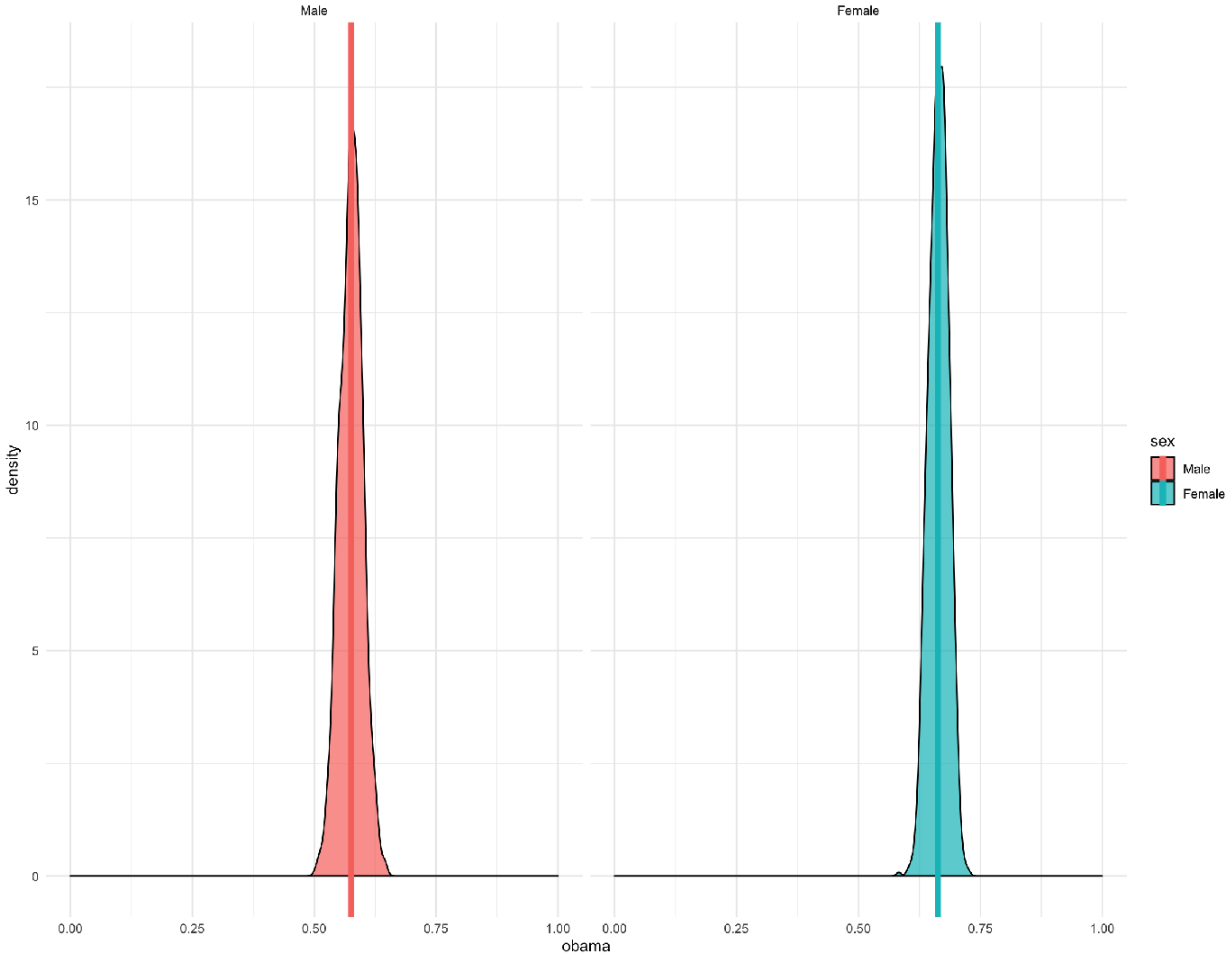
Now take a **sample** of 20 people and calculate the same percent again


Male
Female
density
obama
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
sex
Male
Female


Male
Female
density
obama
0
1
2
3
0.00
0.25
0.50
0.75
1.00
sex
Male
Female



Male
Female
density
4
3
2
1
0
0.00
0.25
0.50
0.75
1.00
obama
sex
Male
Female



Male
Female
density
obama
0
5
10
15
0.00
0.25
0.50
0.75
1.00
sex
Male
Female

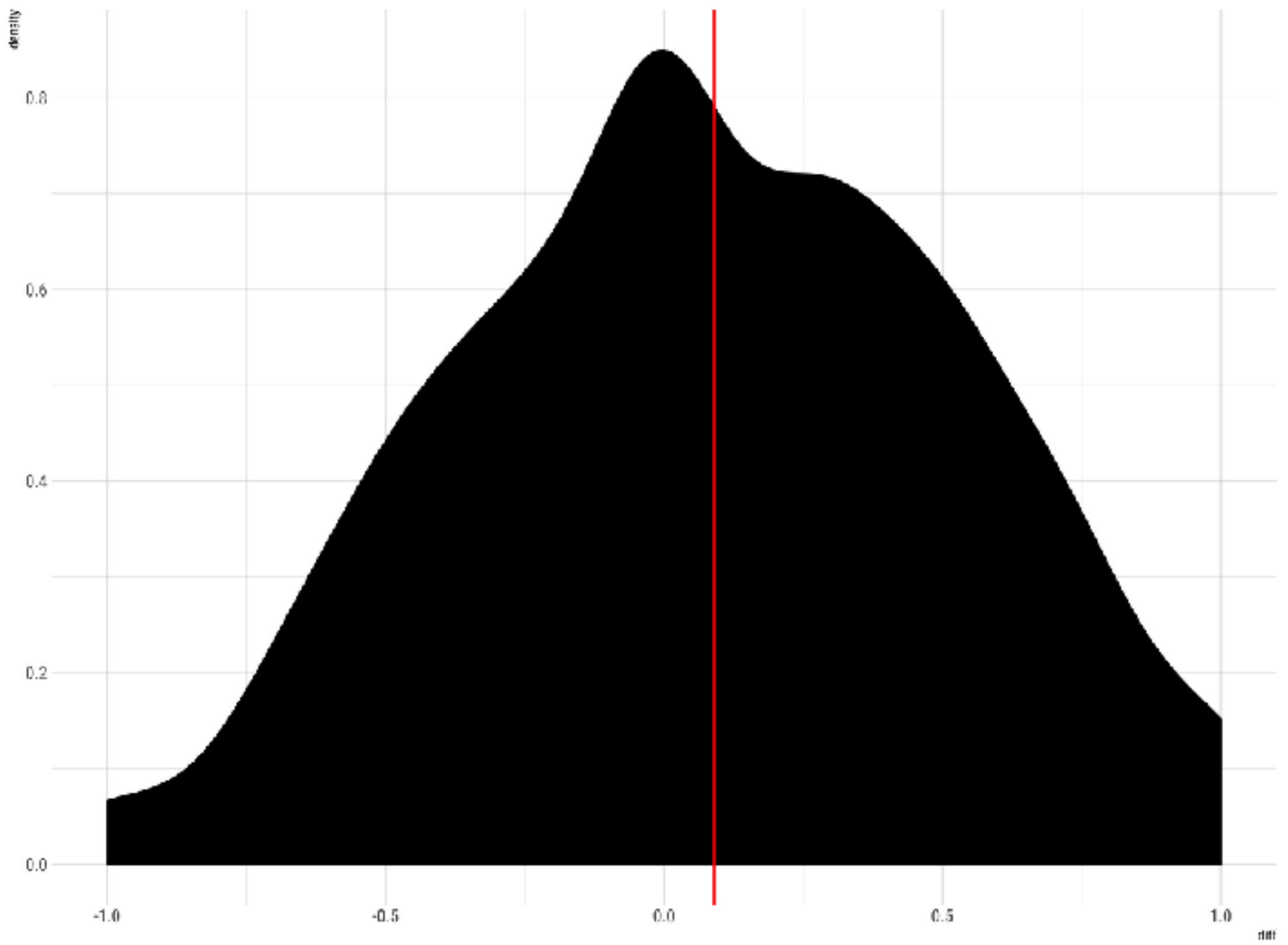
What’s happening?

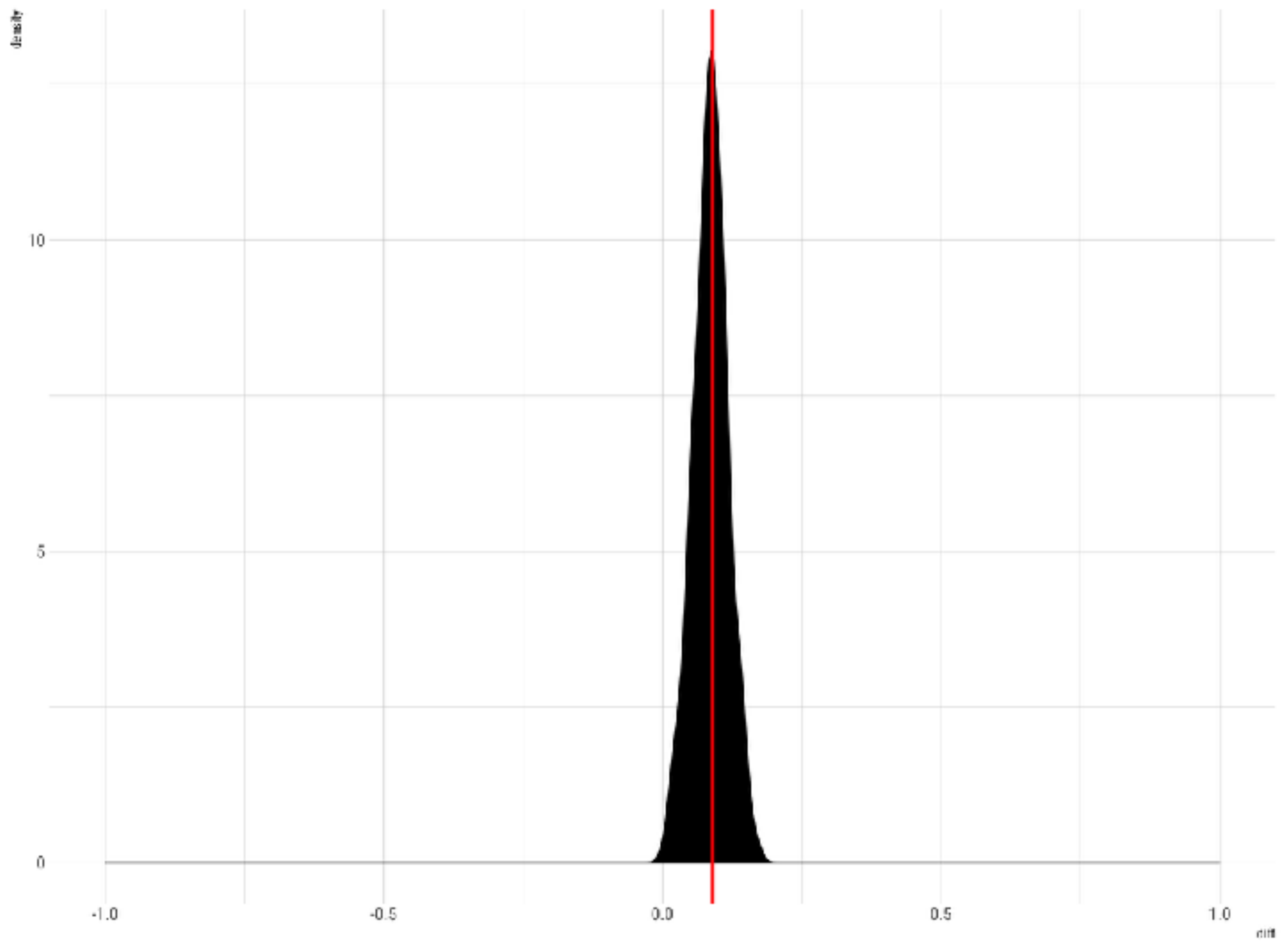
The distribution of **sample means** is roughly centered around the **population mean**

As samples get bigger, the sample means **vary less and get closer to population mean**

Detour on variation

How to measure this decrease in variation?





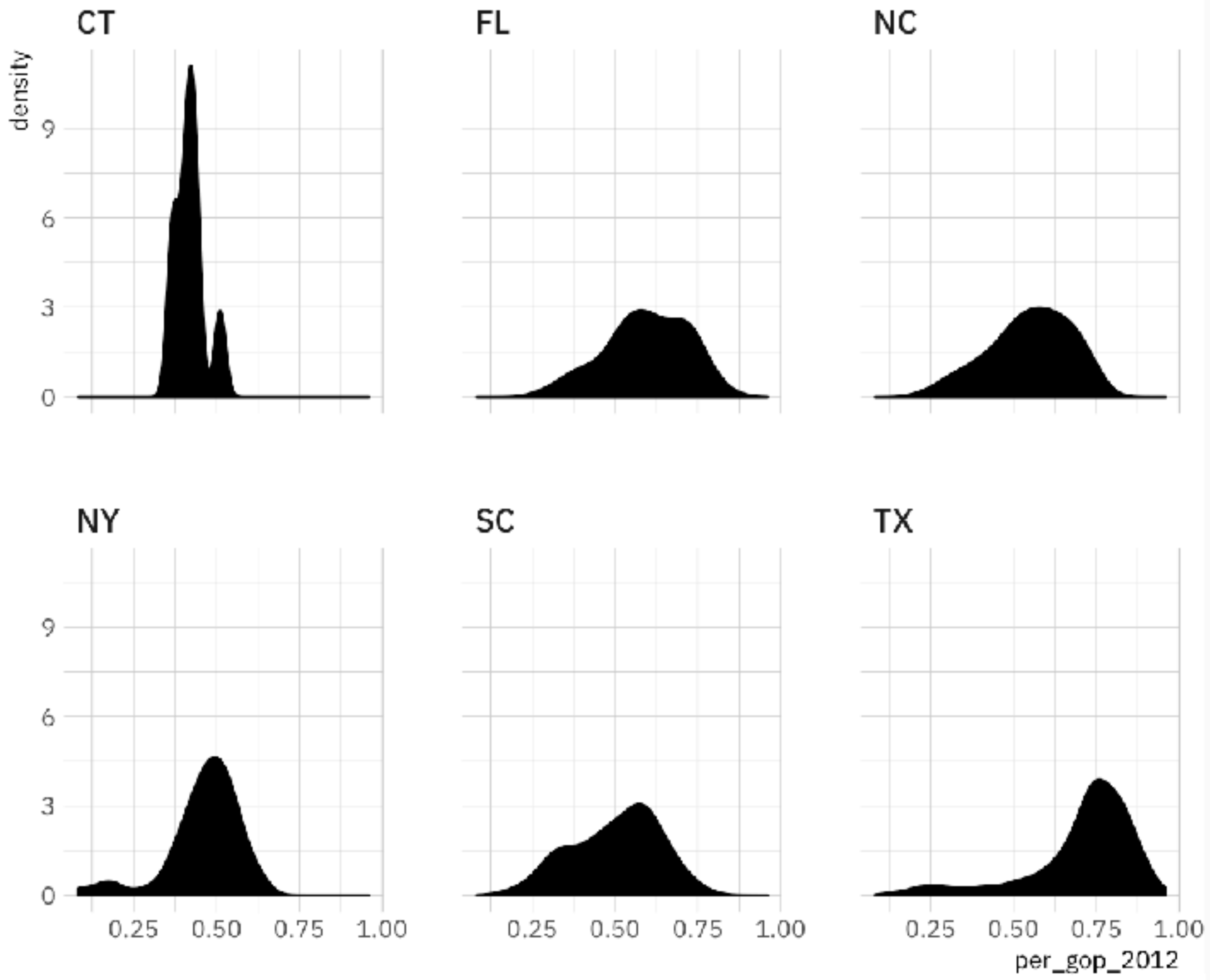
Standard deviation

Measures the spread of the data around its mean

$$Standard\,deviation = \sqrt{\frac{(x_1 - Mean)^2 + (x_2 - Mean)^2 + \cdots + (x_n - Mean)^2}{n-1}}$$

In the context of sampling we call the spread of sample means around the true mean the **standard error**

```r
# variation, standard deviation
county_data %>%
  group_by(state) %>%
  select(per_gop_2012) %>%
  summarise(mean = mean(per_gop_2012, na.rm = TRUE),
            sd = sd(per_gop_2012, na.rm = TRUE))
```




```
  state per_gop_2012
  <fct>        <dbl>
1 CT          0.0455
2 FL          0.123
3 NC          0.122
4 NY          0.109
5 SC          0.126
6 TX          0.159
```

What’s happening?

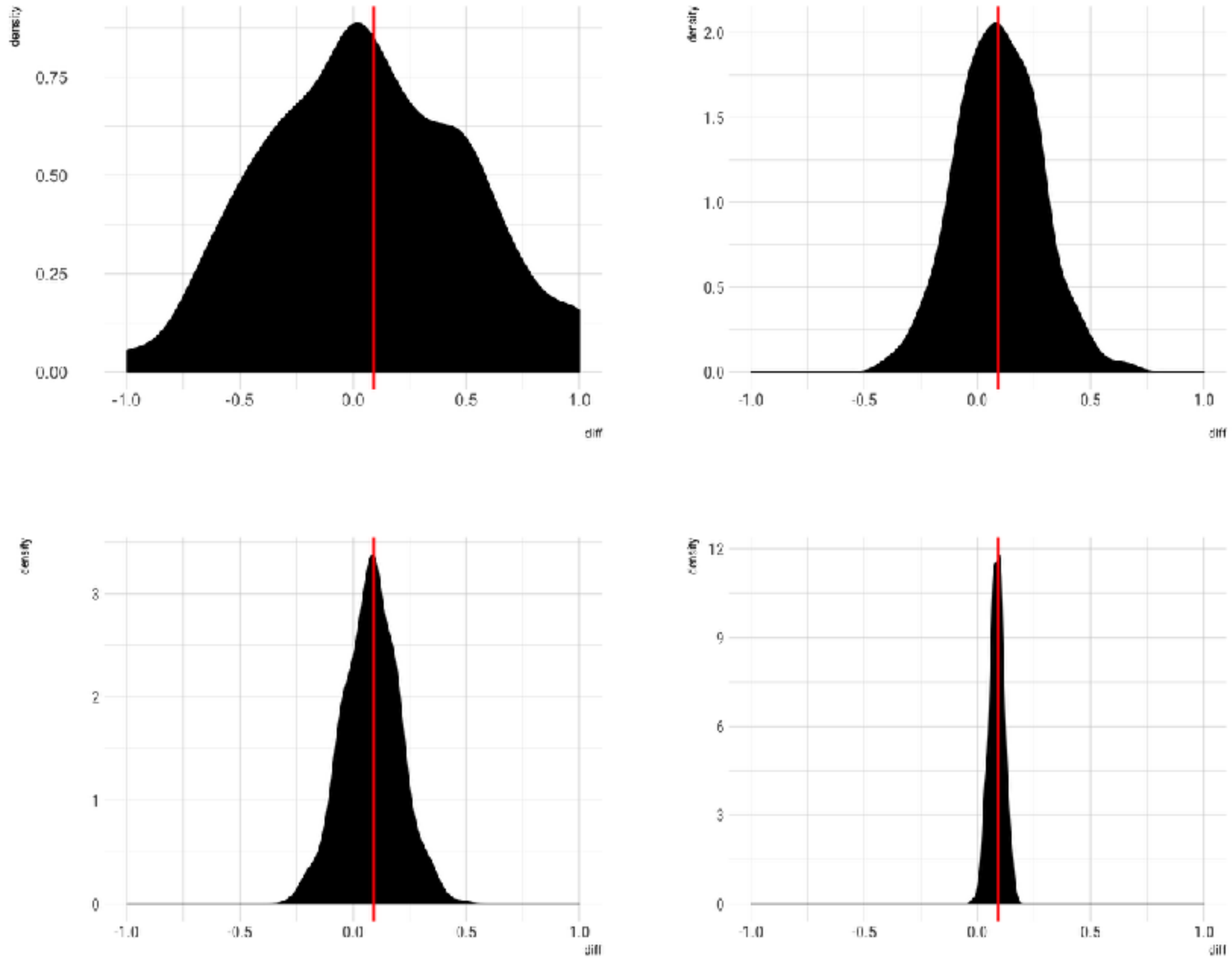
We don’t have all the data, we have a sample

As sample size increases, standard error goes down and our estimates become more **precise**

So long as our sample is **good and big**, we will have a pretty “good guess” of the true value

This is true not just of **averages** but also **coefficient estimates**, **medians**, and any other sample statistic

```
gss_sm %>%
  sample_n(100, replace = TRUE) %>%
  lm(obama ~ sex, data = .) %>%
  get_regression_table()
```



Good samples

There are good and bad samples in the world

A good sample is **representative** of the population and **unbiased**

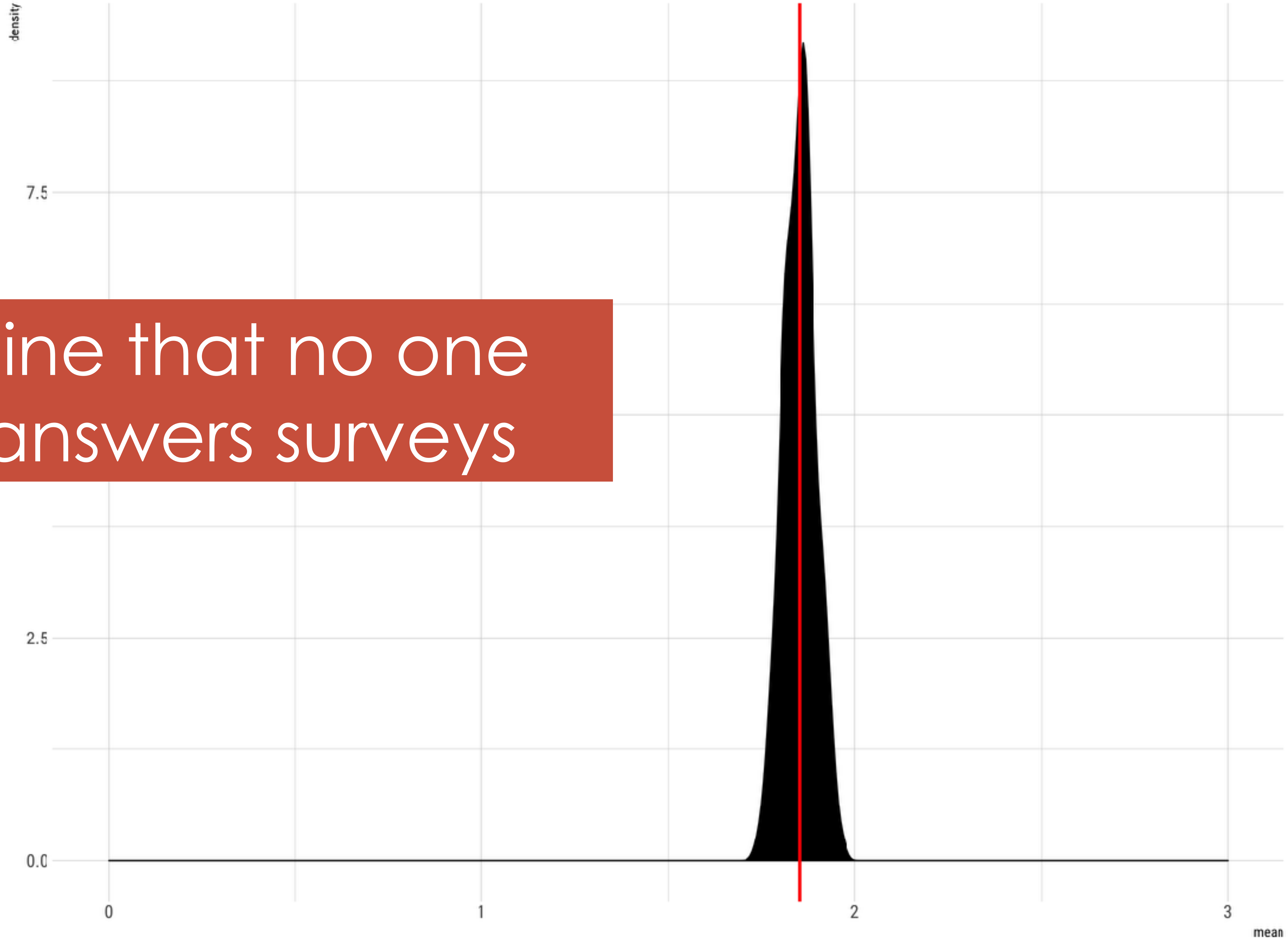
What does this mean?

Average number of kids

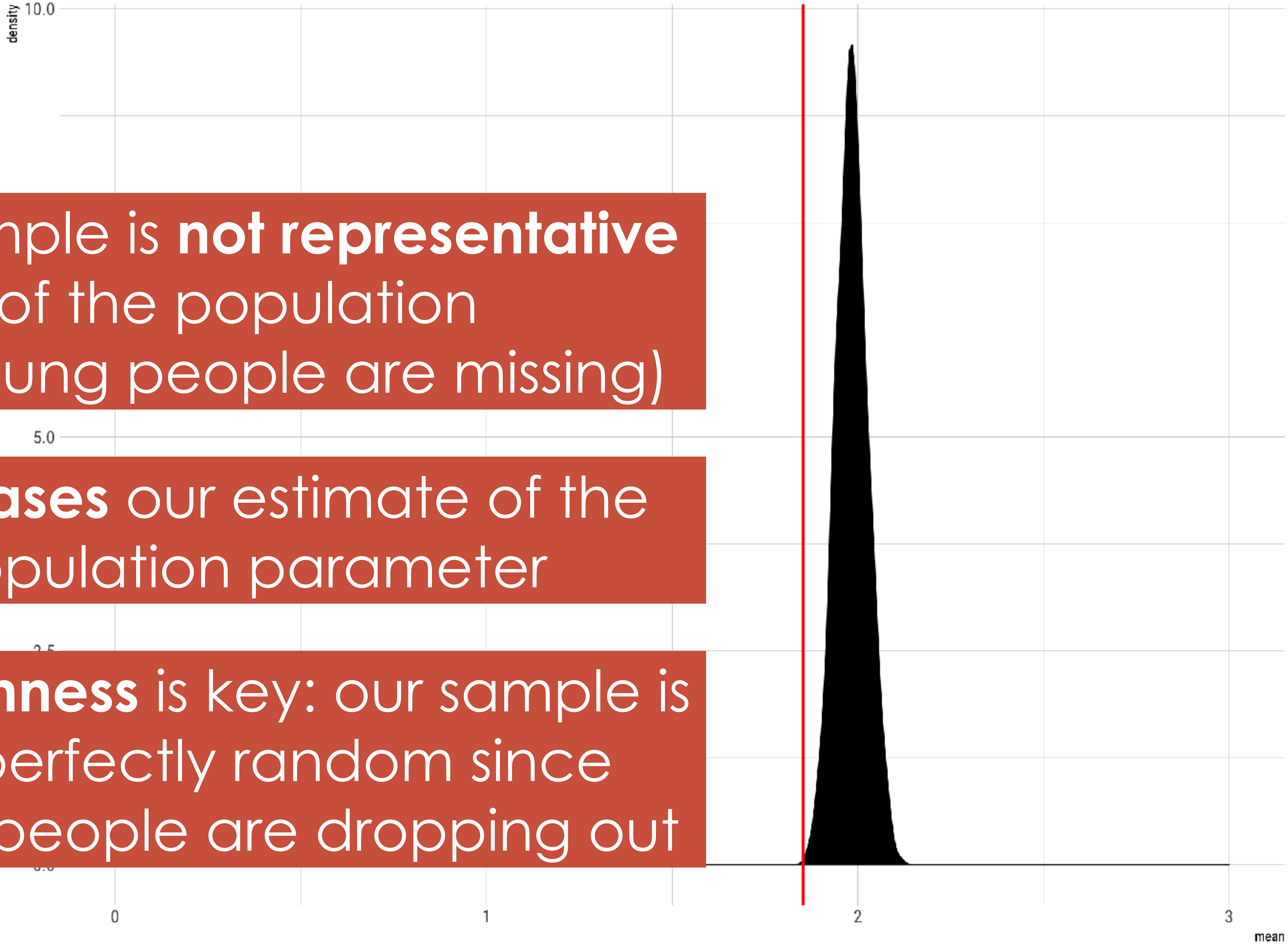
Find average number of kids in survey

Take samples with a big size and big n, and draw histogram

Average number of kids



Now imagine that no one under 25 answers surveys



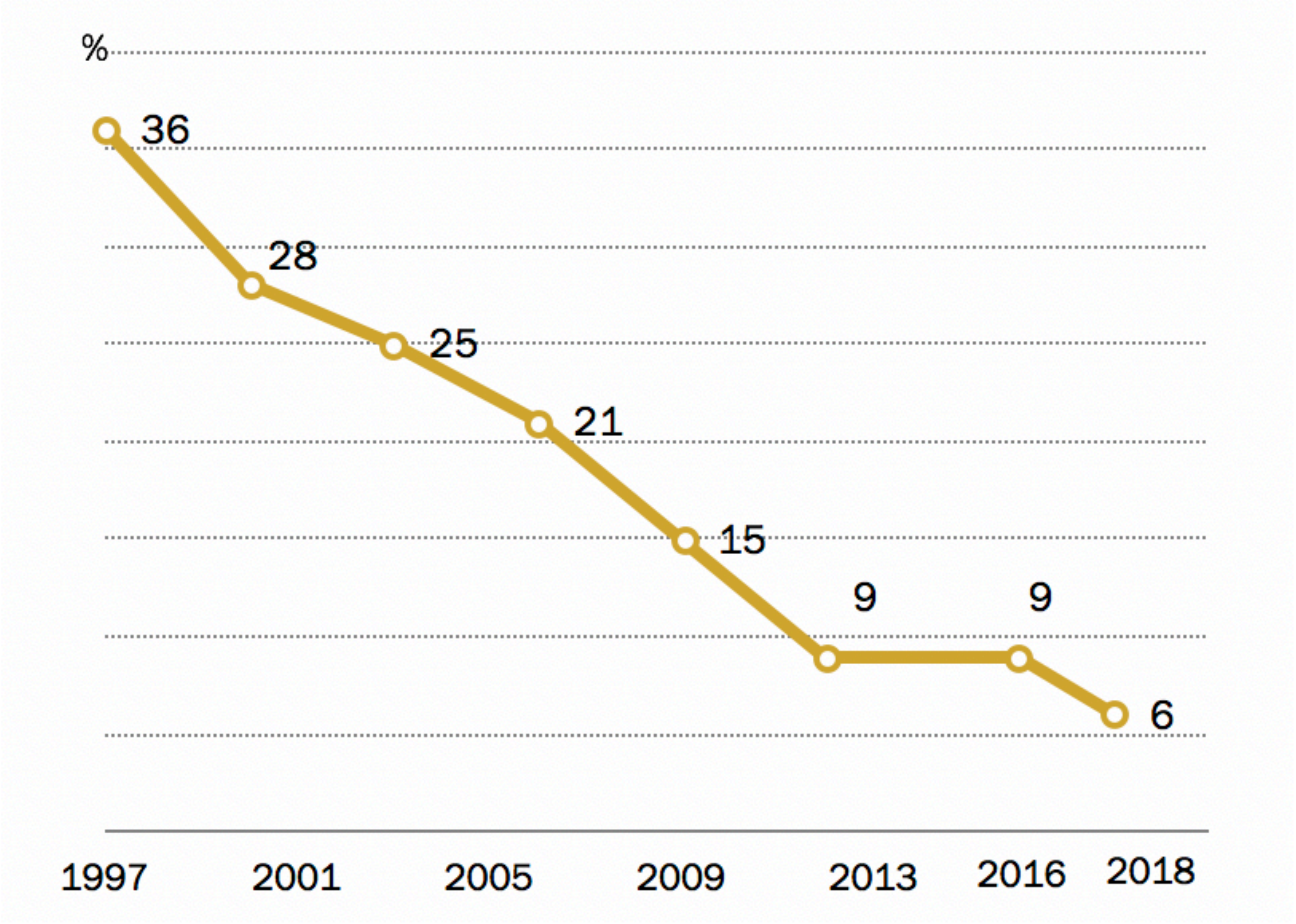
The sample is **not representative** of the population (the young people are missing)

This **biases** our estimate of the population parameter

Randomness is key: our sample is not perfectly random since young people are dropping out

After brief plateau, telephone survey response rates have fallen again

Response rate by year (%)



Note: Response rate is AAPOR RR3. Only landlines sampled 1997-2006. Rates are typical for surveys conducted in each year.
Source: Pew Research Center telephone surveys conducted 1997-2018.

PEW RESEARCH CENTER

Key takeaways

As **sample size** increases, **sample means** approach **population mean**

This makes sense: more data, more certain of the pattern

All **depends** on whether we have a **representative** sample; no amount of data in the world will correct for sampling bias

This is weird

We've shown that if we take 1,000 samples most of the averages of those samples will be close to **population mean**

But in real life we only ever have **one** sample (i.e., the one Iowa poll)

How do we account for the fact that each sample will give a (slightly) different mean?

Solution

We can get a sense for this variability by **resampling from our sample**

Mimics the way each sample mean is different from pop. mean: each bootstrapped sample mean will be "off" as well

This is reasonable because the sample is the **only information we have about the population**

And, most random samples will look like the population, so likely ours does too

Housekeeping

PS9 due this Tuesday… but I won't finish it till tomorrow

No class Tuesday

PS10: due this Tuesday, or Tuesday/ Thursday after Thanksgiving?

Variability

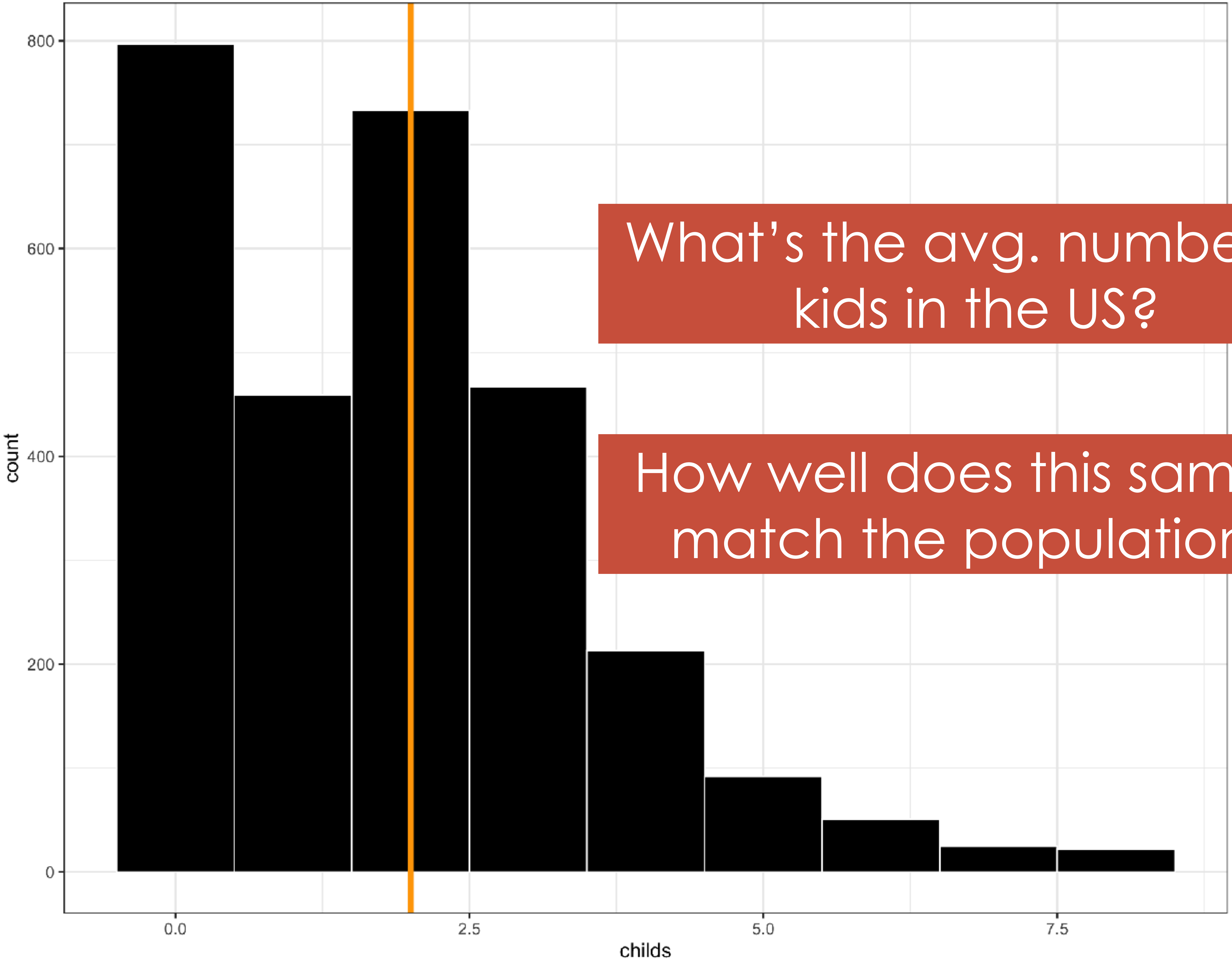
How much you expect the mean to vary from sample to sample

With infinite resources, we could take thousands of samples (or a census), and get the exact population parameter and its variability

This is where **bootstrapping** comes in

Use the sample you have to make many new samples

You can use the distribution of these bootstrapped samples to quantify your uncertainty in the estimate



What's the avg. number of kids in the US?

How well does this sample match the population?

How to bootstrap

Take a bootstrap sample

Sample with replacement; same size as original

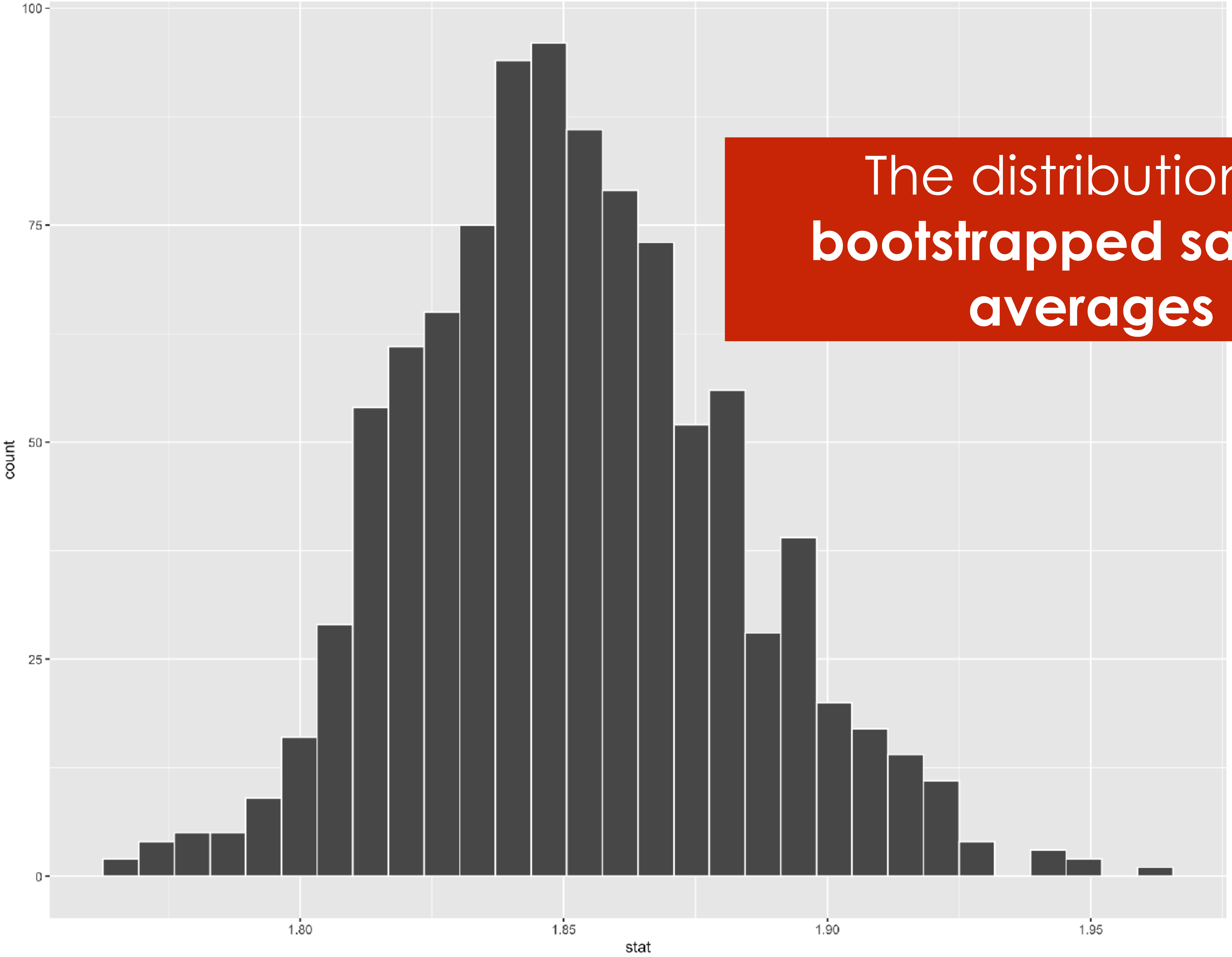
Calculate bootstrap estimate

Mean, median, etc.

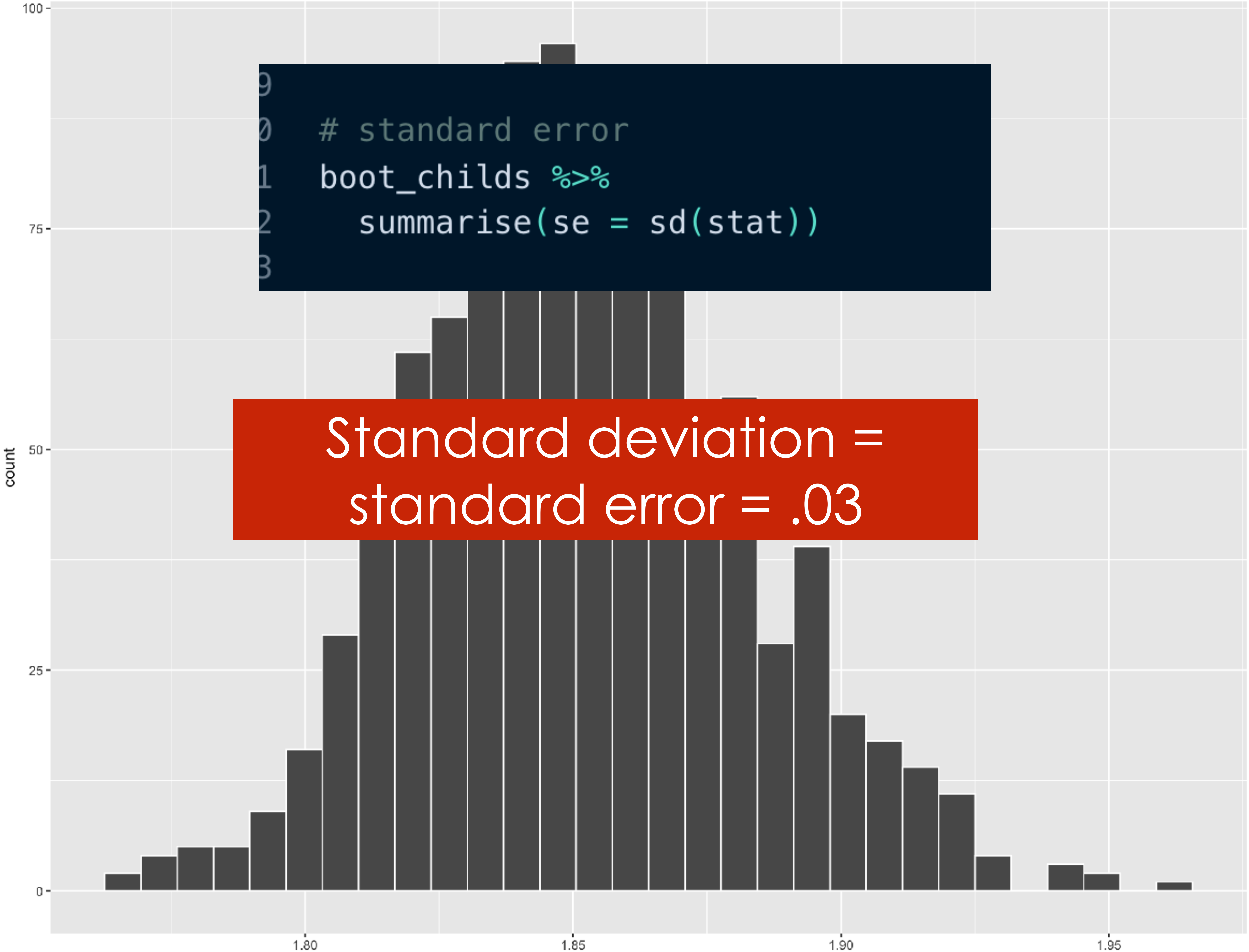
Repeat N times

Infer package

Follow along in R



The distribution of **bootstrapped sample averages**




```
# standard error
boot_childs %>%
  summarise(se = sd(stat))
```

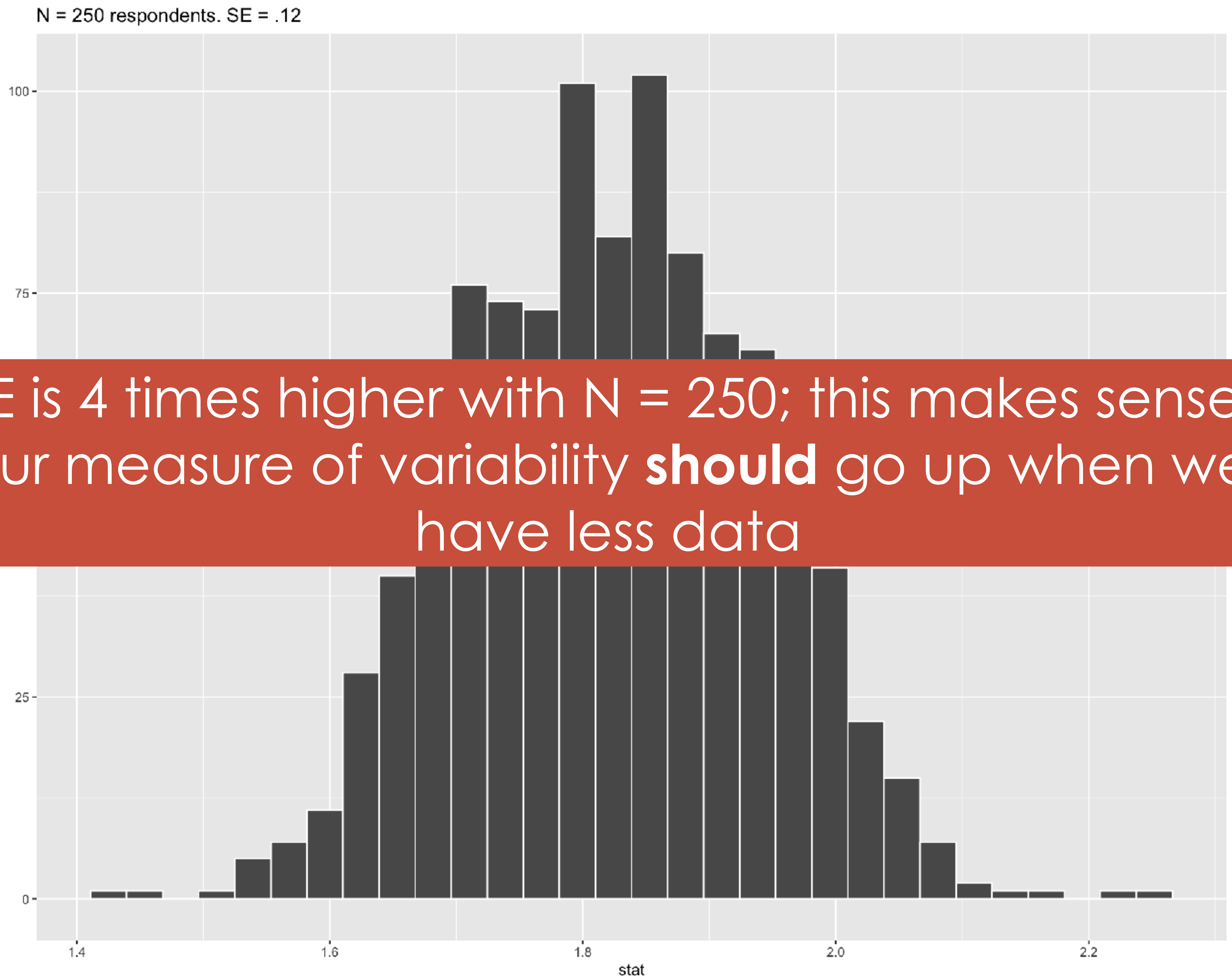
Standard deviation =
standard error = .03

What's happening?

Taking many re-samples from our sample mimics taking many samples from a population

Sampling **with replacement** is what allows us to do this; we can get many different looking re-samples from our sample

If our sample is **representative** of population, it will have an approximate shape and spread, which we can use to quantify **variability**



SE is 4 times higher with N = 250; this makes sense: our measure of variability **should** go up when we have less data

Inferring other statistics

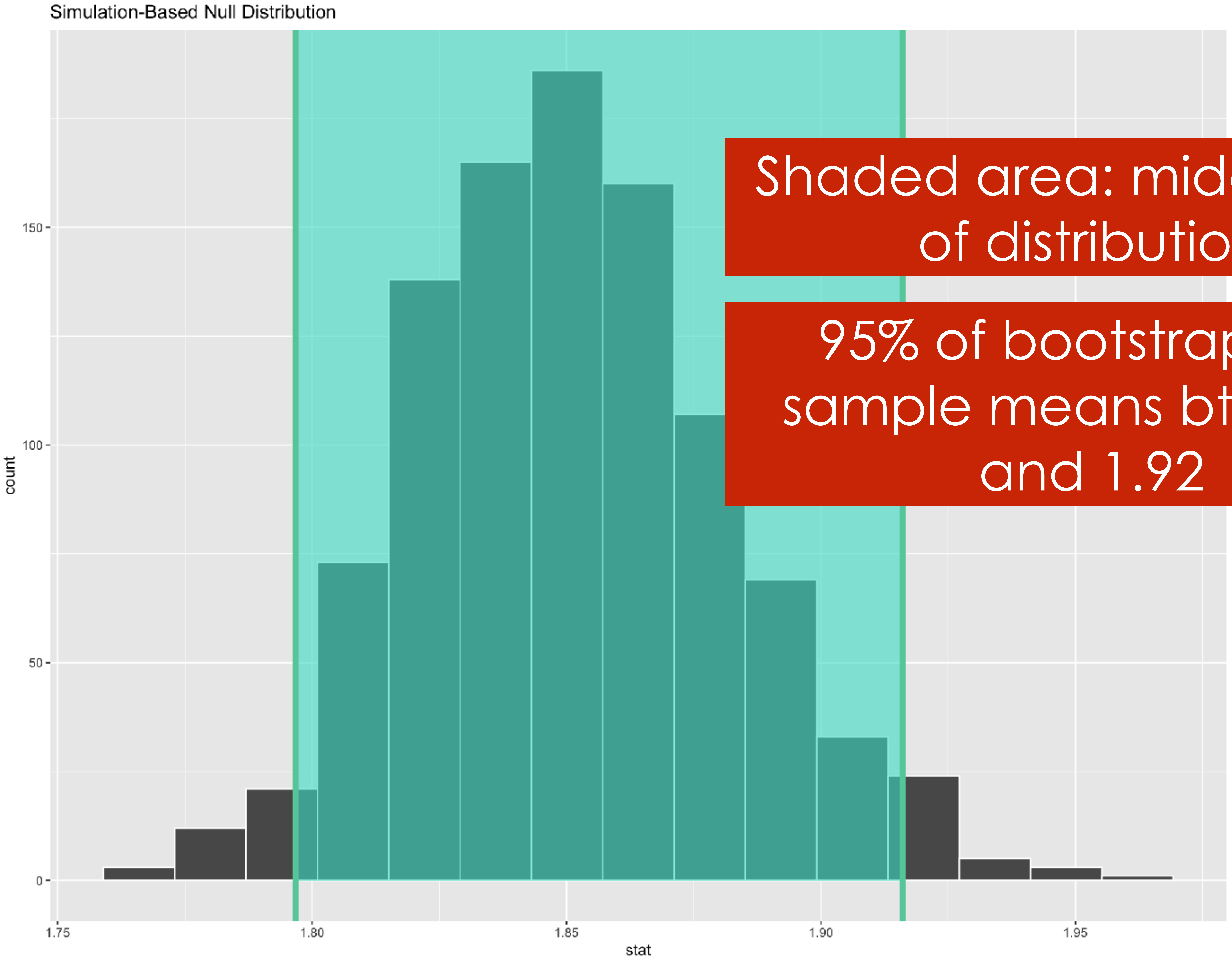
We can use infer with other things, like proportions, medians, means, coefficient estimates

Follow along in R

Confidence intervals

A range of highly plausible values for our parameter

95% confidence interval is the middle 95% of the the distribution

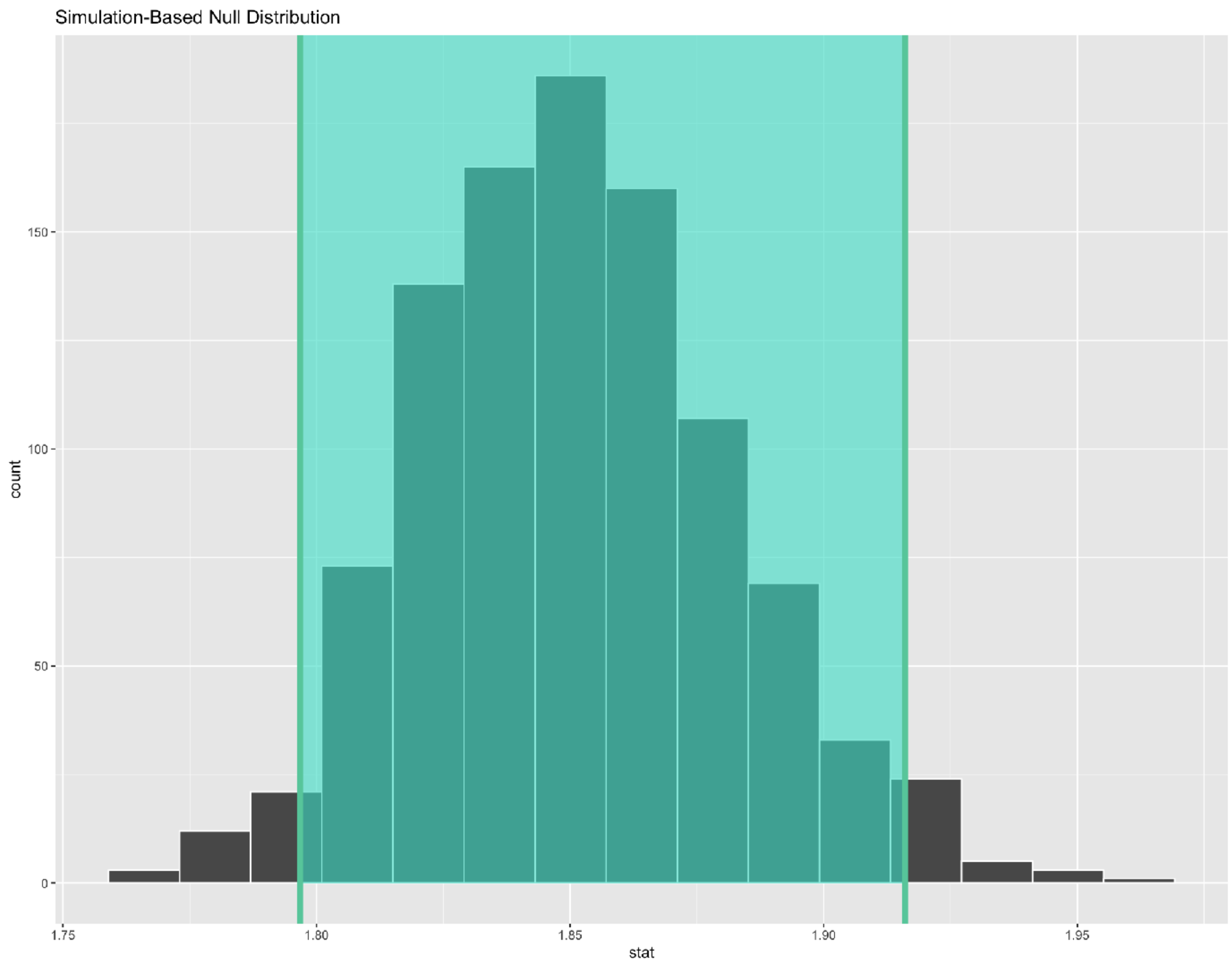


Shaded area: middle 95% of distribution

95% of bootstrapped sample means btwn 1.8 and 1.92

Confidence interval

The **confidence interval** gives us a sense of the variability in our estimate



But how do we interpret it? What does it mean?

How (not) to interpret confidence intervals - THE SKEPTICAL ...
www.timvanderzee.com › not-interpret-confidence-intervals ▾
Oct 19, 2017 - **Confidence intervals** give a range of values which will "in the long run" .... overestimate the amount of error and thus **incorrectly** assume that two ...

(Mis)Interpreting Confidence Intervals | Sherman's Head
rynesherman.com › blog › misinterpreting-confidence-intervals ▾
Nov 17, 2014 - In a recent paper Hoekstra, Morey, Rouder, & Wagenmakers argued that **confidence intervals** are just as prone to **misinterpretation** as ...

Misconceptions about Confidence Intervals - Statistics Solutions
https://www.statisticssolutions.com › misconceptions-about-confidence-inte... ▾
Jan 9, 2017 - Some of the most common misconceptions about **confidence intervals** are: "There is a 95% chance that the true population mean falls within the **confidence interval**." (**FALSE**) "The mean will fall within the **confidence interval** 95% of the time."

The Correct Interpretation of Confidence Intervals - SAGE ...
https://journals.sagepub.com › doi › pdf
by SH Tan - 2010 - Cited by 9 - Related articles
Confidence intervals (CI) are a key output of many ... in the **interpretation** of estimates of parameters. .... is an **incorrect interpretation** of 95% CI because the.

Why is the "wrong" interpretation of confidence intervals still ...
https://math.stackexchange.com › questions › why-is-the-wrong-interpretat... ▾
1 answer
May 6, 2015 - Certainly the probability that the **interval** you will get contains the population mean is 0.95, but the conditional probability given the numbers that ...

Confidence intervals are **frequently** misinterpreted

The probability that the average number of children in the US is between 1.8 and 1.92 is 95%

95% of the time, the average number of children in our sample is between 1.8 and 1.92

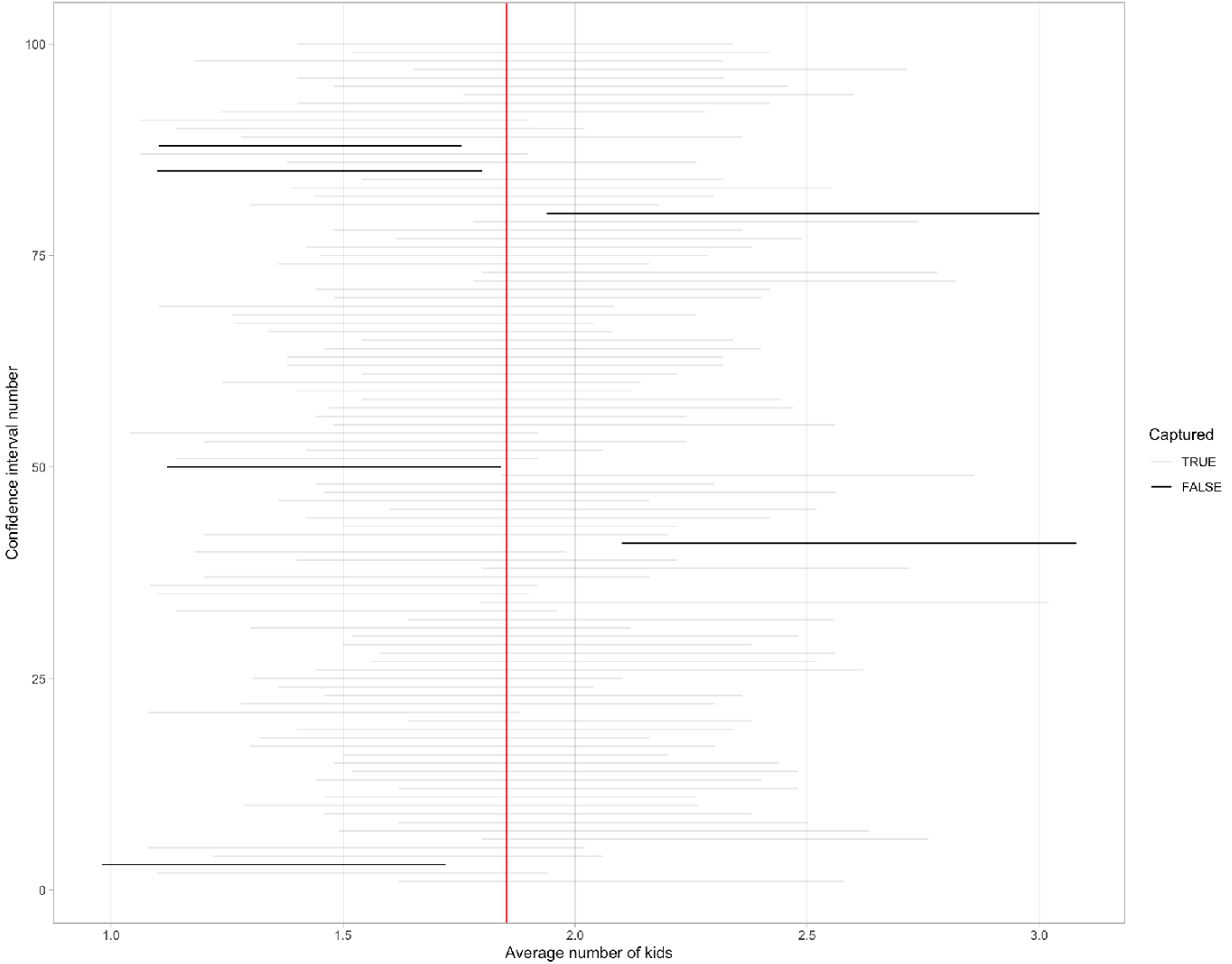
95% of all respondents in our sample have between 1.8 and 1.92 kids

We are 95% “confident” that the average number of kids is between 1.8 and 1.92

“95% confident”

Confidence intervals are a “net”:

Let’s take 100 samples, each of size = 50 and calculate 95% confidence intervals for each



Confidence interval number
100
75
50
25
0
1.0
1.5
2.0
2.5
3.0
Average number of kids
Captured
TRUE
FALSE

"for every 100 95% confidence intervals, we expect that 95 of them will capture population mean and that 5 of them won't."

The "95% confident" part pertains to the **procedure** itself and not to the confidence interval we calculated

Loosely though: confidence intervals are our "best guess" of the mean number of kids in the US

Very common mistake

"There is a 95% chance that the true avg. number of kids in the US is between 1.8 and 1.92"

Wrong: the true avg. number is either within that interval or it isn't, there's no chance here

Only legal interpretation

"We are 95% confident that the mean number of kids in the US falls between 1.8 and 1.92"

If this sounds insane to you

It also does to many other people

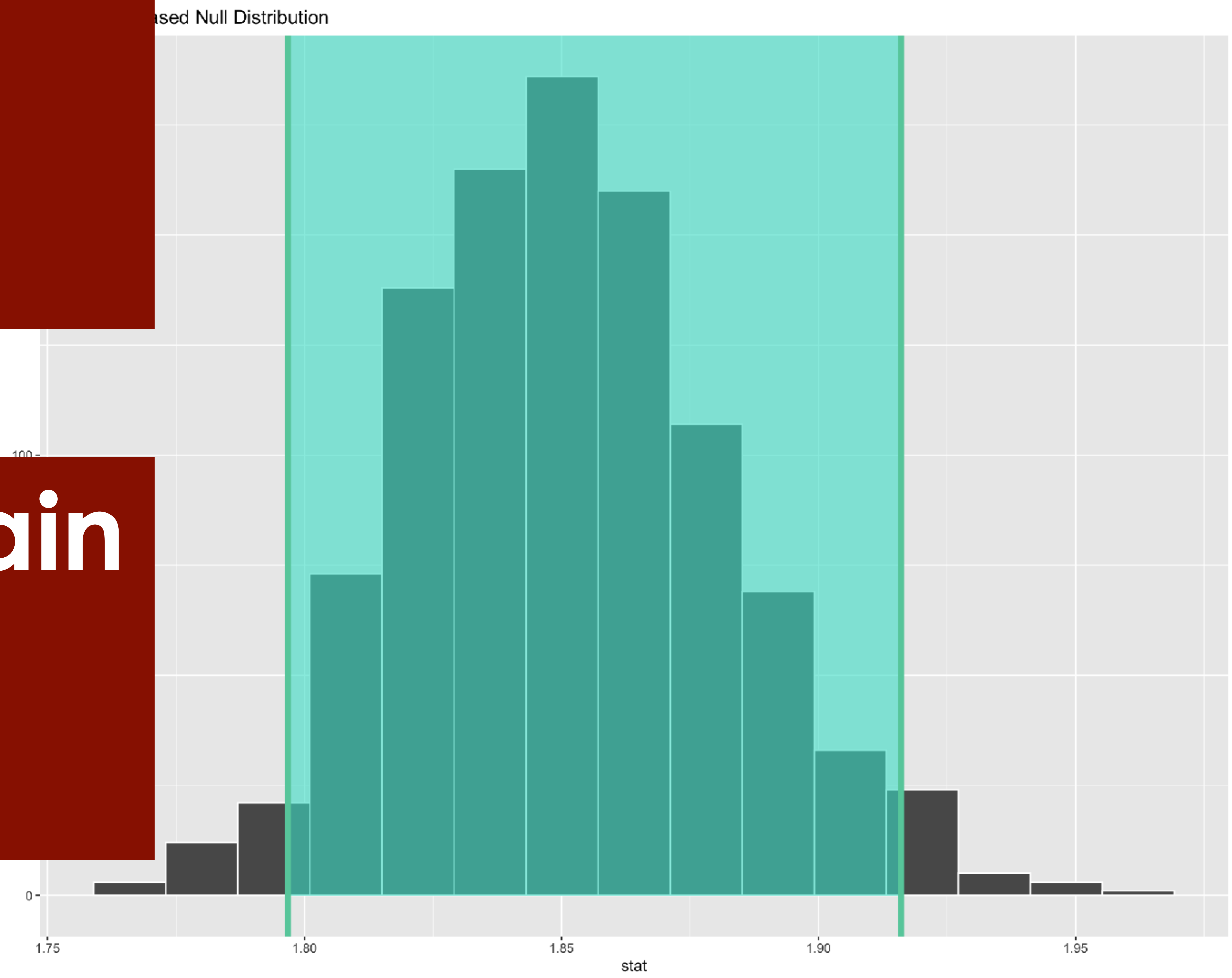
A different approach to statistics known as **Bayesian statistics** devises a more intuitive answer

Precision vs. accuracy

95% is the standard but we could pick other percentiles

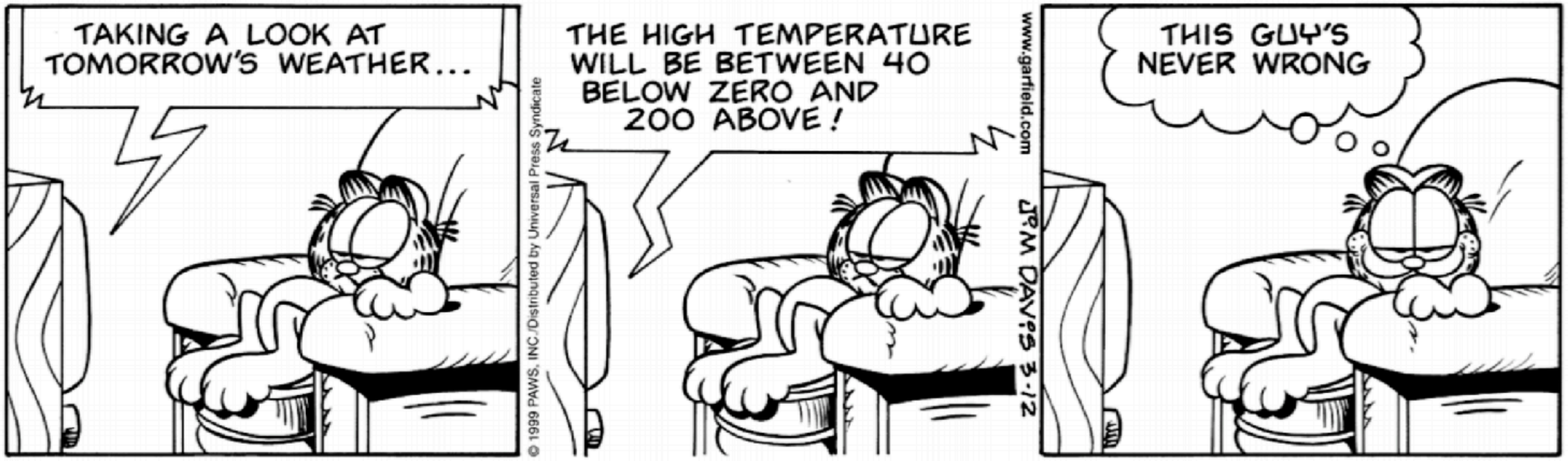
The **bigger** the percent, the **more certain** we are, and the **wider** the confidence interval

The **smaller** the percent, the **less certain** we are, and the **narrower** the confidence interval



```
> boot_childs %>%
+   get_ci(level = 0.95, type = "percentile")
# A tibble: 1 x 2
  `2.5%` `97.5%`
   <dbl>   <dbl>
1   1.80    1.92
> boot_childs %>%
+   get_ci(level = 0.99, type = "percentile")
# A tibble: 1 x 2
  `0.5%` `99.5%`
   <dbl>   <dbl>
1   1.78    1.94
> boot_childs %>%
+   get_ci(level = 0.50, type = "percentile")
# A tibble: 1 x 2
  `25%` `75%`
  <dbl> <dbl>
1  1.83  1.87
```

Always a trade-off between precision and accuracy



But 95% is standard

RECAP

Sample statistic will always be "off" from population

If sample is good, it will be close

Report estimate with confidence interval

Width of interval depends on how variable sample statistics would be from different samples

Use bootstrap to get confidence interval

Bootstrap lets us get at variability

HYPOTHESIS TESTING

No, Having Bigger Biceps Does Not Make You More Conservative

Walt Hickey May 17, 2013, 12:34 PM

A study by a group of researchers about biceps and politics that has made waves over the past couple of days following its publication in *Psychological Science* has been widely misinterpreted by several news outlets.

Getty

Specifically, men with greater upper-body strength should more strongly favor redistribution if they are poor but oppose it if they are wealthy. [...] First, for males of low [socioeconomic status], physical strength should positively correlate with support for redistribution; second, for males of high [socioeconomic status], physical strength should negatively correlate with support for redistribution

Science is like the courts

Innocent until proven guilty

Accuser (plaintiff) has burden of proof

Judge/jury decide guilt on amount of evidence

Innocence is never proven! Accuser tries (and fails) to reject innocence

Science is like the courts

Preponderance of evidence

Clear and convincing evidence

“Beyond a reasonable doubt”

Why do we have these different levels?

We fear punishing someone who is innocent (Type 1 error) more than letting someone go who is guilty (type 2 error)

Science is like the courts

Presume no effect of X on Y (the **null hypothesis**)

You have burden of proving **there is an effect**

Decide if there is an effect based on **amount of evidence**

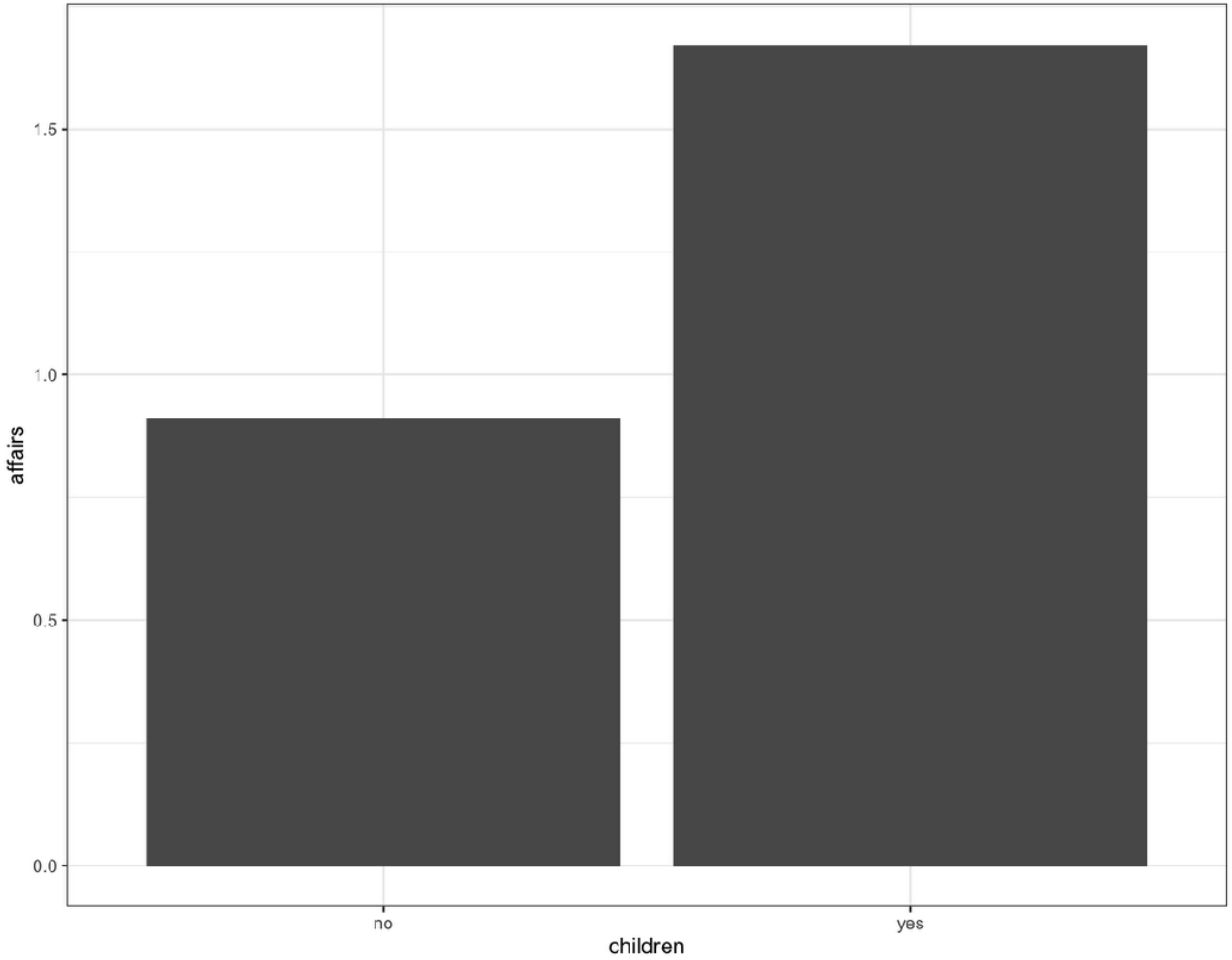
Never prove that null is true; we try and fail to reject the null

Let's say we think people with children are more likely to have affairs than people without children

Example

Null hypothesis: people with and without children have affairs at same rate

Alternative (our) hypothesis: people with children have affairs at higher rates than child-less people



The problem

People with children have on average .76 more affairs

That's an 80% increase; substantial

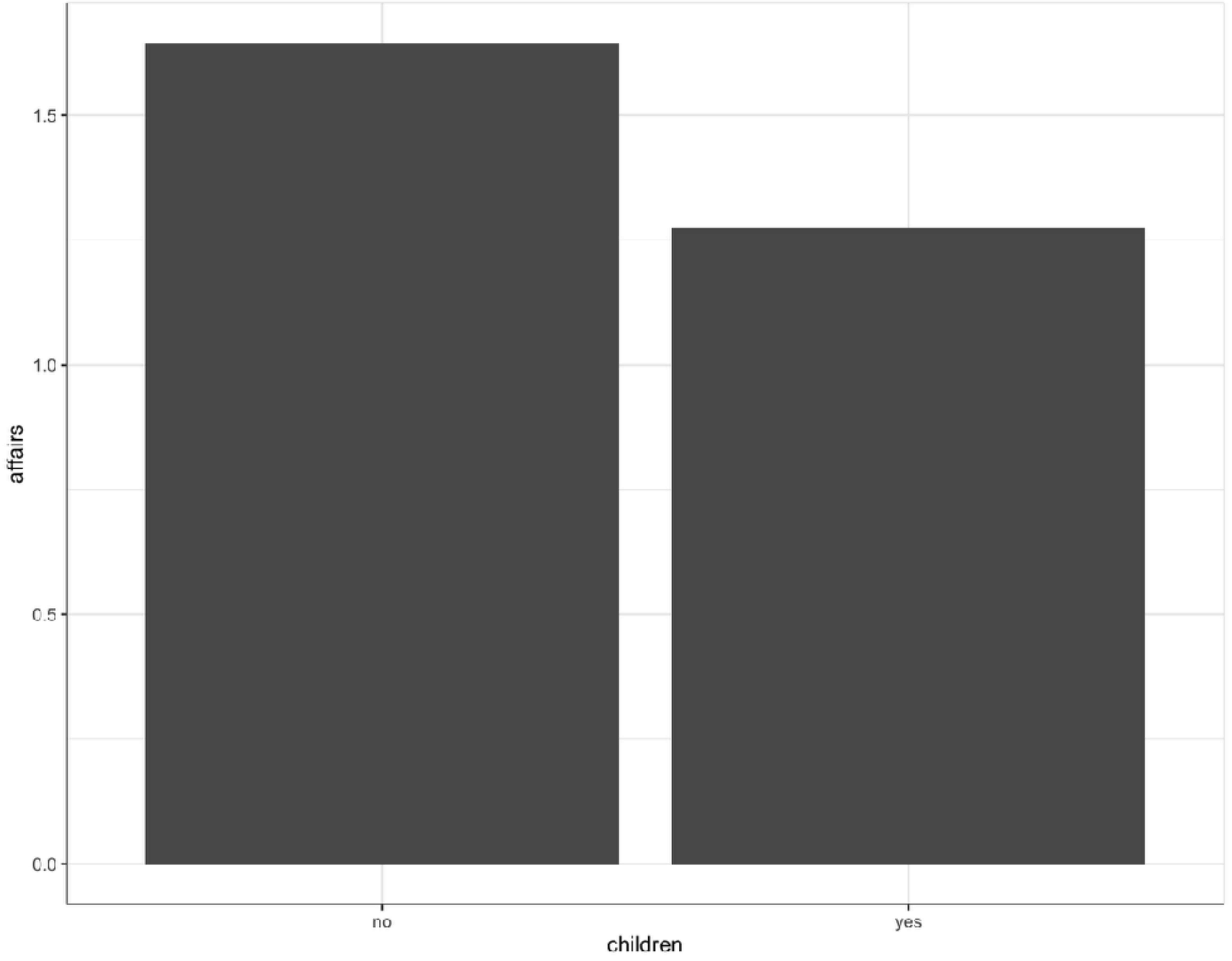
What if we just got a weird sample though?

Could we have gotten this estimate **by chance**?

Imagine **randomly shuffling** whether respondent has kids

```
> Affairs %>%
+   select(affairs, children) %>%
+   mutate(shuffled_kids = sample(c("yes", "no"), size = n(), replace = TRUE))
   affairs children shuffled_kids
1        0       no            no
2        0       no           yes
3        0      yes           yes
4        0      yes           yes
5        0       no            no
6        0       no           yes
7        0       no            no
8        0      yes            no
9        0      yes           yes
10       0       no           yes
11       0      yes            no
```

Now: those without kids have .37 *more* affairs!

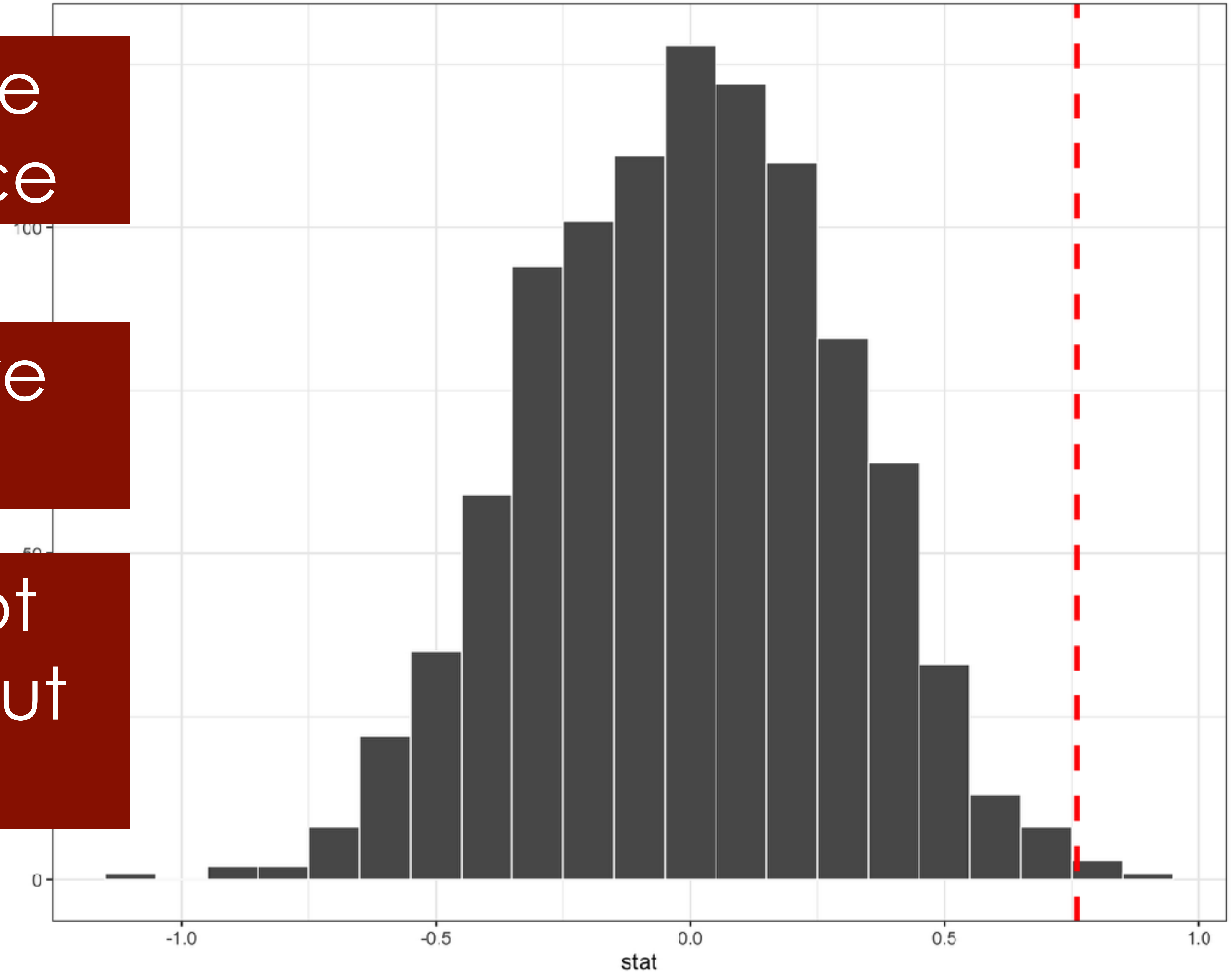


Repeat 1000x

Null distribution: what we might observe by chance

Sample statistic: what we actually observed

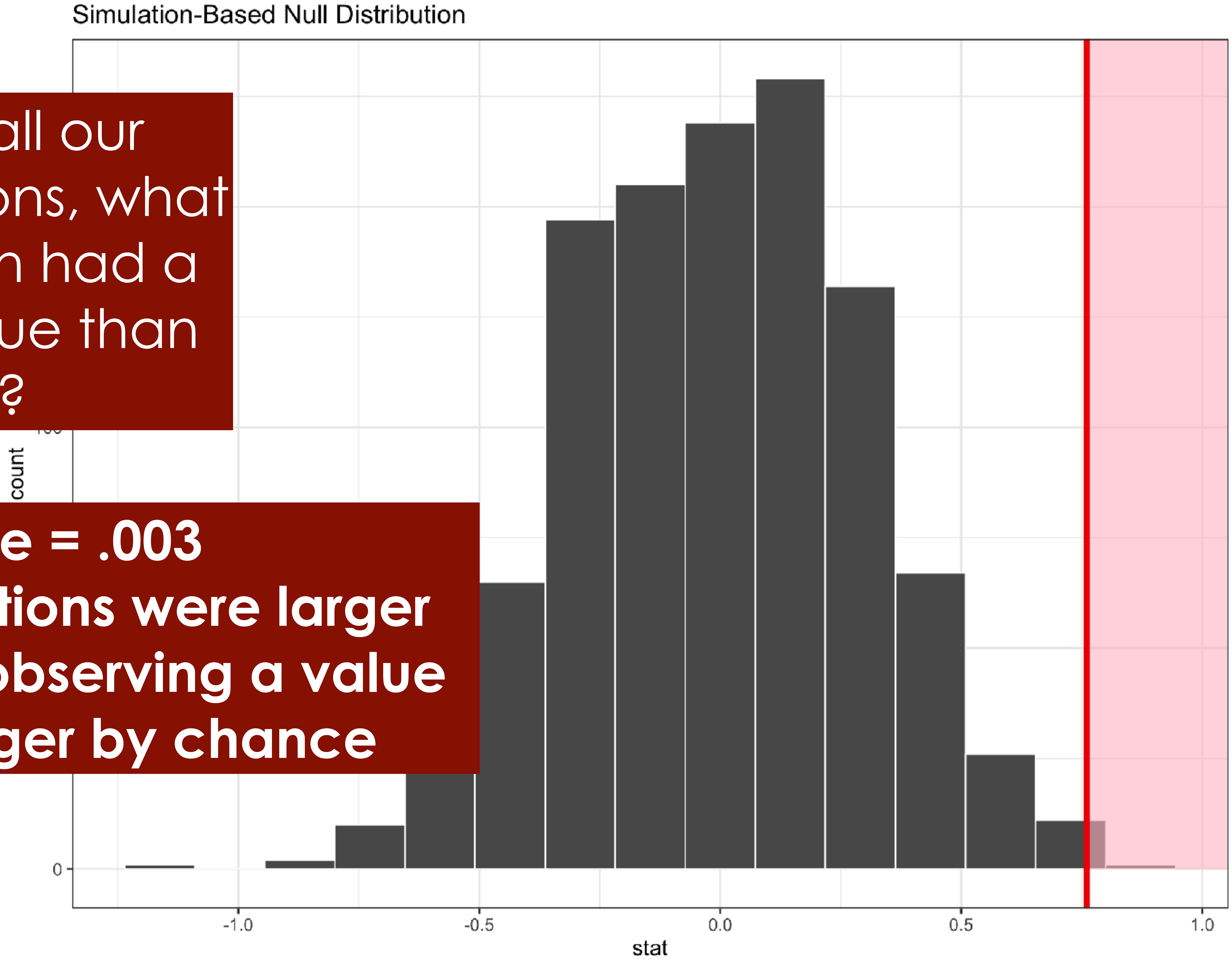
What we observed is not impossible by chance, but very unlikely



What do we mean by “very unlikely”?

P-values

The probability of observing an effect at least that large when no effect exists



Out of all our permutations, what proportion had a larger value than .76?

p-value = .003
.3% of permutations were larger
.3% chance of observing a value this big or larger by chance

OK, but how unlikely does something have to be before we reject the null ("declare guilty")?

Statistical significance

A threshold for deciding if enough evidence to safely reject the null

Typically: p-value needs to be less than .05