EDS222 Week 8

Hypothesis Testing

November 19, 2024

Agenda

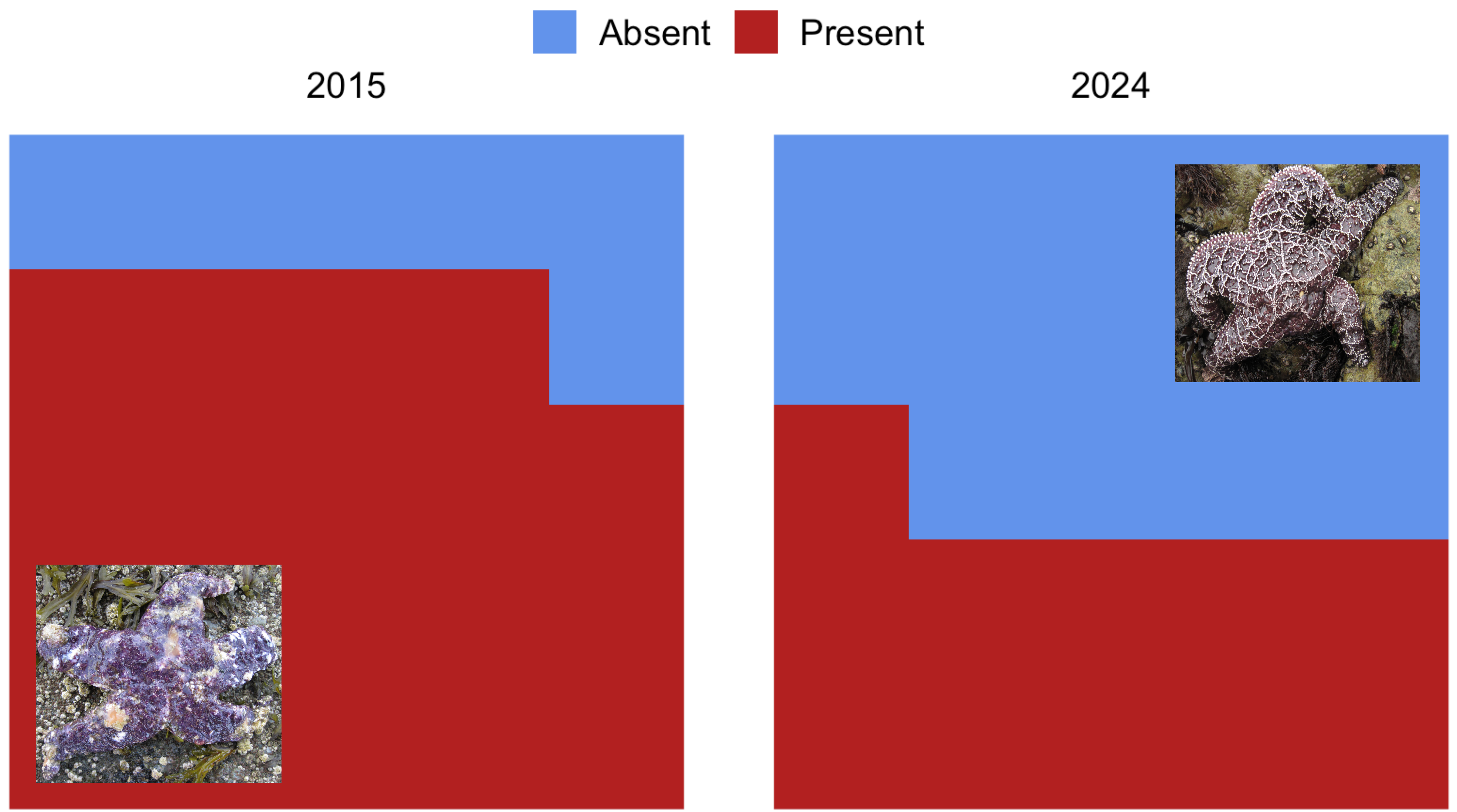
- **Hypothesis testing by randomization**
  - Null and alternative hypotheses
  - Sample statistics and sampling distributions
  - P-values and rejecting the null
- **Hypothesis testing in practice**
  - Central limit theorem
  - Standard errors
- **Confidence intervals**
  - Interpretation
  - Effect sizes

Hypothesis testing by randomization

Sea star wasting sydrome

Hypothesis testing by randomization

Sea star wasting sydrome



Hypothesis testing by randomization

Overview

- **Overall question**
  - Did sea star wasting syndrome incidence decrease from 2015 to 2024?
- **Procedure**

1.

2.

3.

4.

5.

Hypothesis testing by randomization

Key terms

Null and alternate hypotheses

Sample statistic

Point estimate

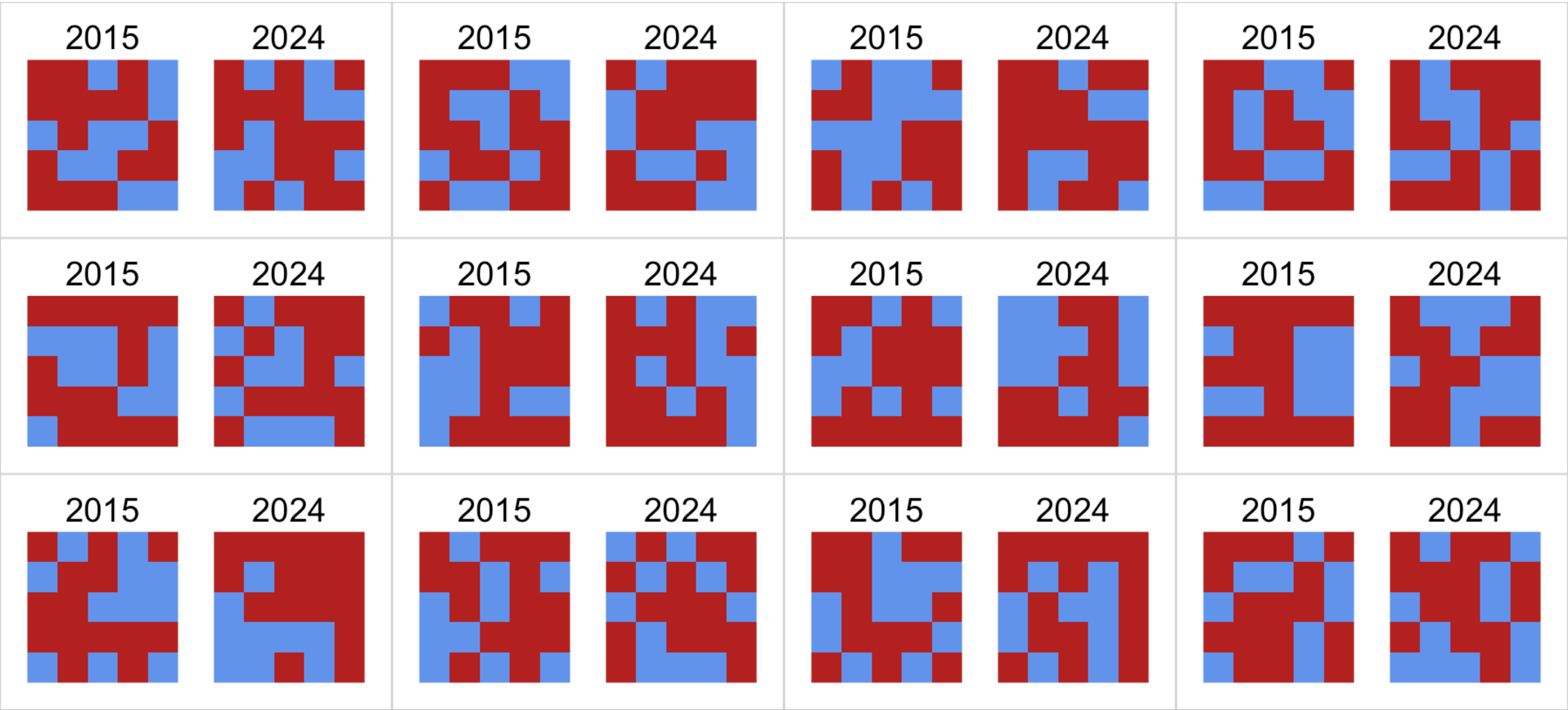
Sampling distribution

Hypothesis testing by randomization

Sea star wasting sydrome

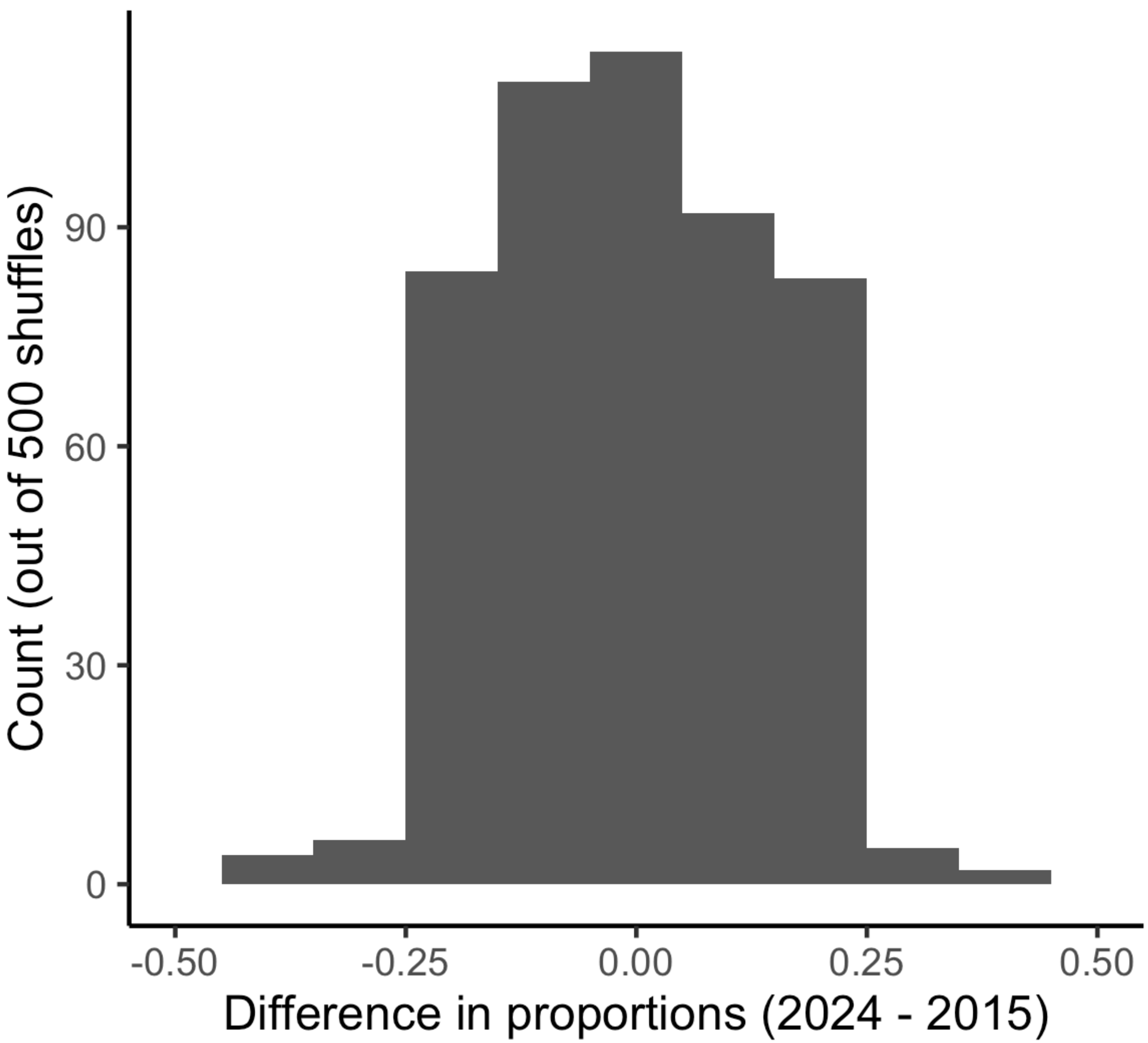
Hypothesis testing by randomization

Quantify uncertainty by shuffling



Hypothesis testing by randomization

Probability of point estimate under the null



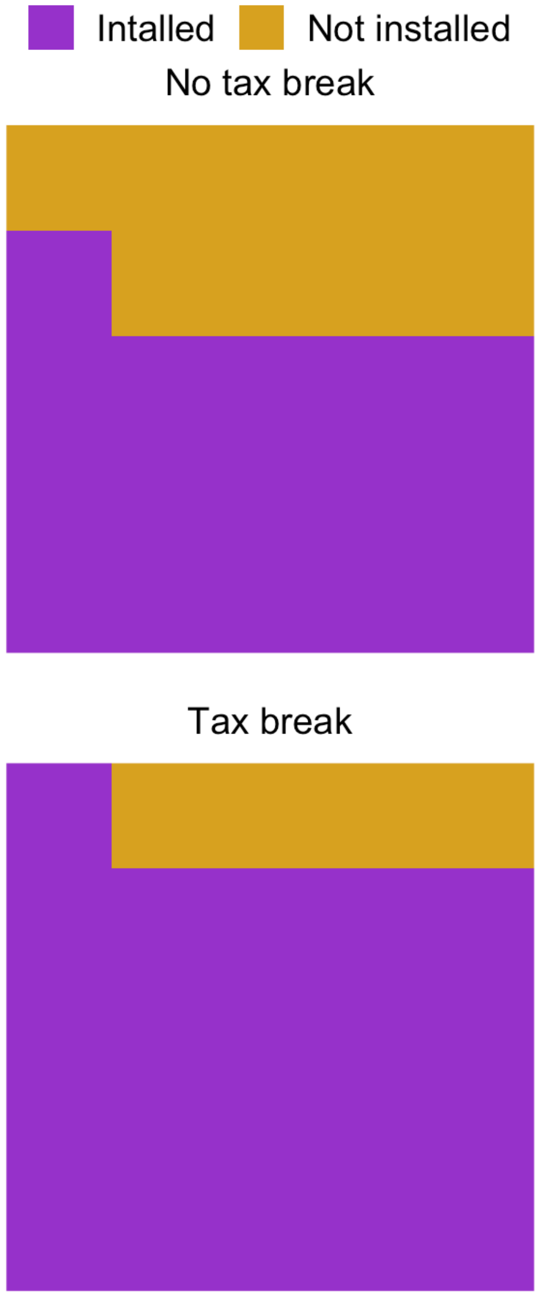
Hypothesis testing by randomization

Reject or fail to reject the null?

Hypothesis testing by randomization

Your turn

Do tax breaks incentivize solar panel installation?

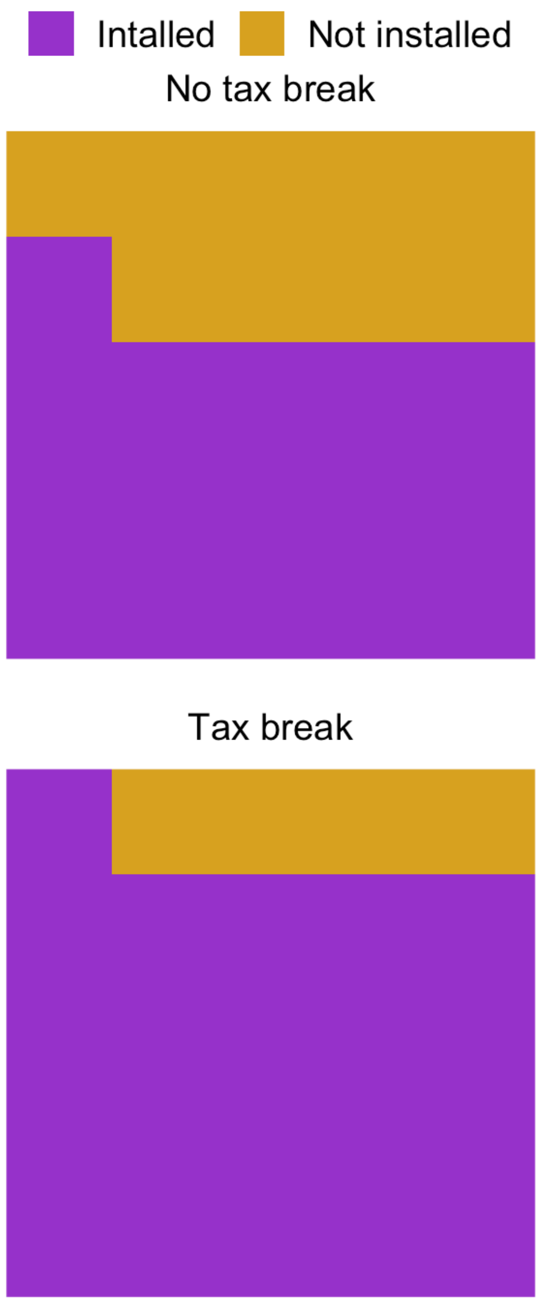


1. Define the null and alternate hypotheses
2. Calculate the point estimate of the sample statistic
3. Quantify the uncertainty in the sampling distribution
4. Calculate probability of point estimate under the null
5. Reject or fail to reject null

Hypothesis testing by randomization

Your turn

Do tax breaks incentivize solar panel installation?



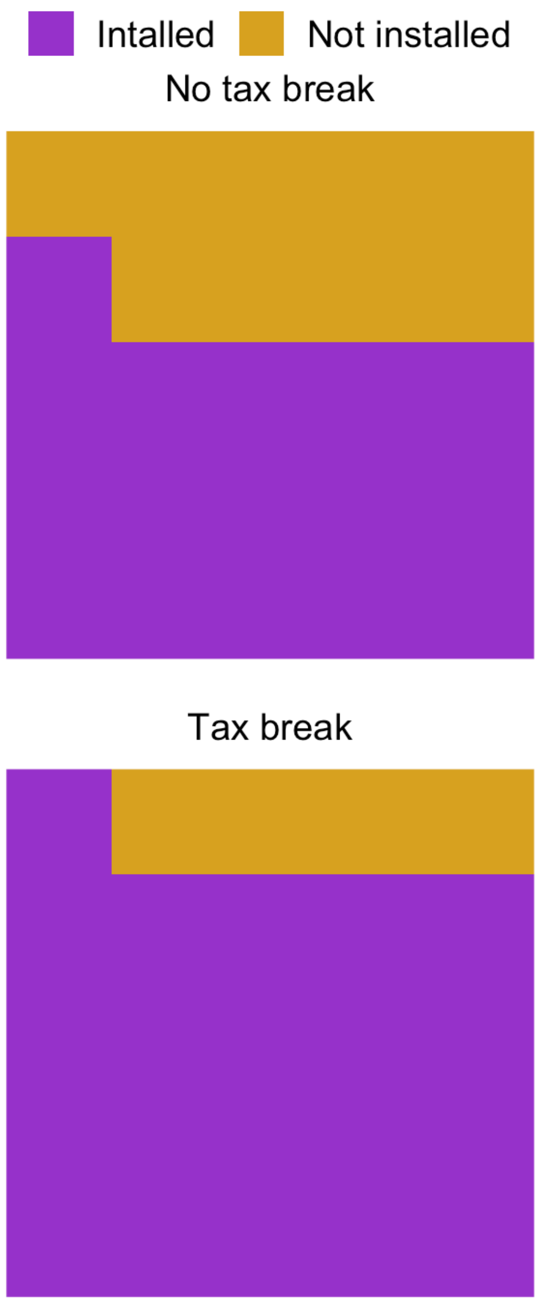
1. Define the null and alternate hypotheses

   What are H_0 and H_A?

Hypothesis testing by randomization

Your turn

Do tax breaks incentivize solar panel installation?



2. Calculate the point estimate of the sample statistic

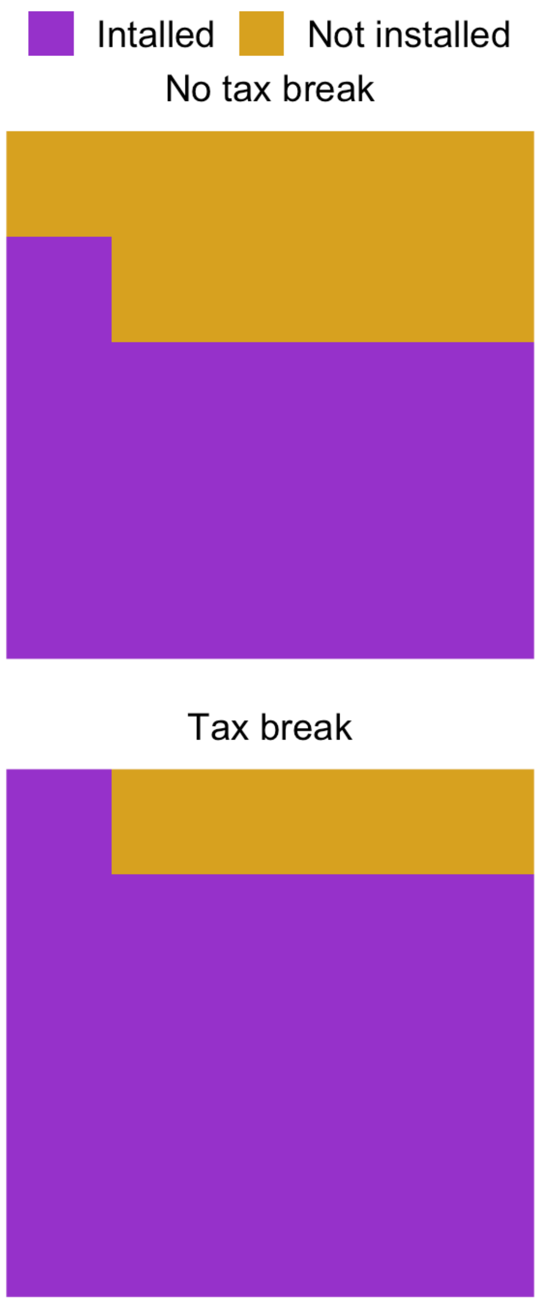
   Draw lines from *a,b,c,d* in the equation below to the corresponding parts of the figure on the left.

   $$\text{difference in proportions} = \frac{a}{b} - \frac{c}{d}$$

Hypothesis testing by randomization

Your turn

Do tax breaks incentivize solar panel installation?



3. Quantify the uncertainty in the sampling distribution

Which R function will help?

A) `rnorm()`

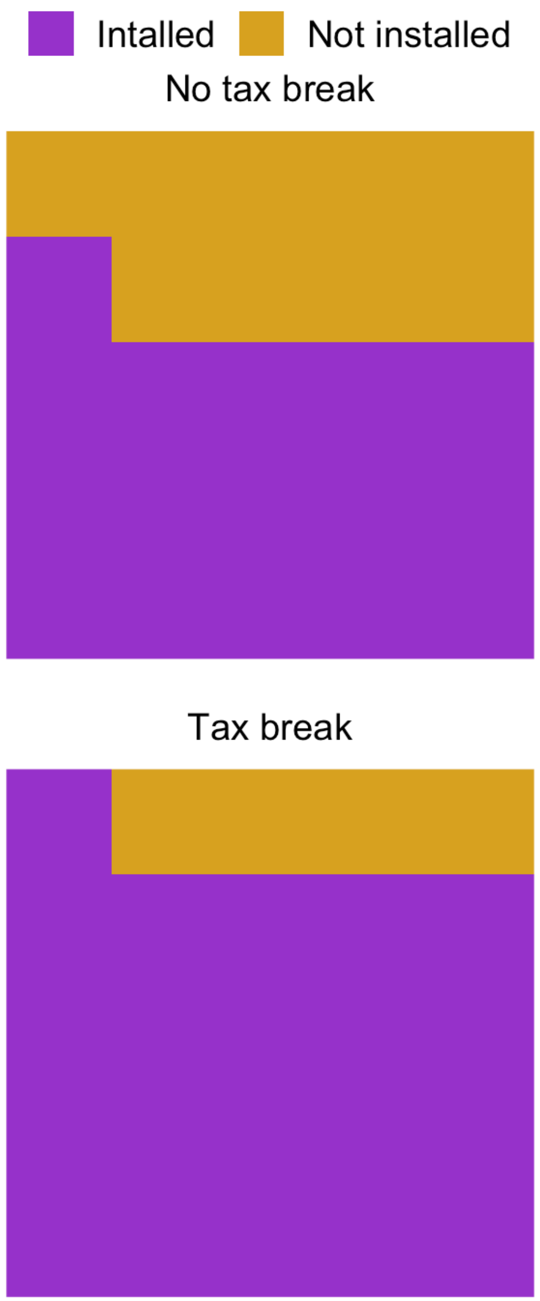
B) `sample()`

C) `dnorm()`

Hypothesis testing by randomization

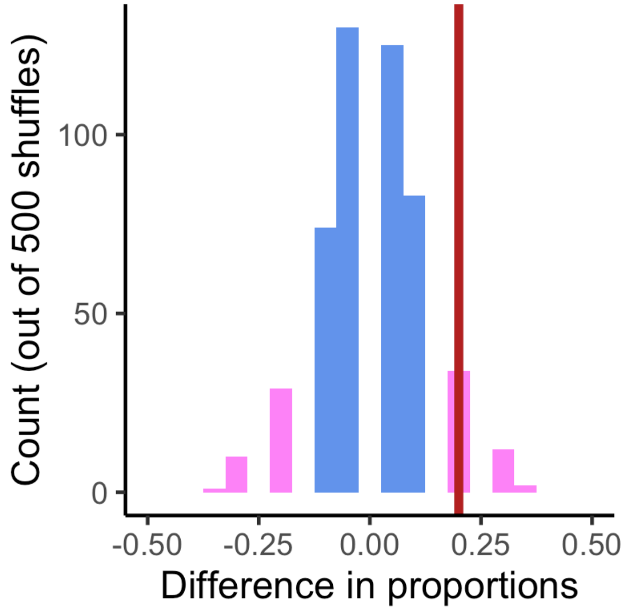
Your turn

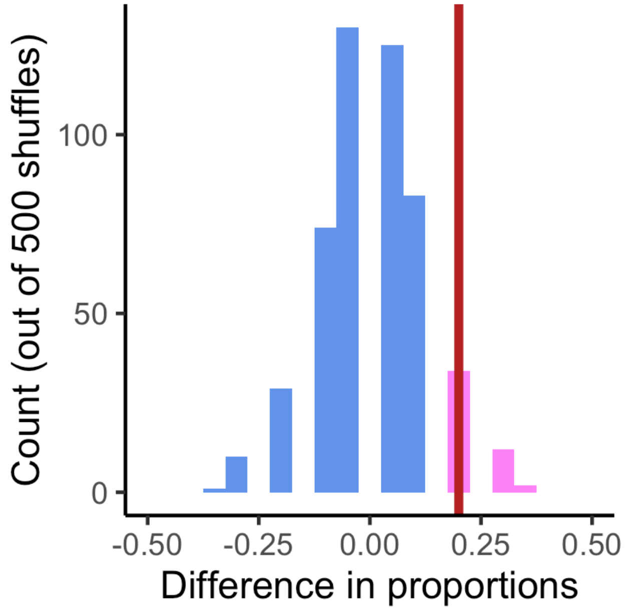
Do tax breaks incentivize solar panel installation?

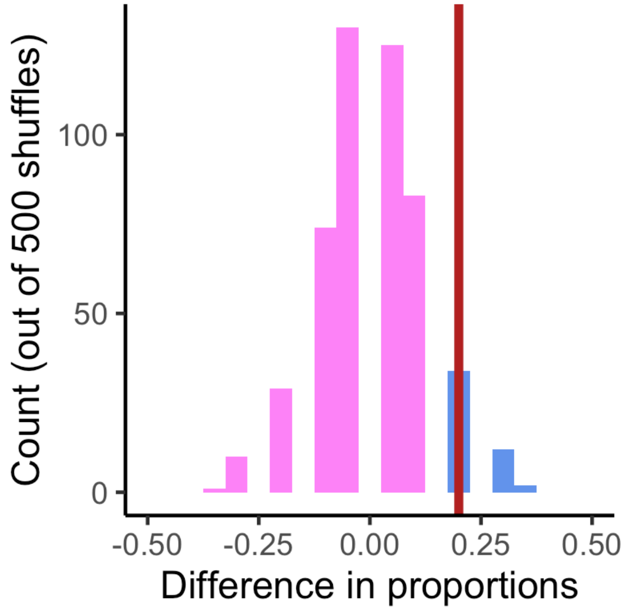


4. Calculate probability of point estimate under the null

   The histograms below show the results of randomization and the red line is the observed difference. Which figure shows the p-value in pink?



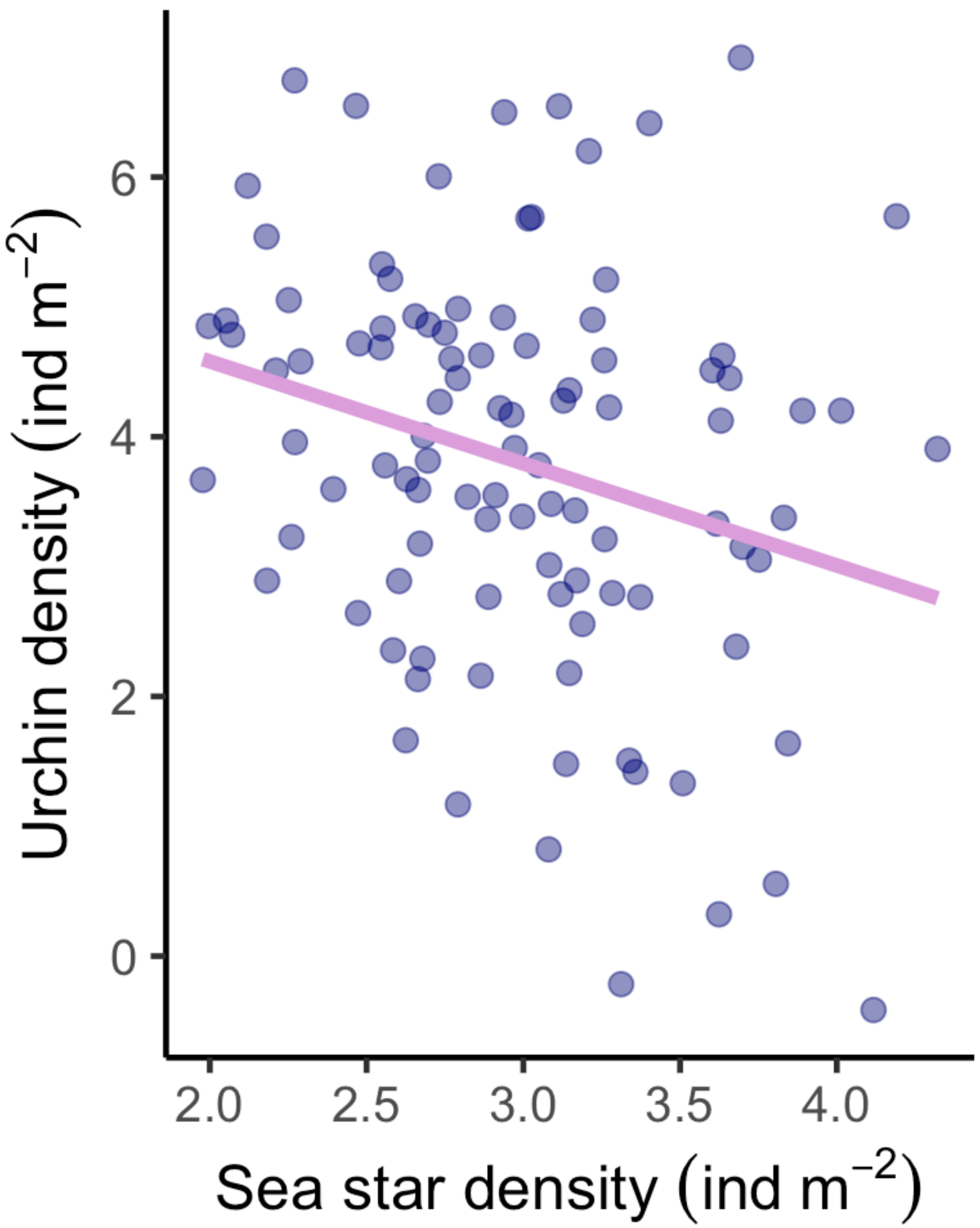




Hypothesis testing by randomization

Applicable to regression and other models

Do sea stars reduce urchin populations?

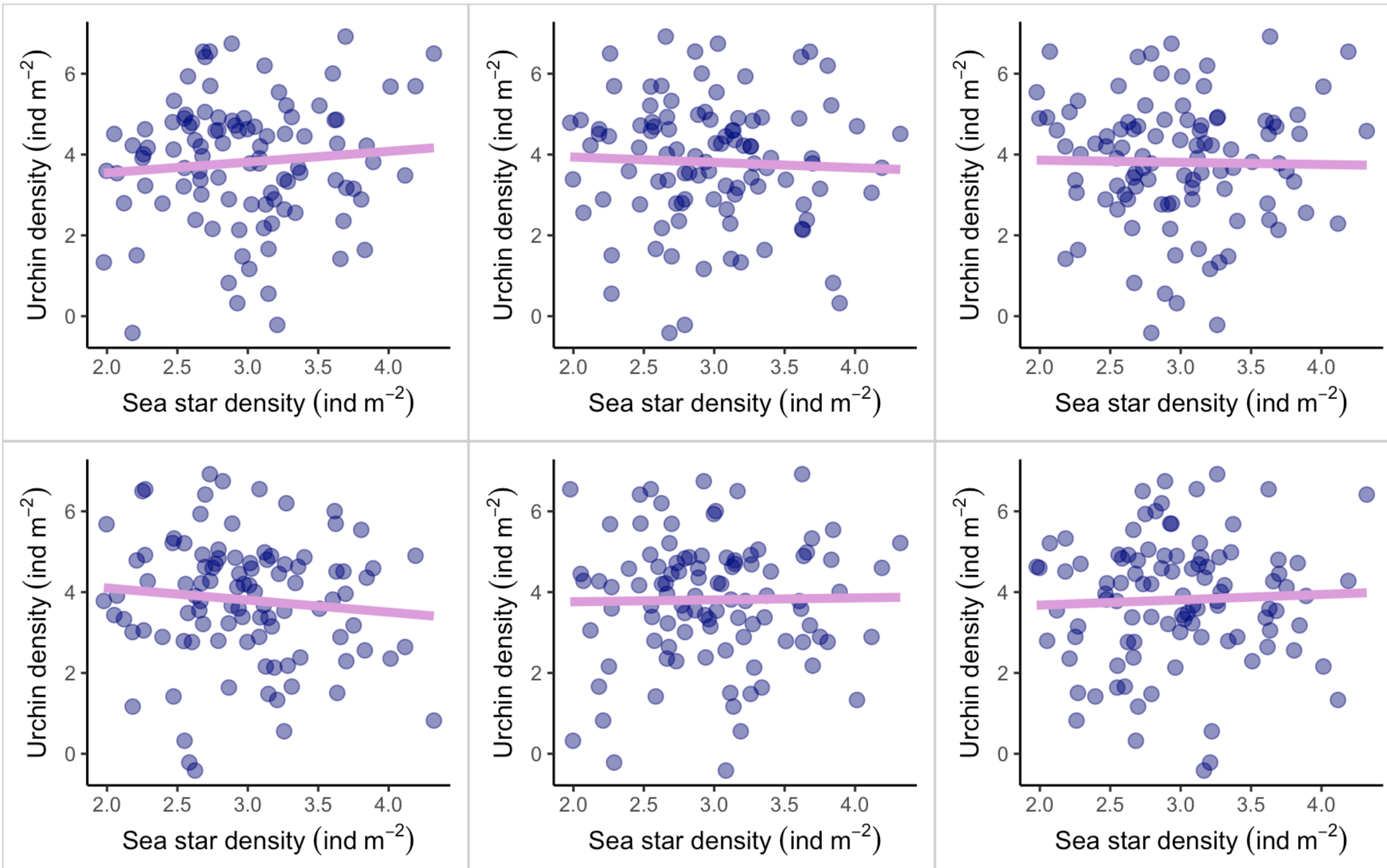


1. What are the null and alternate hypotheses?

2. What's the relevant sample statistic?

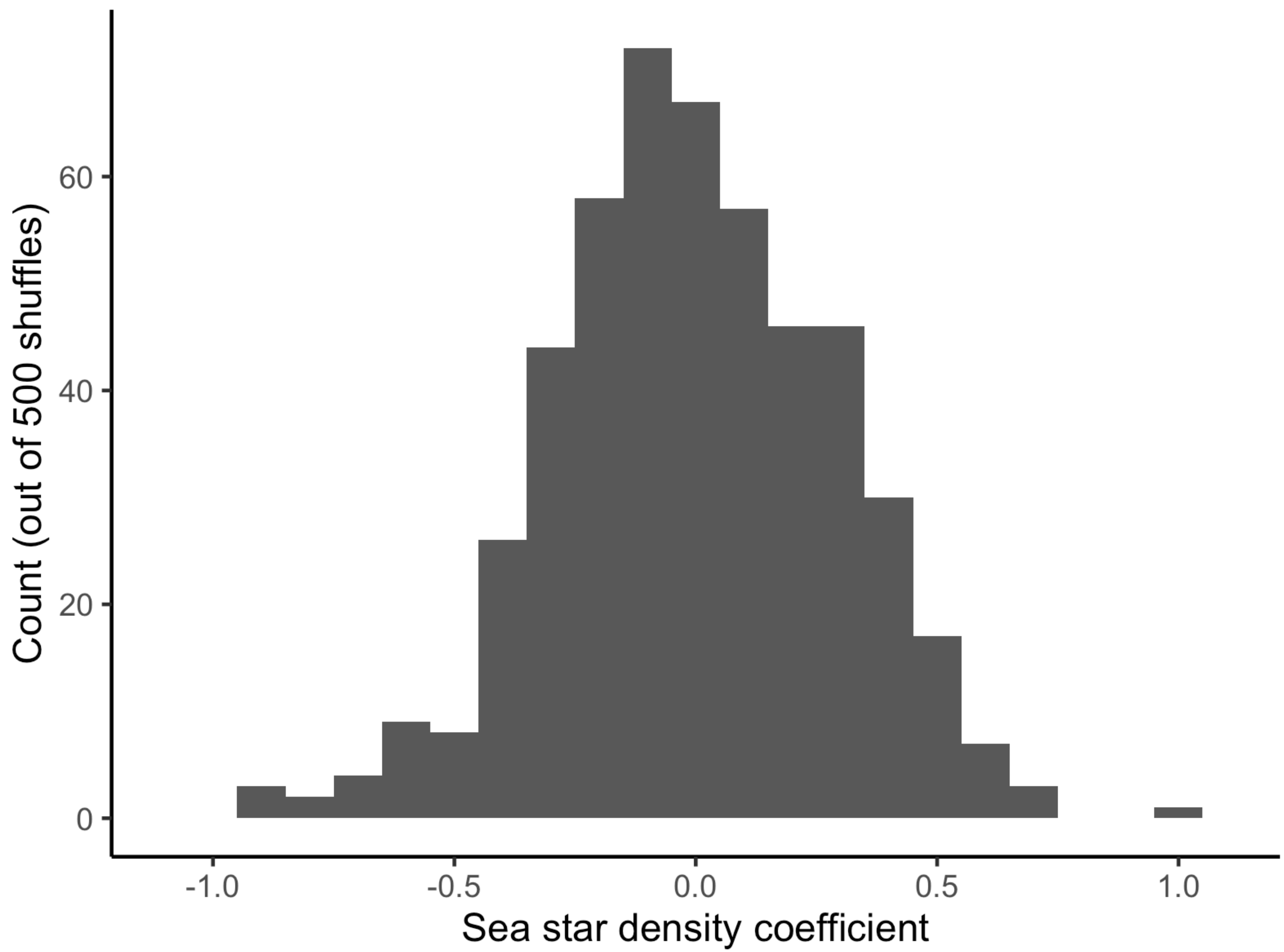
Hypothesis testing by randomization

Applicable to regression and other models



Hypothesis testing by randomization

Applicable to regression and other models



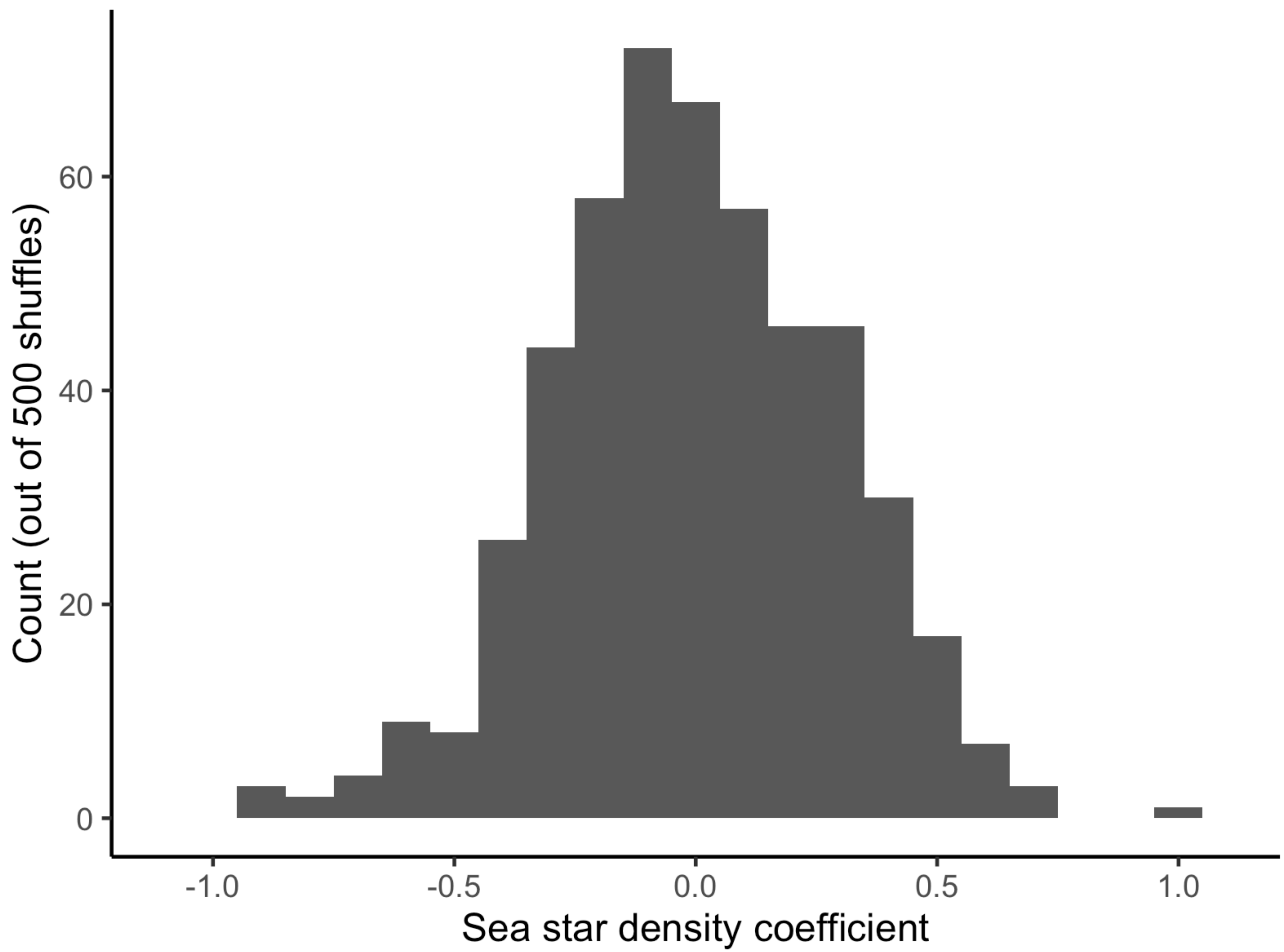
Hypothesis testing by randomization

Recap

1. Formulate your hypotheses
   H_0 = no effect, H_A = some effect
2. Calculate point estimate
   Difference in means, regression coefficient, etc
3. Quantify uncertainty in sampling distribution
   Shuffle data, recalculate point estimate, repeat
4. Calculate p-value
   Probability of point estimate <u>if null is true</u>
5. Reject or fail to reject the null
   Is $p \leq \alpha$?

Hypothesis testing in practice

Motivation



Hypothesis testing in practice

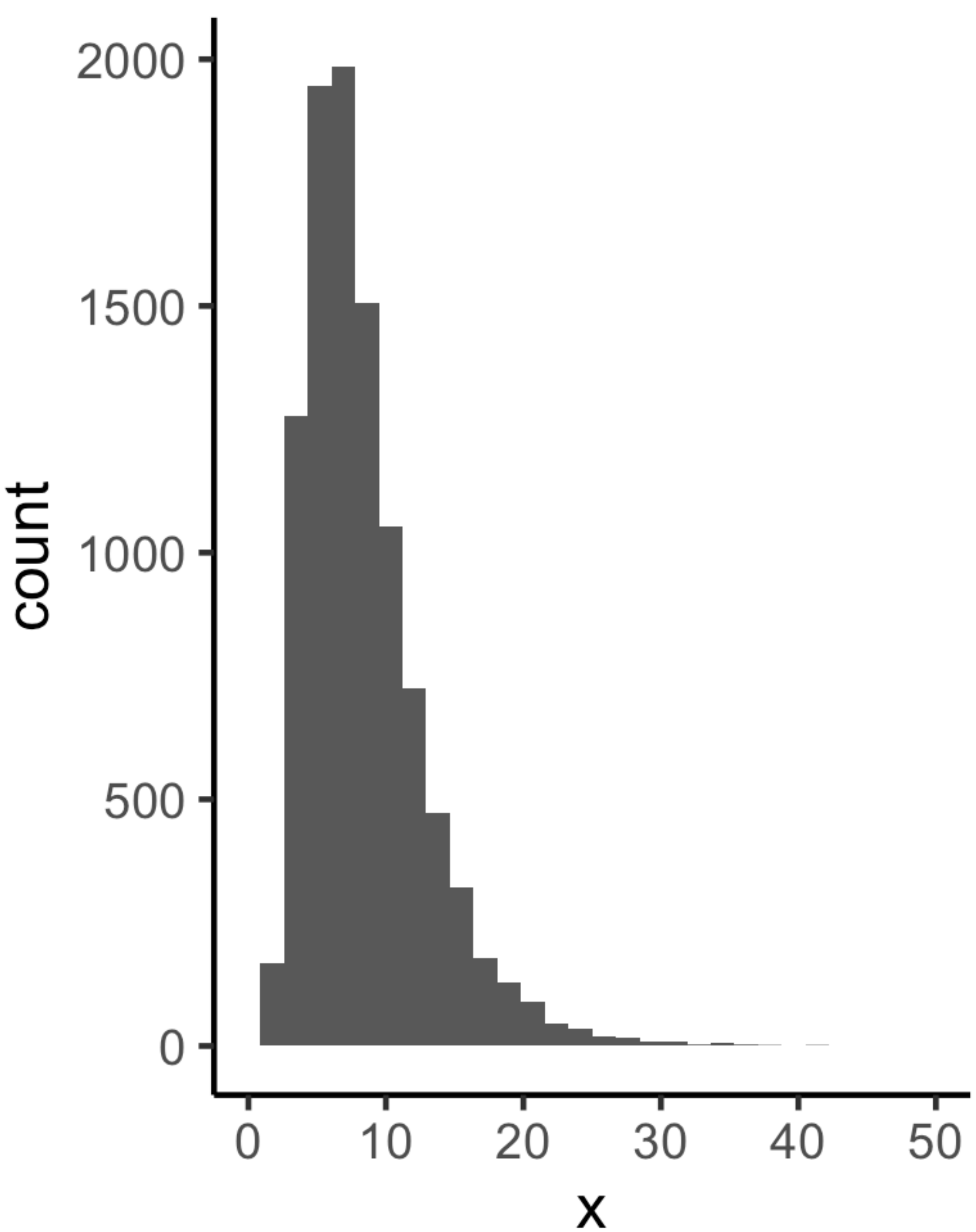
Central limit theorem

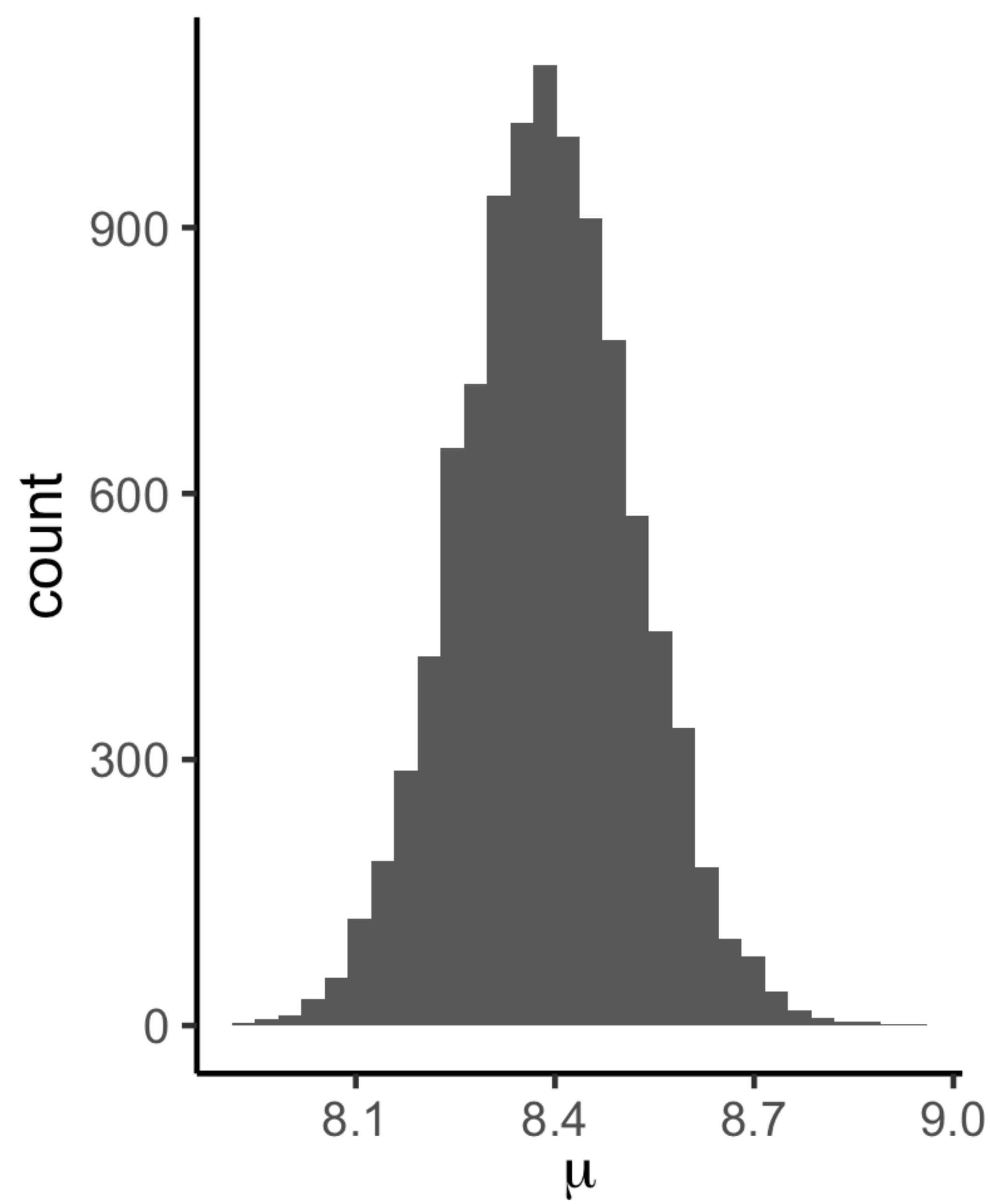
The Central Limit Theorem states:

If your sample size is large enough, then the sampling distribution for many sample statistics (difference in proportions, regression coefficients, etc) are approximately normal

Hypothesis testing in practice

Central limit theorem





Hypothesis testing in practice

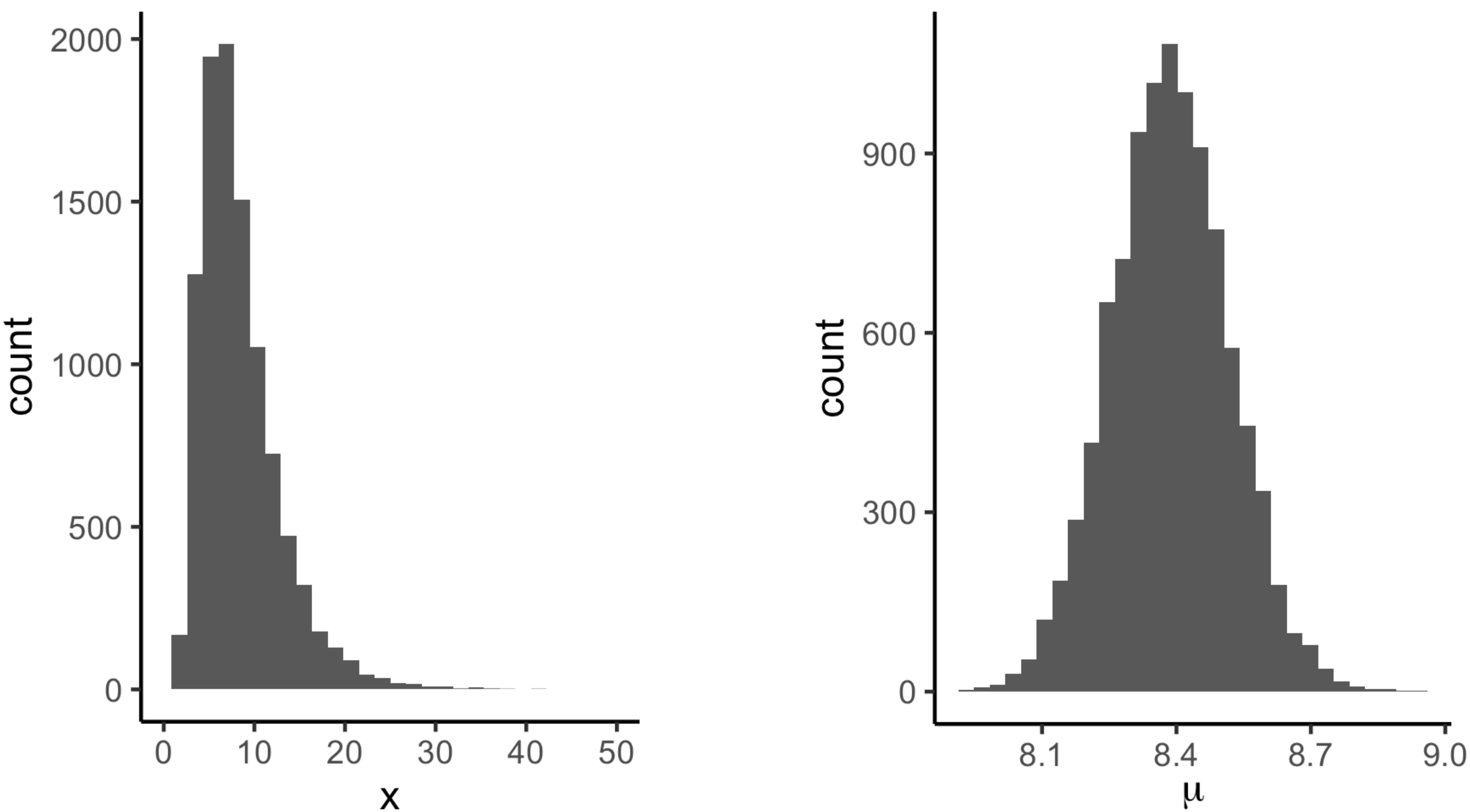
Central limit theorem

Try it on your own

```
# Roll a dice 10,000 times to get a non-normal population
# It's not even continuous!
x <- sample(1:6, 1e4, replace = TRUE)
ggplot(tibble(x), aes(x)) +
  geom_histogram(binwidth = 1, color = "blue", fill = NA) +
  theme_classic()
# Simulate the sampling distribution of the mean
# Do the following 1000 times
#   1. Sample 50 values from your non-normal population
#   2. Calculate the sample mean
mean_x <- replicate(
  1e3,
  mean(sample(x, size = 50))
)
ggplot(tibble(mean_x), aes(mean_x)) +
  geom_histogram(bins = 15, color = "blue", fill = NA) +
  theme_classic()
# Looks pretty normal!
```

Hypothesis testing in practice

Standard errors



Hypothesis testing in practice

Standard errors

Standard error
Standard deviation of the sampling statistic.

Problem
We only get one sample! Can't get the standard deviation of one data point.

Solution
Someone else solves the central limit theorem for you.

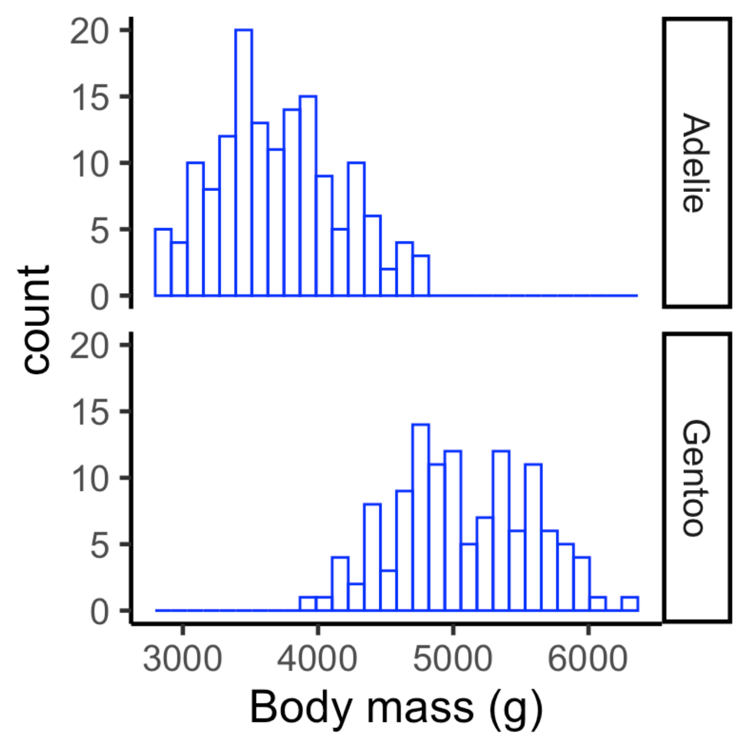
Note
Don't memorize equations! Demonstration purposes only.

Hypothesis testing in practice

Standard error of the *difference of means*

Population

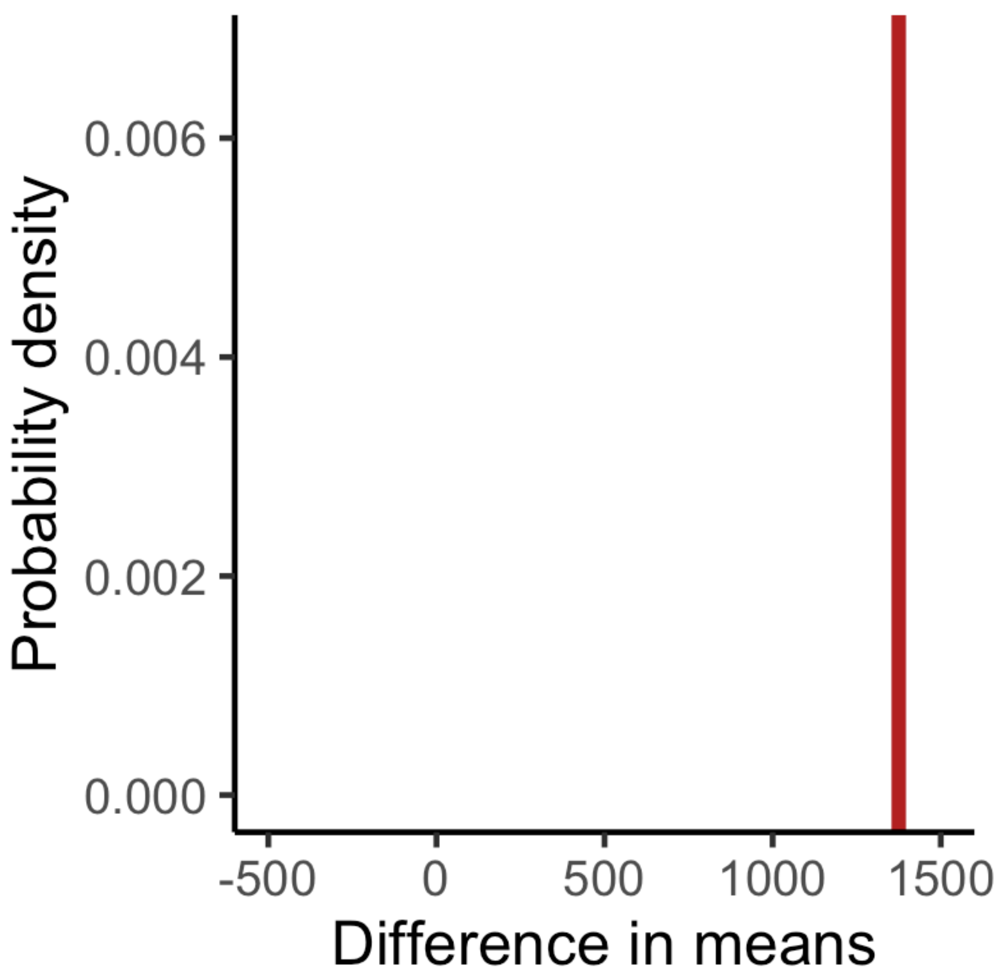
Sample



Sample statistic

Hypothesis testing in practice

Standard error of the *difference of means*



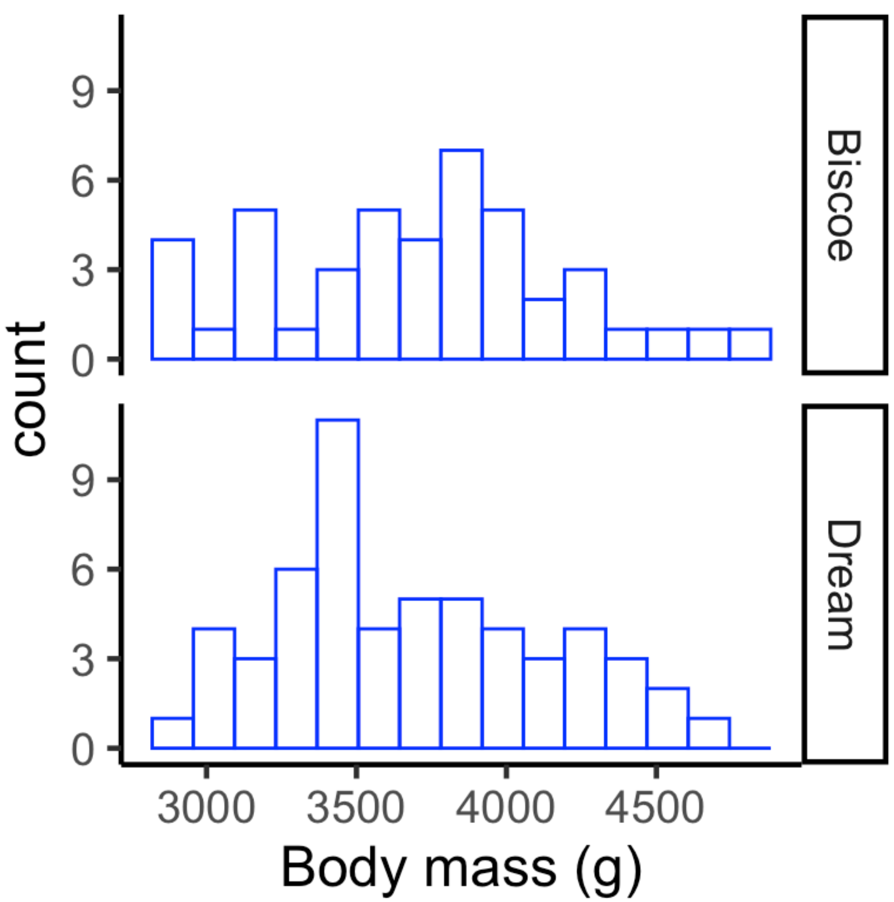
Hypothesis testing in practice

Standard error of the *difference of means*

1. Formulate your hypotheses
   H_0 = no effect, H_A = some effect
2. Calculate point estimate
   Difference in means, regression coefficient, etc
3. Quantify uncertainty in sampling distribution
   ~~*Shuffle data, recalculate point estimate, repeat*~~
   Approximate sampling distribution using standard error
4. Calculate p-value
   Probability of point estimate if null is true
   *2 * pnorm(-abs(observed), mean = 0, sd = se)*
5. Reject or fail to reject the null
   Is $p \leq \alpha$?

Hypothesis testing in practice

Your turn




```
adelie_biscoe <- with(penguins,
                      body_mass_g[species == "Adelie" &
                                  island == "Biscoe"])
adelie_dream <- with(penguins,
                     body_mass_g[species == "Adelie" &
                                 island == "Dream"])

obs_diff <- mean(adelie_biscoe - adelie_dream)
obs_diff <- mean(adelie_biscoe) - mean(adelie_dream)

se <- function(a, b) {
  a <- na.omit(a)
  b <- na.omit(b)
  sqrt(sd(a)^2 / length(a) + sd(b)^2 / length(b))
}

se_diff <- se(adelie_biscoe, adelie_dream)

pval <- 2 * pnorm(-abs(observed_difference),
                  mean = 0,
                  sd = se_difference)
pval <- 2 * pnorm(0,
                  mean = -abs(observed_difference),
                  sd = se_difference)
```

Hypothesis testing in practice

Your turn

1. Which `obs_diff` is the difference of the means?

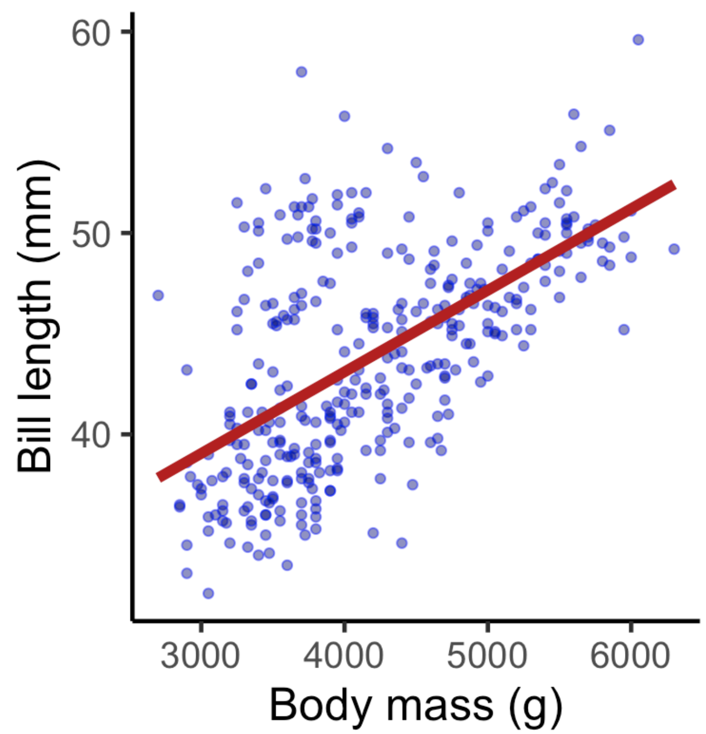
2. Which `pval` is the probability of the observed difference, if the null is true?

3. Sketch the null distribution of the sample statistic. Indicate the observed difference, the standard error, and the p-value.

Hypothesis testing in practice

Standard error of a *regression coefficient*

Population | **Sample** | **Sample coefficient**



Hypothesis testing in practice

Standard error of a *regression coefficient*

```
Call:
lm(formula = bill_length_mm ~ body_mass_g, data = penguins)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.690e+01  1.269e+00   21.19   <2e-16 ***
body_mass_g 4.051e-03  2.967e-04   13.65   <2e-16 ***
```

Hypothesis testing in practice

Recap

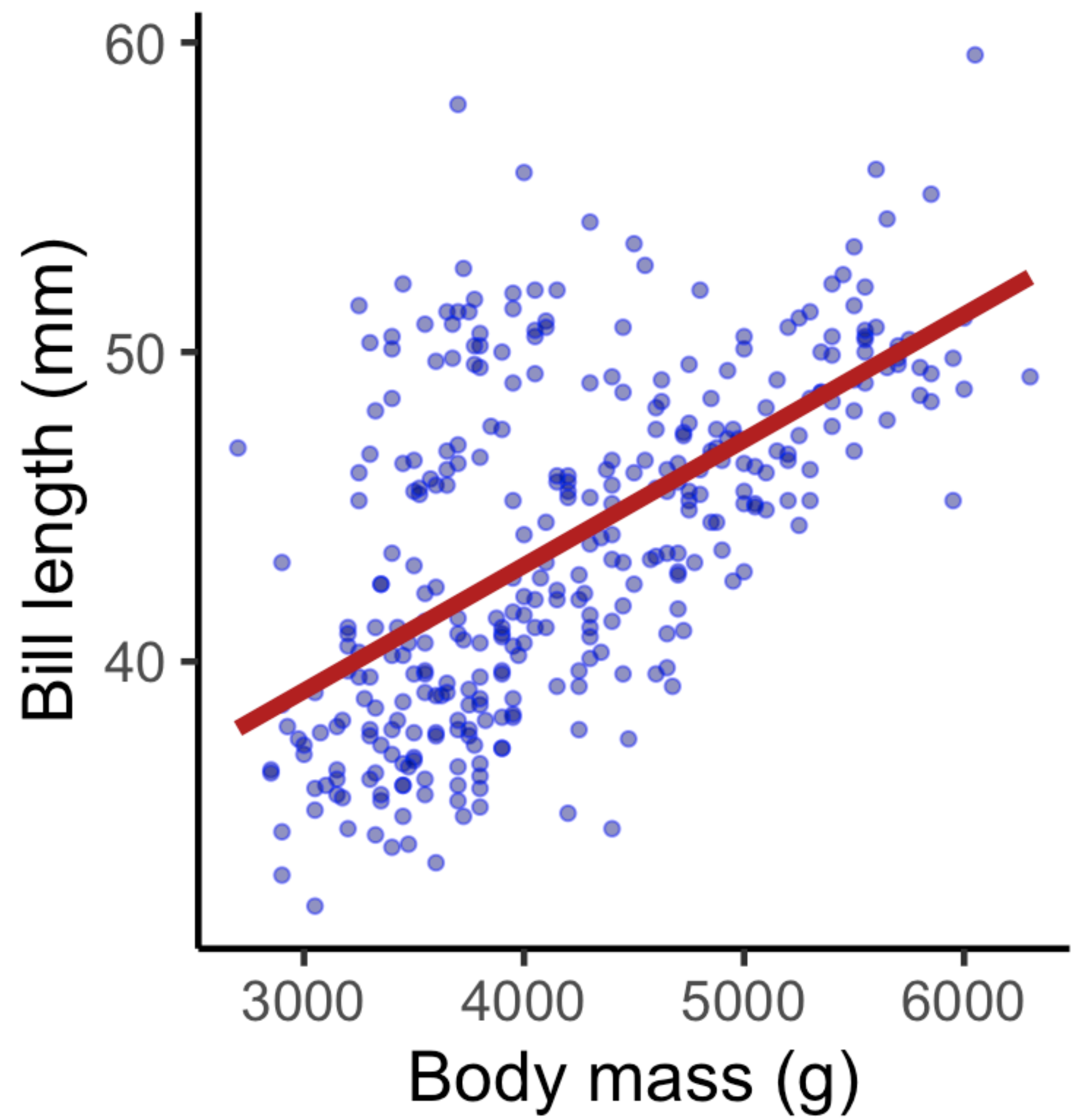
1. Sampling statistics are approximately normally distributed
2. From the central limit theorem, we can get the standard error of the sampling distribution *from just one sample*
3. R will tell you the point estimate and the standard error when you fit a model
4. The p-value is the probability of getting a point estimate that many standard errors away from 0

Confidence intervals

Motivation

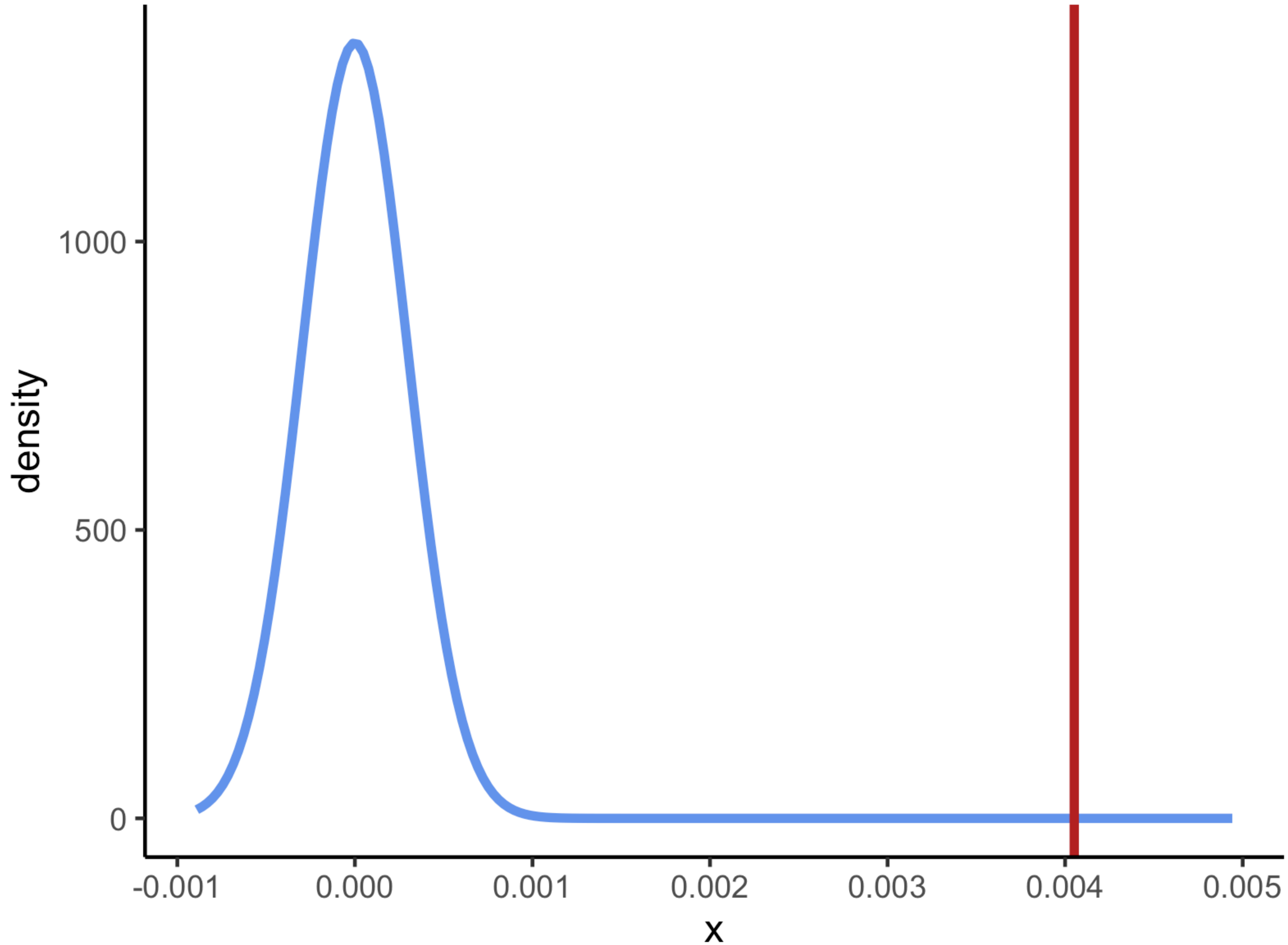
$$\hat{\beta}_1 = 0.0041$$

$$\beta_1 \in (???, ???)$$



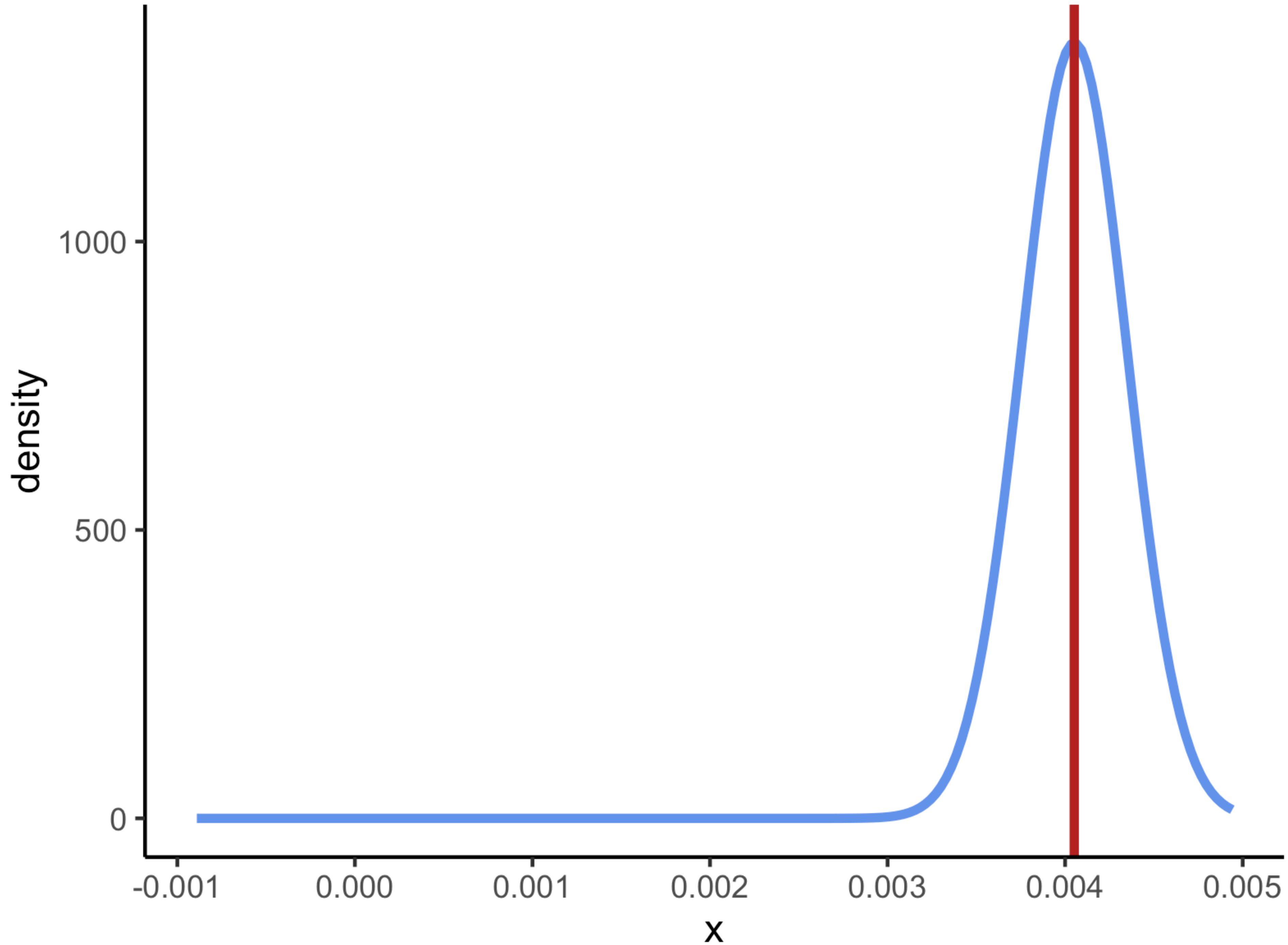
Confidence intervals

Recycling standard errors



Confidence intervals

Recycling standard errors



Confidence intervals

Interpretation

Choose the correct interpretation of the confidence interval:

“We are 95% confident the true coefficient is between 0.0035 and 0.0046.”

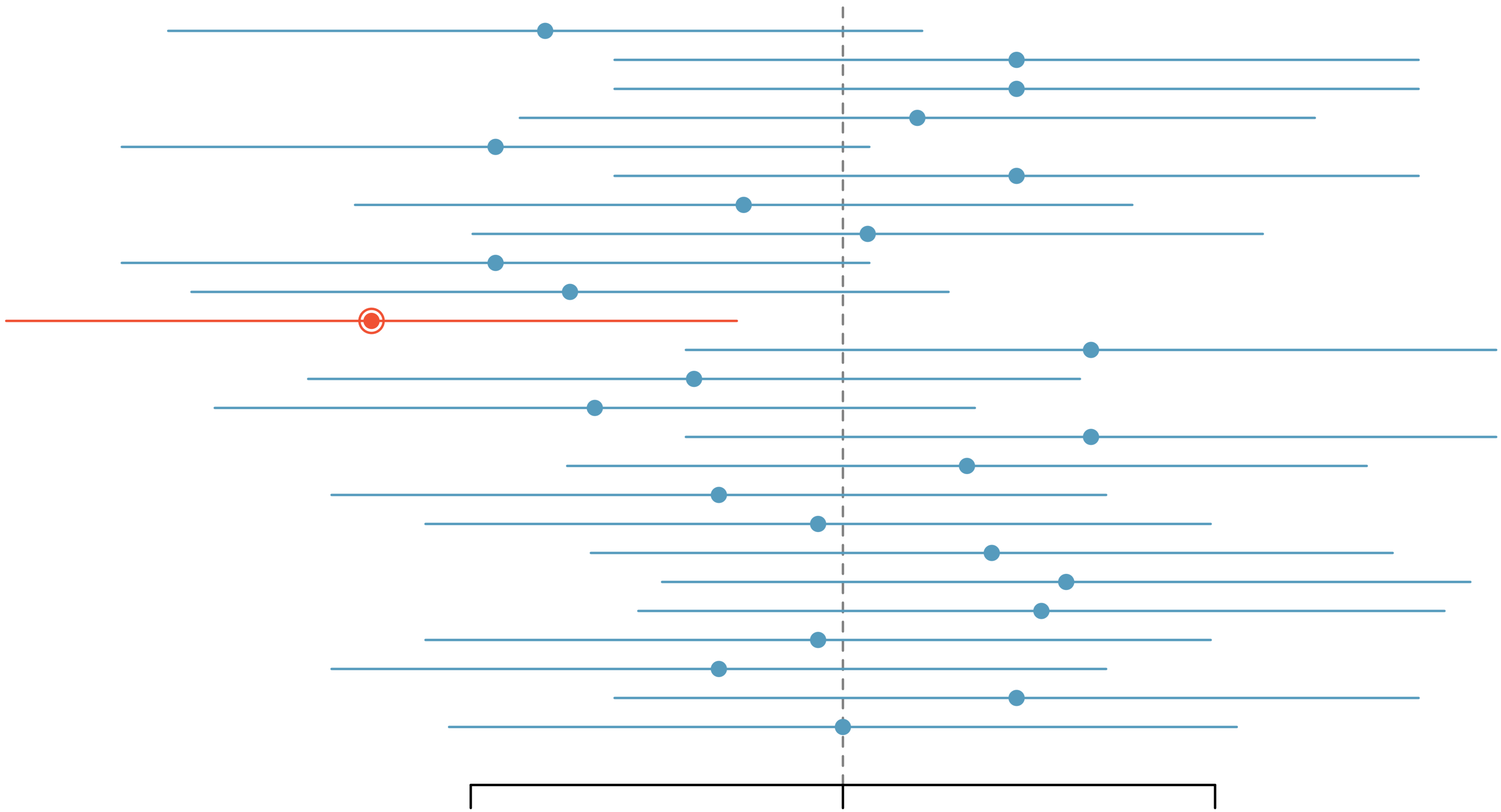
A. We are 95% confident the true coefficient falls in this range.

B. The true coefficient will fall in this range 95% of the time.

C. This range has a 95% probability of containing the true coefficient.

Confidence intervals

Interpretation



Confidence intervals

Recap

1. We know point estimates aren't perfect - confidence intervals provide a useful bounds.
2. Use the standard error again, but center the distribution on the point estimate.
3. Be careful with interpretation! "Confidence" refers to the procedure, not to the probability the CI contains the population parameter.

Summary

Could our sample statistic point estimate be explained just by randomness?

- **Hypotheses**
  - H_0 no effect. H_A some effect.
  - If the point estimate is improbable under the null hypothesis, reject the null. Otherwise, fail to reject.
- **Two methods for estimating null distribution**
  - Randomization.
  - Normal approximation.
- **Confidence intervals**
  - An interval that we are confident contains the population parameter.