A Short History of Experiment Design (SOC 412)

Week 3 Lecture 6

Sherrerd Hall 306

J. Nathan Matias

@natematias

civilservant.io

jmatias@princeton.edu

Department of Psychology

PRINCETON UNIVERSITY

CITP

mit media lab

What we will cover today

History of Randomized Trials, as described by Ann Oakley

(Psychology, Sociology, Education, Criminal Justice, Welfare, the Economy)

Teams and the Final Project

Questions to Discuss

When the stakes are high (substantial risks/benefits) should we be more likely to experiment or less likely?

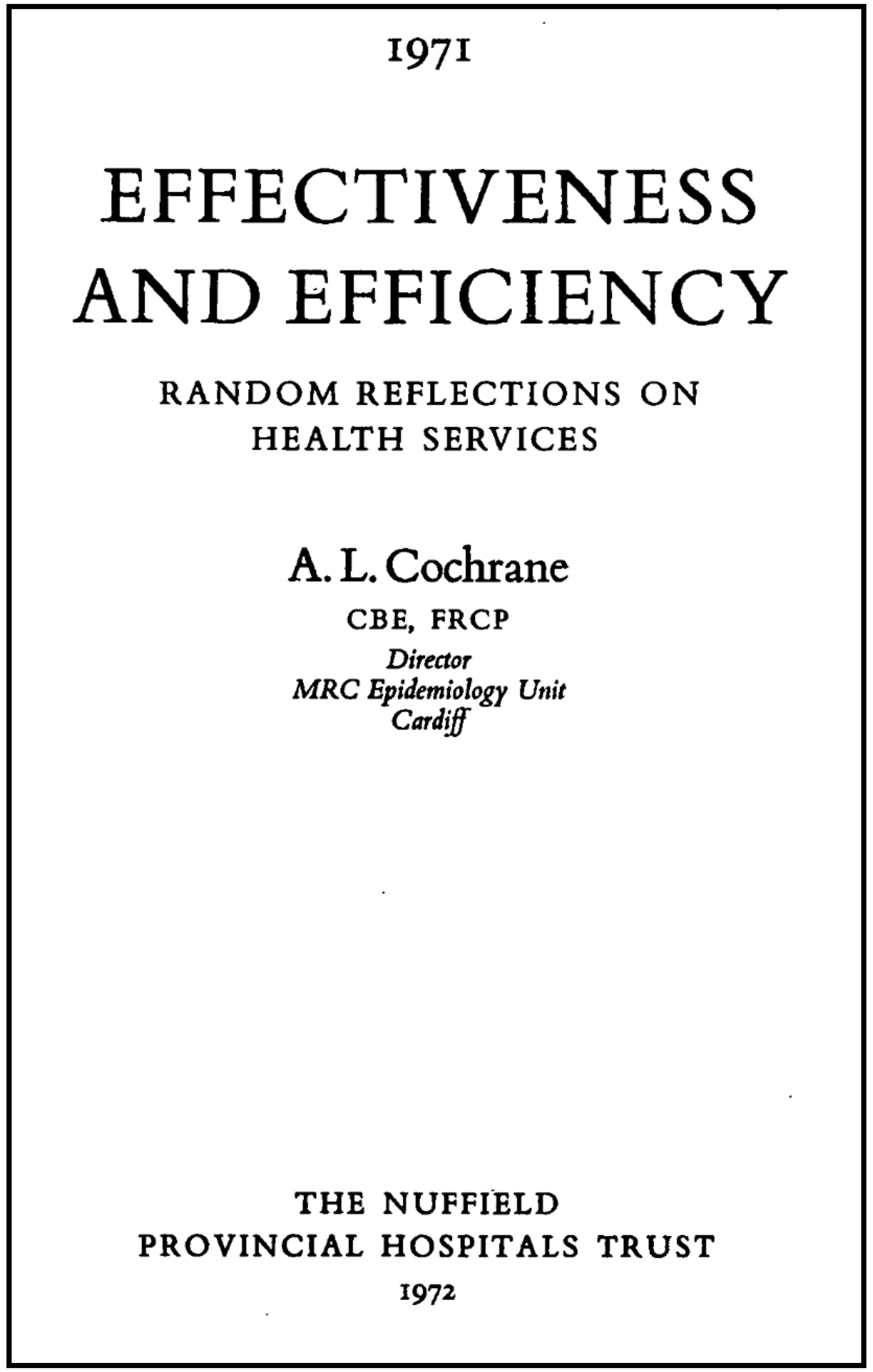
How much do you think it matters who is asking the questions (in science) (in policy) (in business)? What are the pros and cons of having non-experts set research goals? (assuming non-experts are stakeholders)

How can politics derail or enable effective evaluation?

Should randomized trials be a gold standard?

Other Streams of History

1962: Clinical Testing of Drugs Required



1971

EFFECTIVENESS AND EFFICIENCY

RANDOM REFLECTIONS ON HEALTH SERVICES

A. L. Cochrane

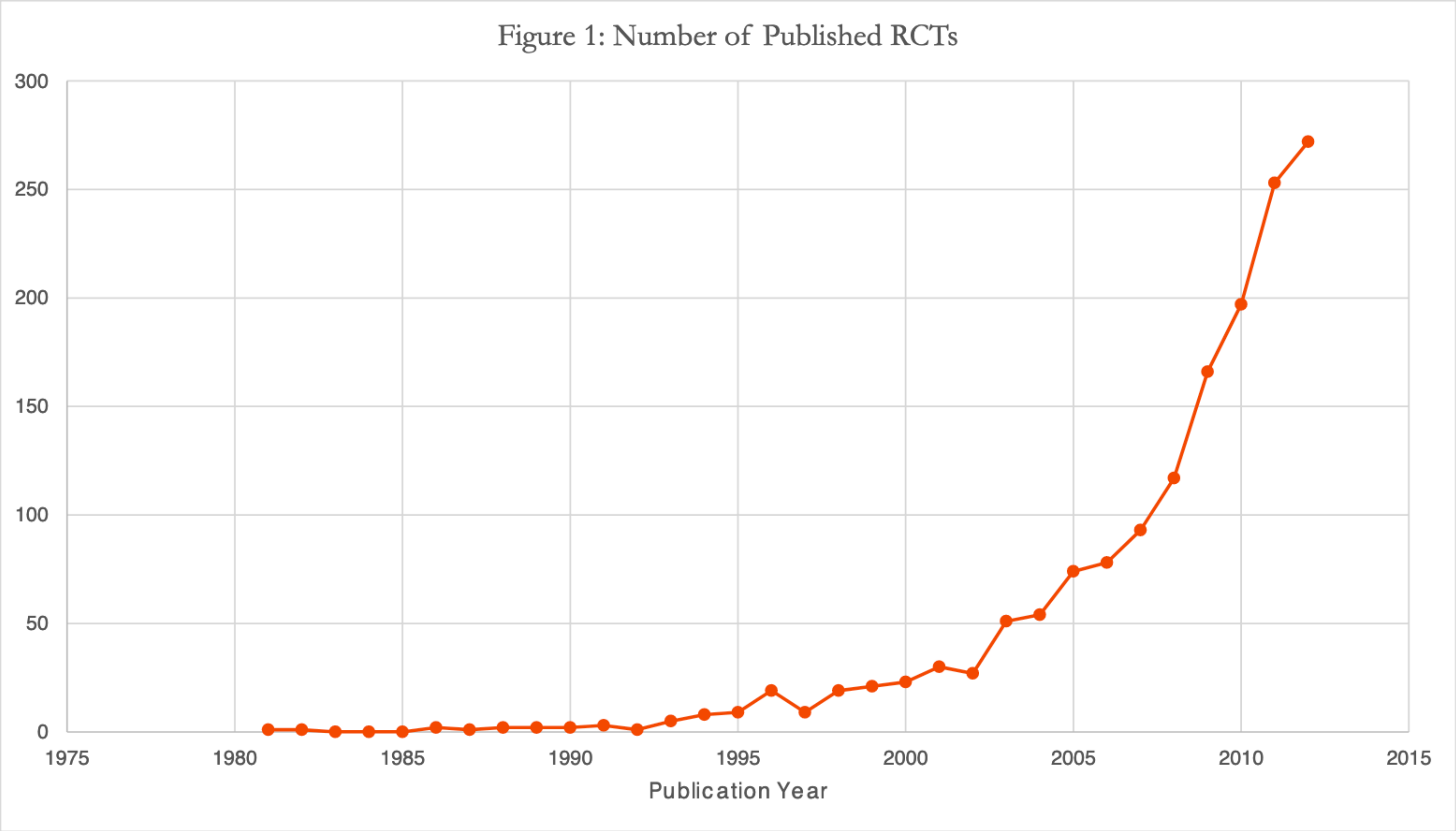
CBE, FRCP

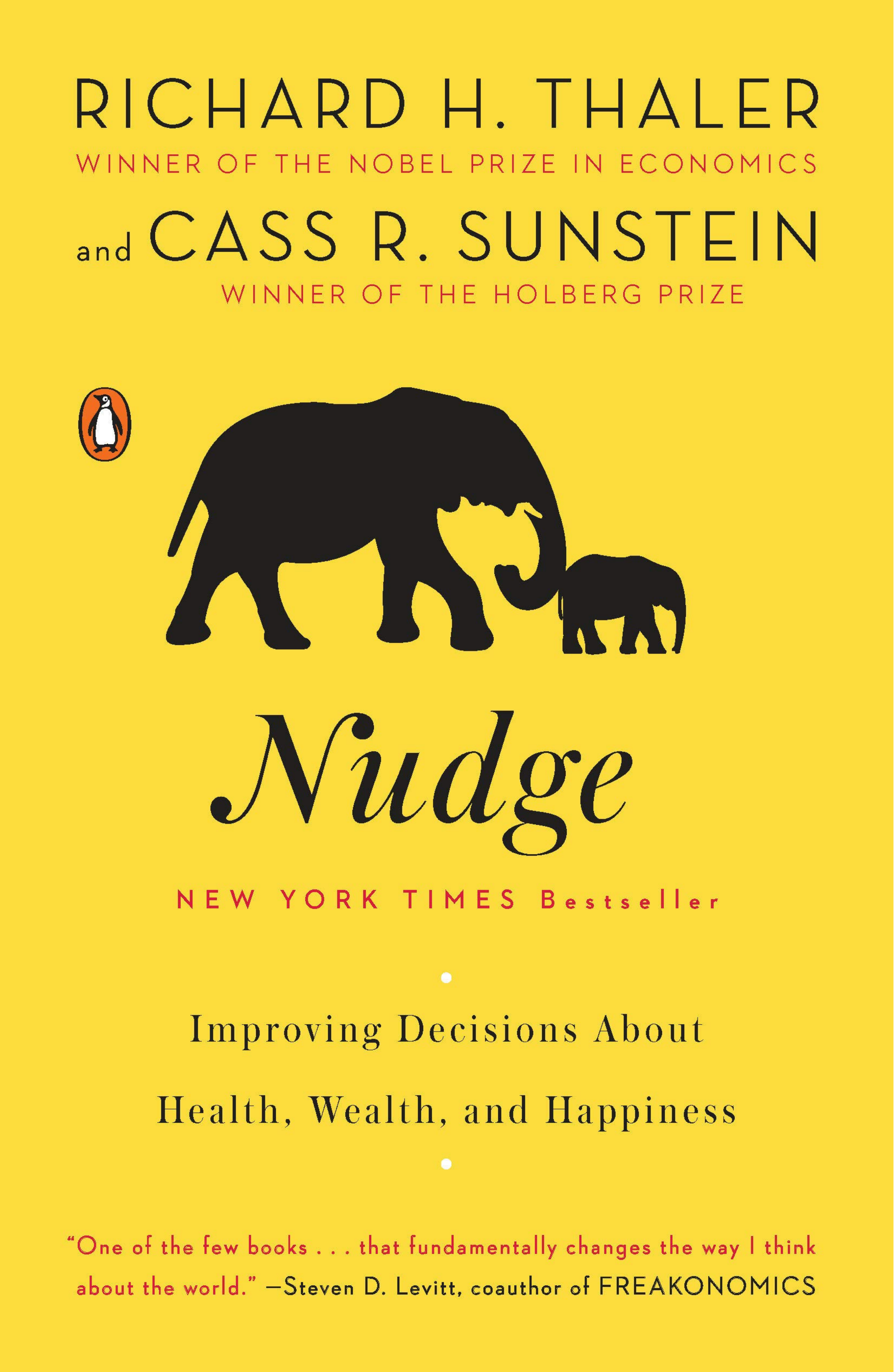
Director
MRC Epidemiology Unit
Cardiff

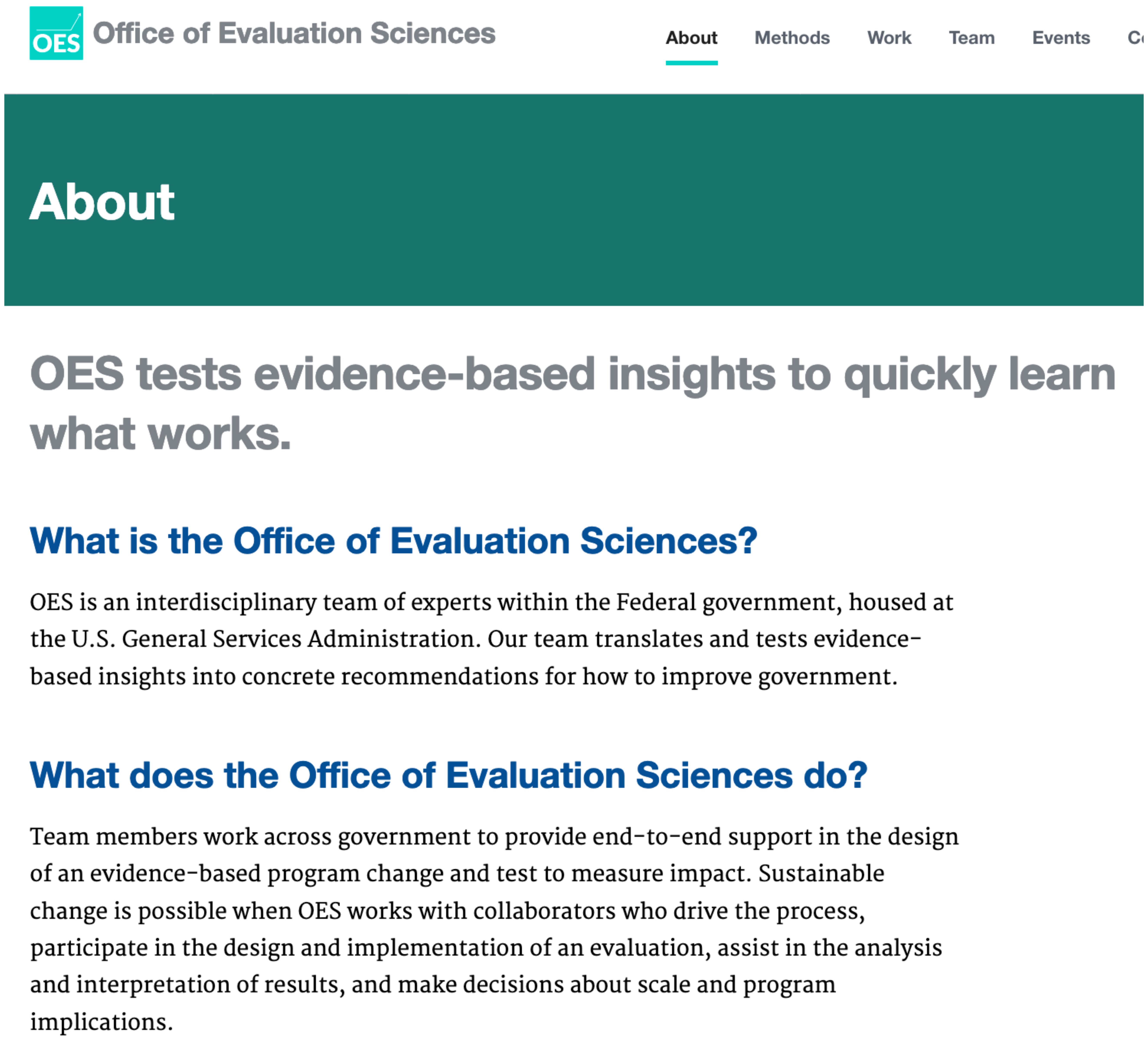
THE NUFFIELD PROVINCIAL HOSPITALS TRUST

1972

Cameron et al (2016): RCT in development







OES Office of Evaluation Sciences

About Methods Work Team Events C

About

OES tests evidence-based insights to quickly learn what works.

What is the Office of Evaluation Sciences?

OES is an interdisciplinary team of experts within the Federal government, housed at the U.S. General Services Administration. Our team translates and tests evidence-based insights into concrete recommendations for how to improve government.

What does the Office of Evaluation Sciences do?

Team members work across government to provide end-to-end support in the design of an evidence-based program change and test to measure impact. Sustainable change is possible when OES works with collaborators who drive the process, participate in the design and implementation of an evaluation, assist in the analysis and interpretation of results, and make decisions about scale and program implications.

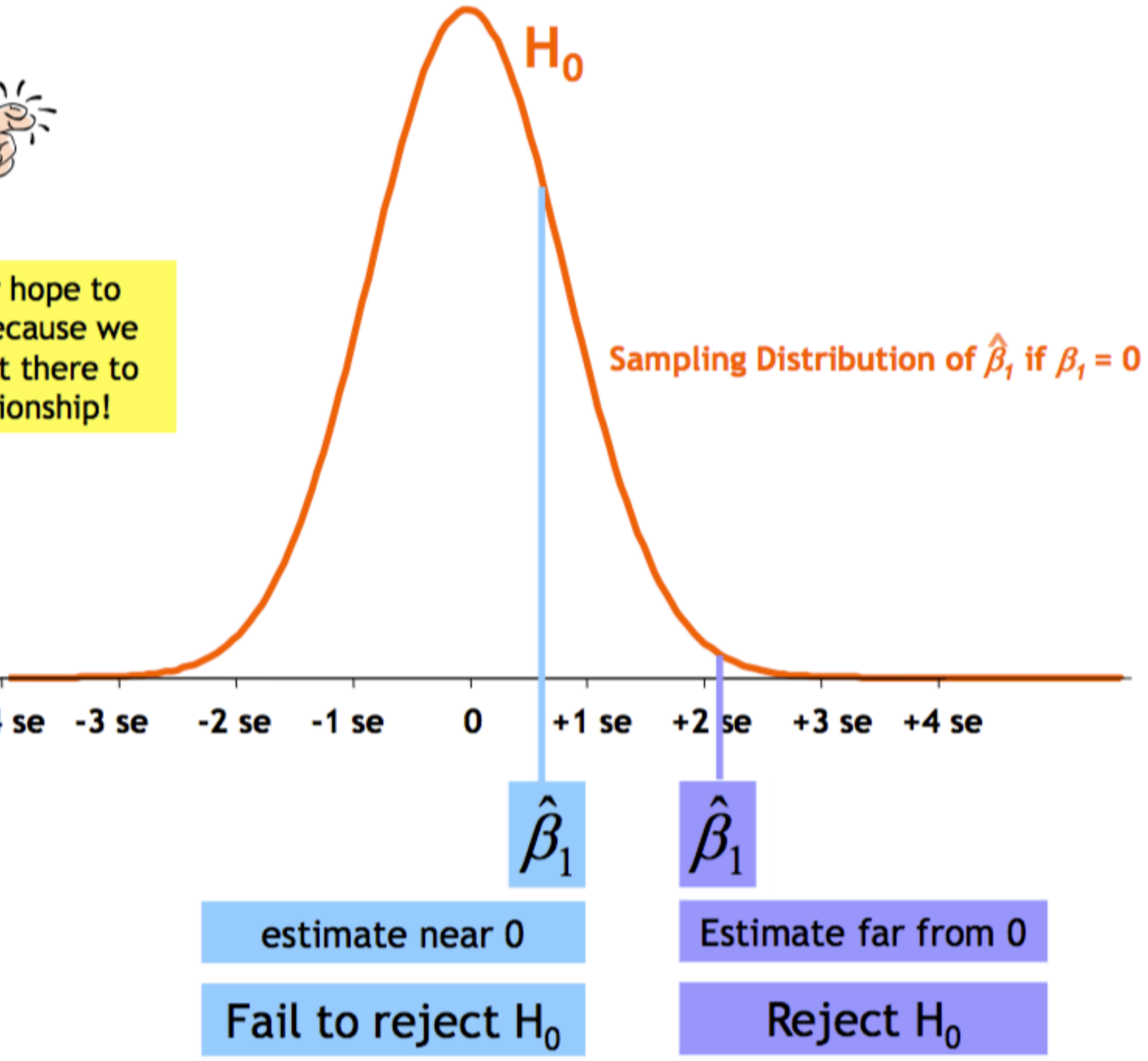
Final Project

If you want to do your own project, please **schedule office hours next week**.
meetme.so/natematias-soc412

Null Hypothesis Testing

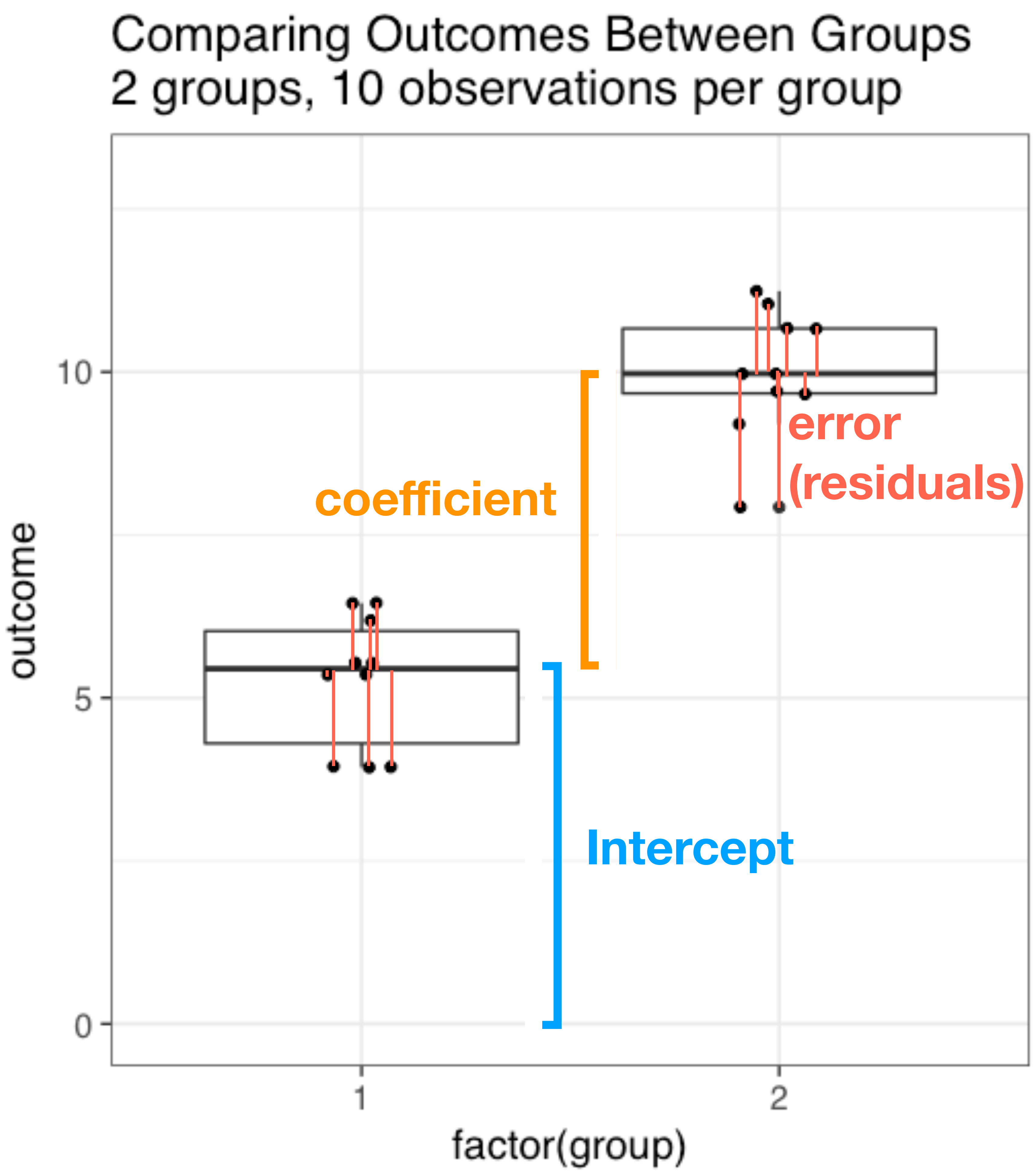
revisiting basics of OLS linear regression

We usually hope to reject H_0 because we usually want there to be a relationship!



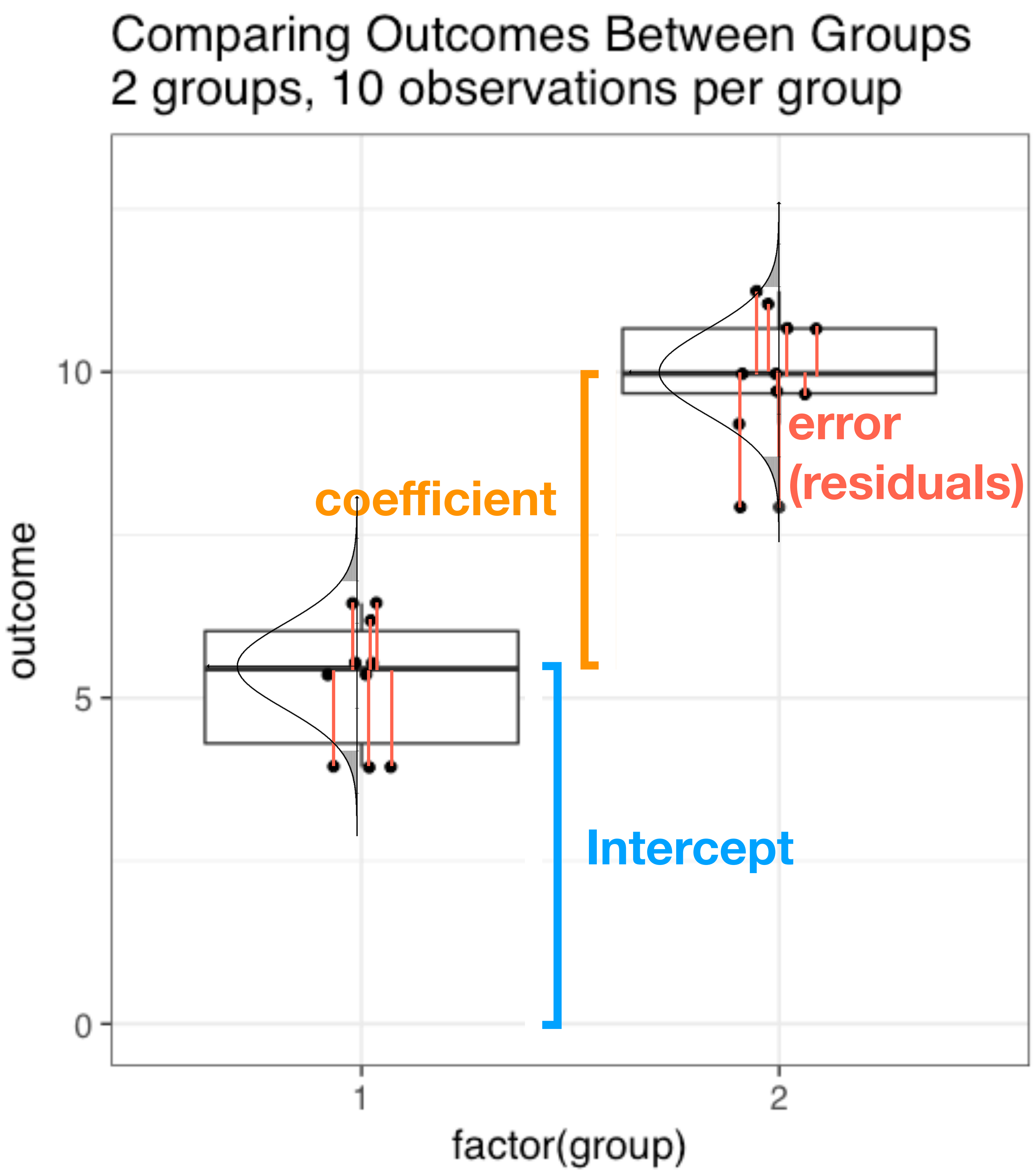
Error

revisiting basics of OLS linear regression



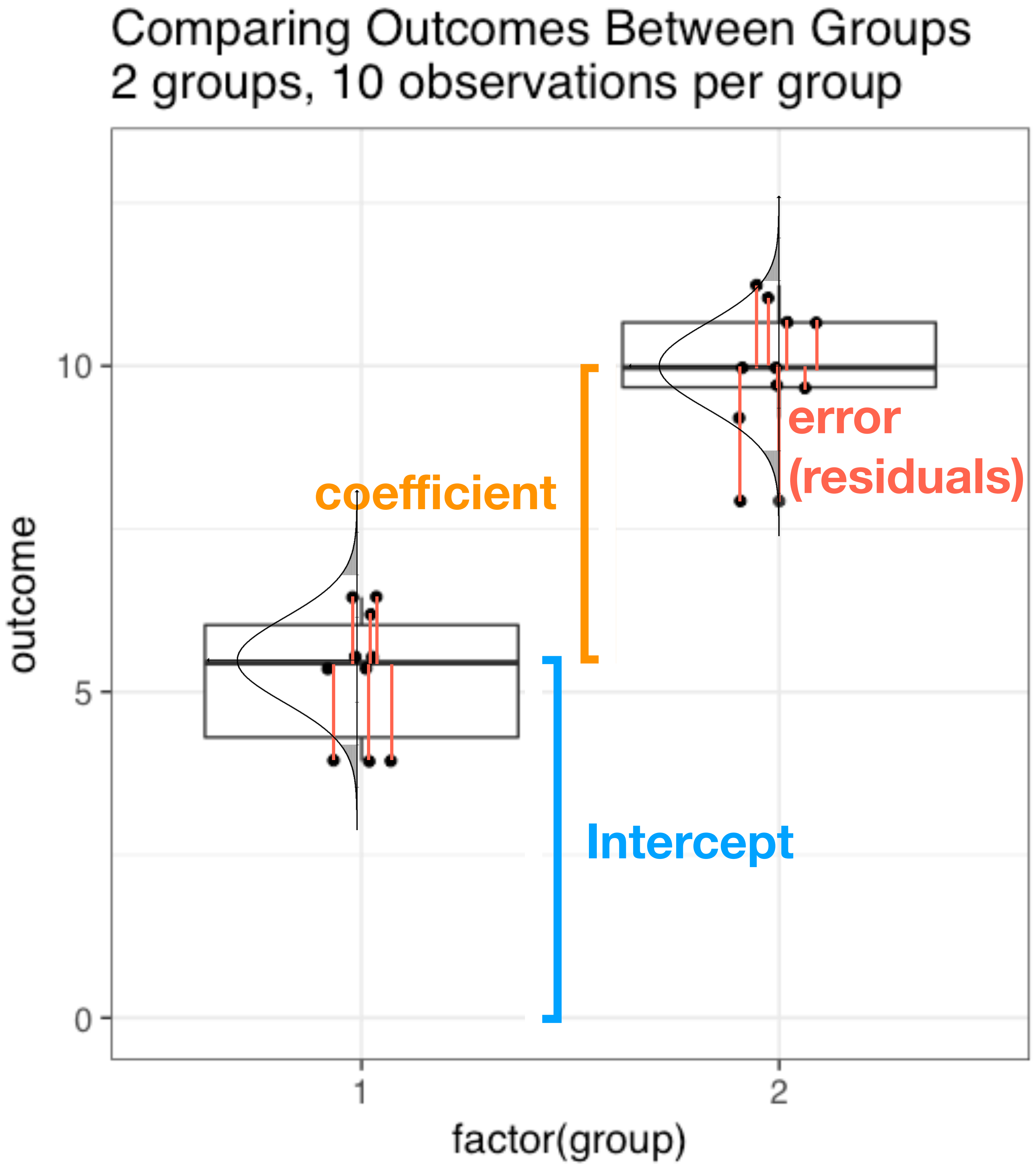
Distribution of Errors

assumption of OLS: error is normally distributed & homoscedastic



Root Mean Square Error

the estimated standard deviation of the residuals

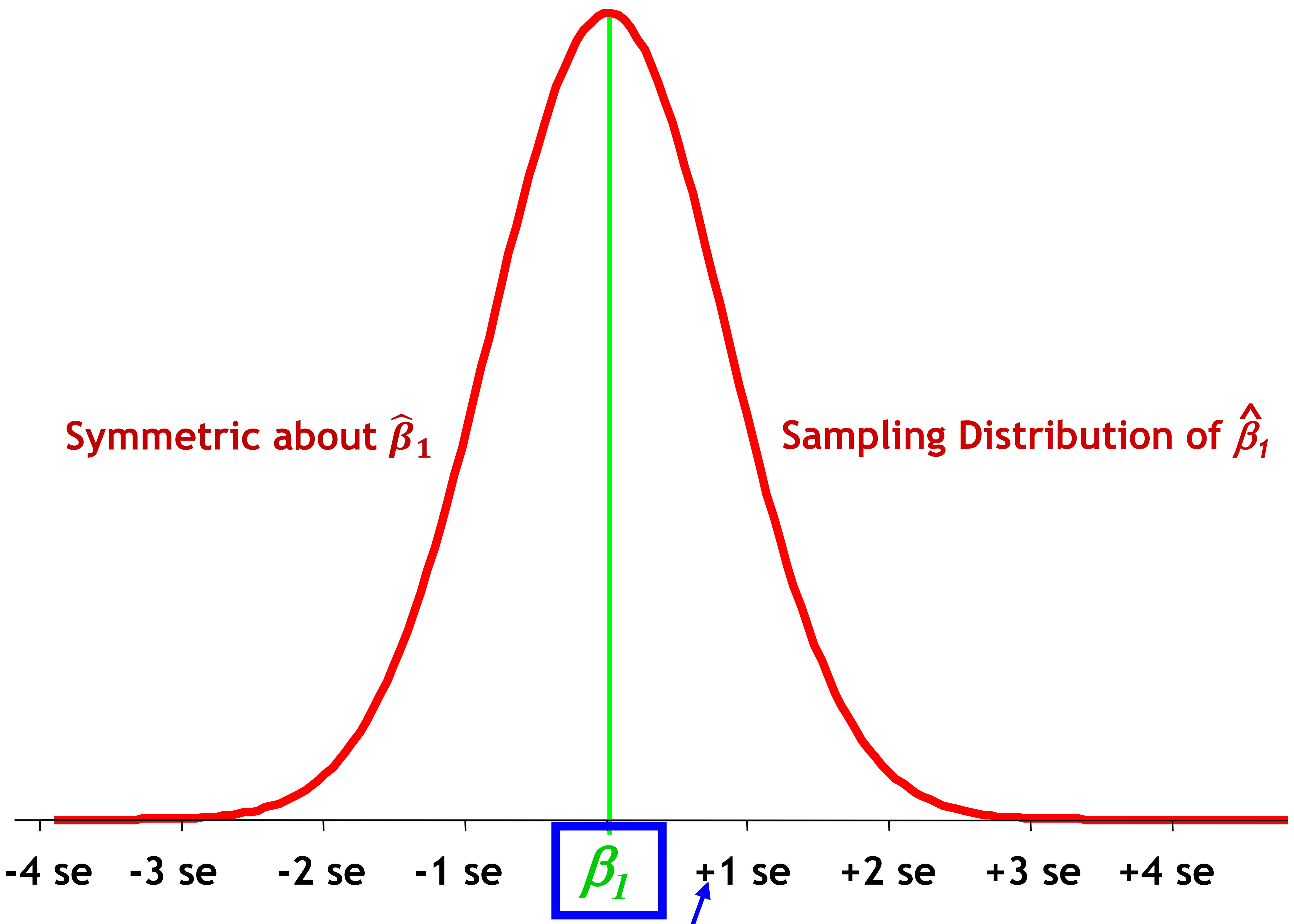



```
m1 <- lm(outcome ~
           group == 2 ,
           data=posts)
```

```
rmse <- sqrt(
 sum(m1$residuals^2) /
    length(m1$residuals)
)
```

Standard Error

The standard deviation of the sampling distribution (of a particular statistic)



Definition: The *standard deviation of a sampling distribution* is known as a "standard error," commonly abbreviated as "se"

$$se(\hat{\beta}) = \frac{\sigma}{\sqrt{(n-1)\times Var(X)}}$$

$\sigma = RMSE$

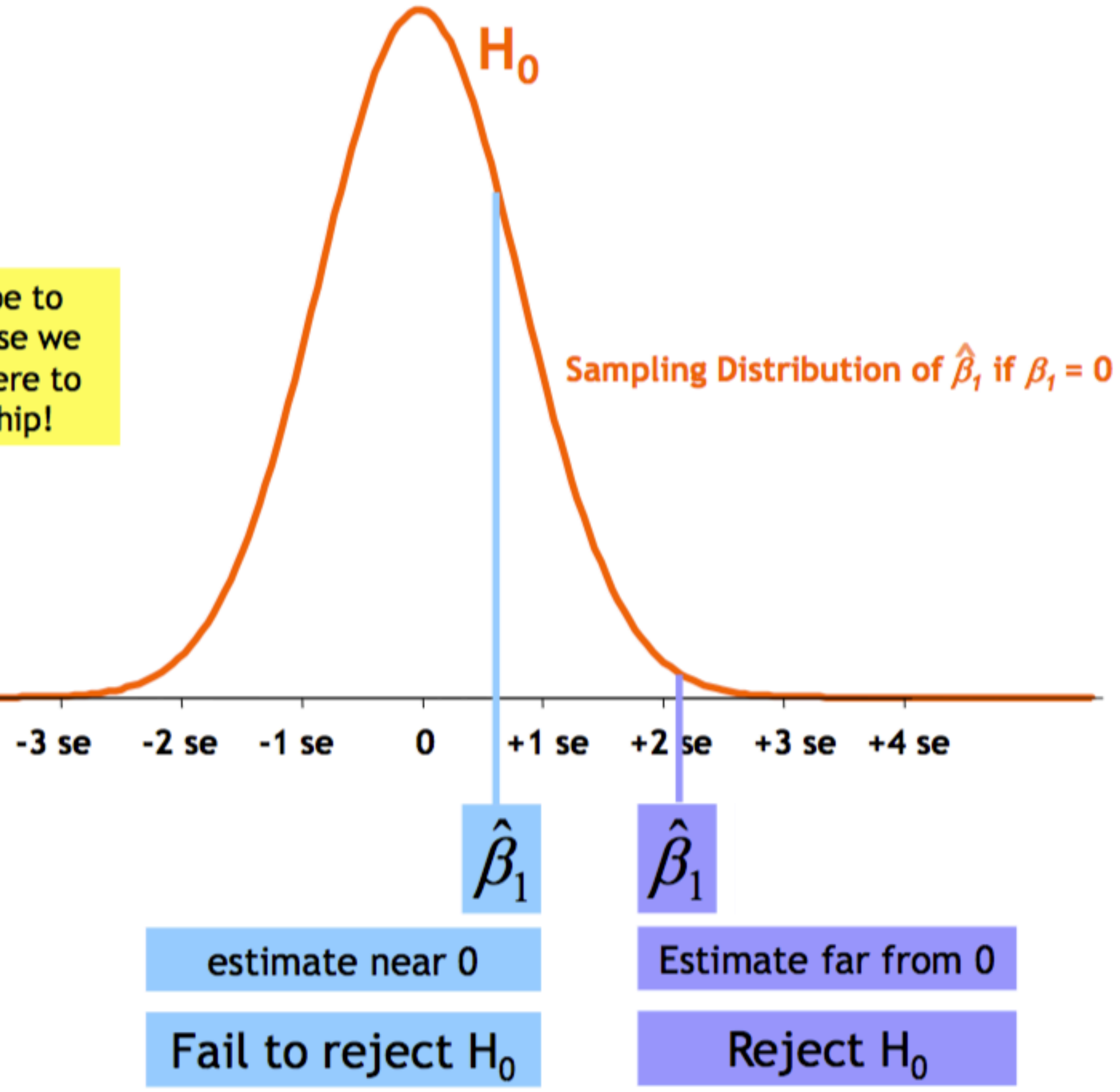
$n = num\ of\ observations$

$Var(X) = \ variance\ of\ predictor\ (TREAT)$

Null Hypothesis Testing

revisiting basics of OLS linear regression

We usually hope to reject H_0 because we usually want there to be a relationship!



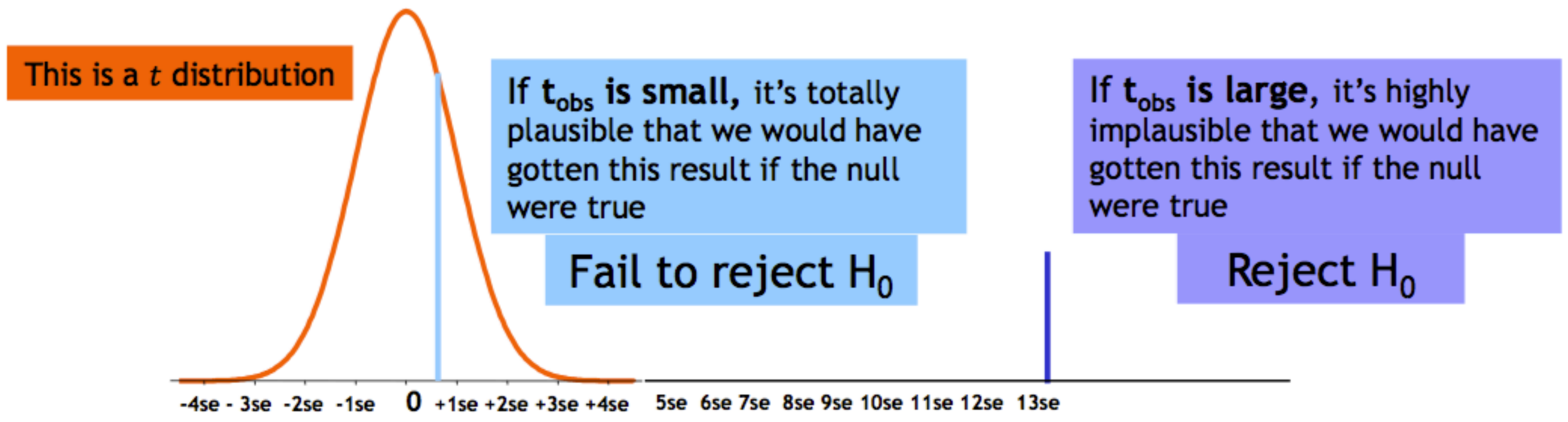
t-statistic

revisiting basics of OLS linear regression

$$t_{obs} = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \quad where\ H_0: \beta = 0$$

So t_{obs} tells us how many standard errors away from 0 our sample estimate is.

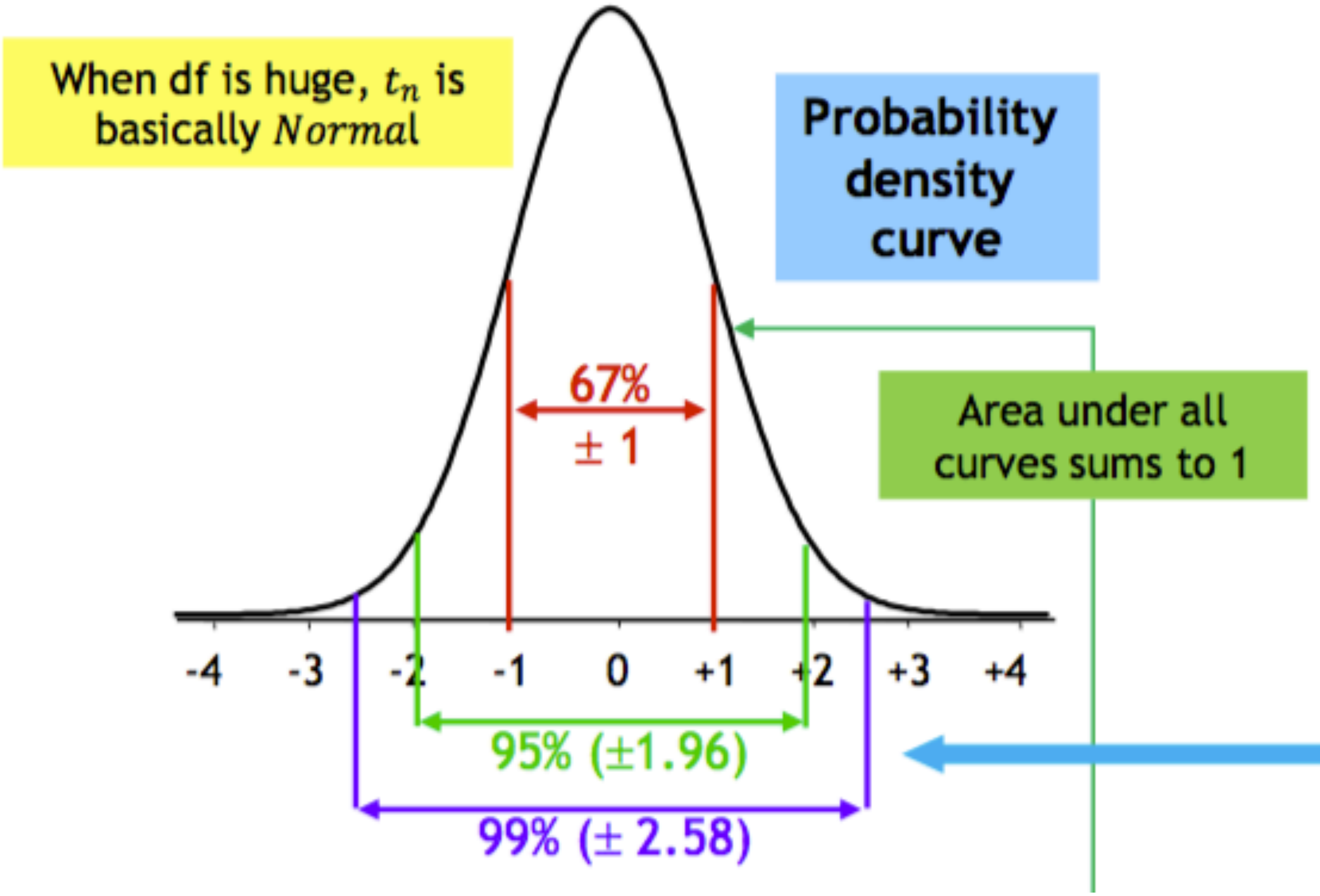
There's nothing magical about $H_0: \beta_1 = 0$, it could be anything (e.g., $H_0: \beta_1 = 1$)



t-distribution

revisiting basics of OLS linear regression

$$t_{obs} = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \quad where\ H_0 \colon \beta = 0$$



How large is large? Critical values of $t_{observed}$			
df	Two-sided probability level, p		
	0.10	0.05	0.01
10	1.81	2.23	3.17
20	1.72	2.09	2.85
30	1.70	2.04	2.75
50	1.68	2.01	2.68
100	1.66	1.98	2.63
infinite	1.64	1.96	2.58

confidence intervals

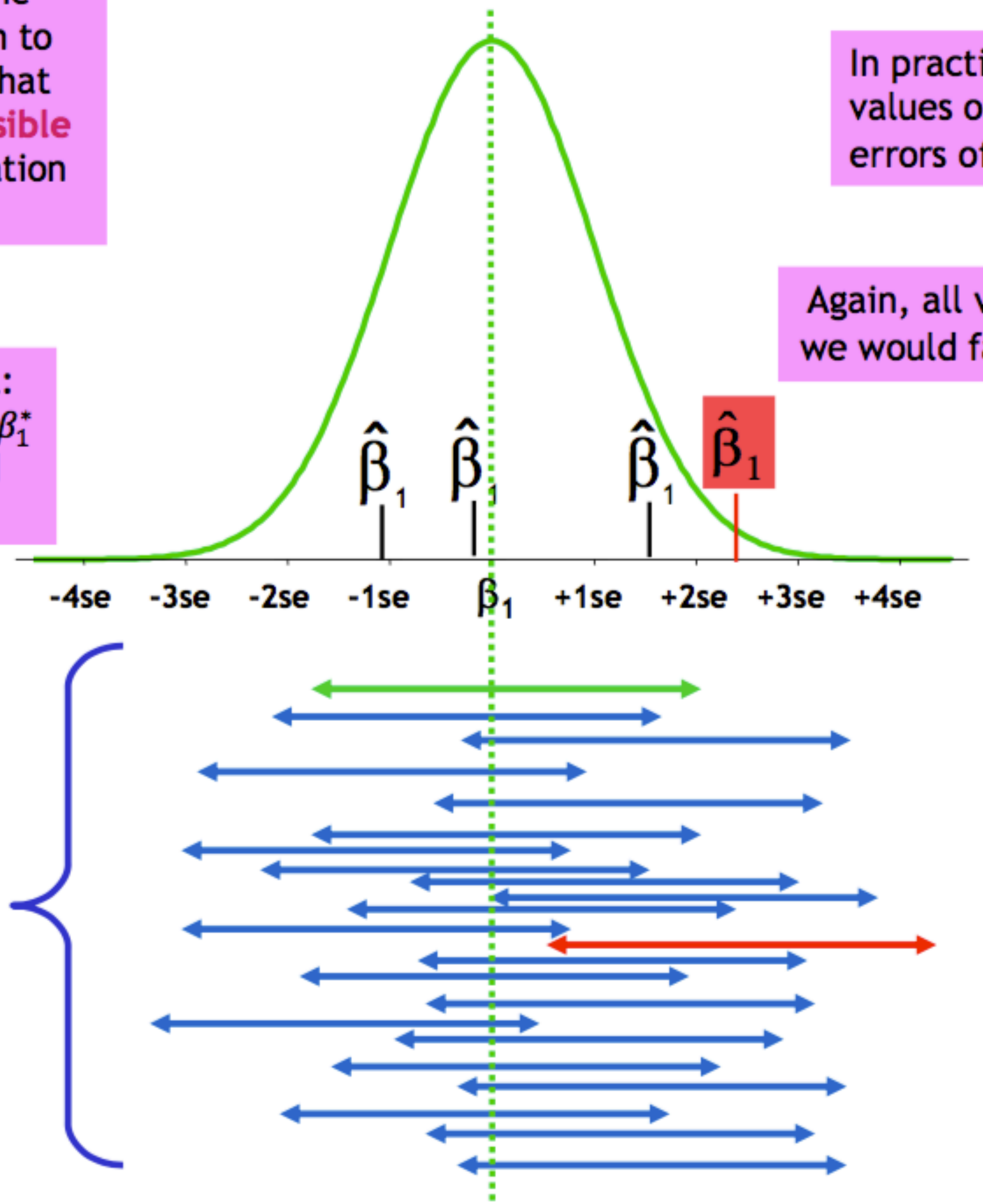
Idea: Can we use the sampling distribution to construct intervals that offer a **range of plausible values** for the population parameter?

Yes!

Confidence interval: Set of all values of β_1^* for which we would reject $H_0: \beta_1 = \beta_1^*$

In practice: Take all possible values of β_1 within ~2 standard errors of $\hat{\beta}_1$

Again, all values of $\beta_{possible}$ for which we would fail to reject $H: \beta = \beta_{possible}$



$$95\%\ conf\ interval = \hat{\beta} \pm 1.96\ se(\hat{\beta})$$

For every 20 95% confidence intervals we construct, we estimate that an average of 1 won't cover the true value of β_1

We don't know whether this is one of the lucky 95% that do cover the true value or the unfortunate 5% that don't

p-value

probability of observing a coefficient this extreme if the null hypothesis were true