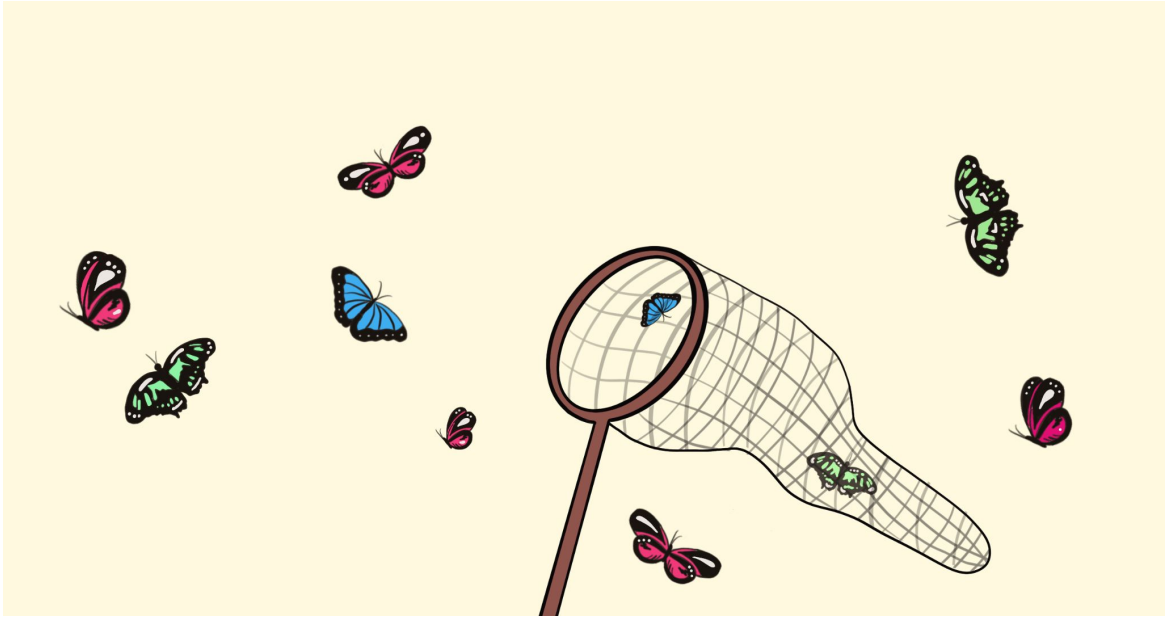
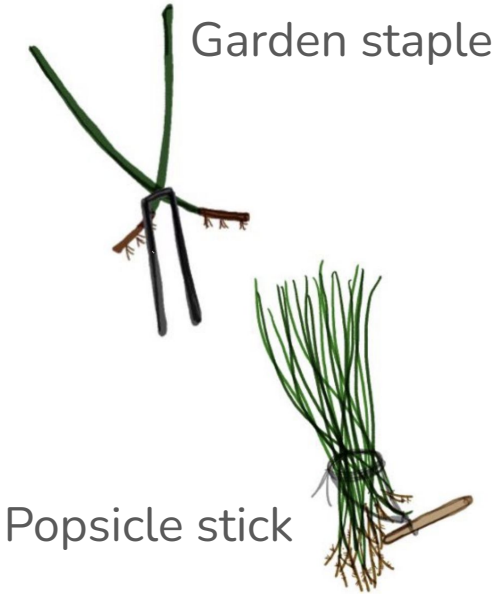
Week 4 Lecture 2:
Inference with mathematical models

EDS 222: *Statistics for Environmental Data Science*



Eelgrass restoration

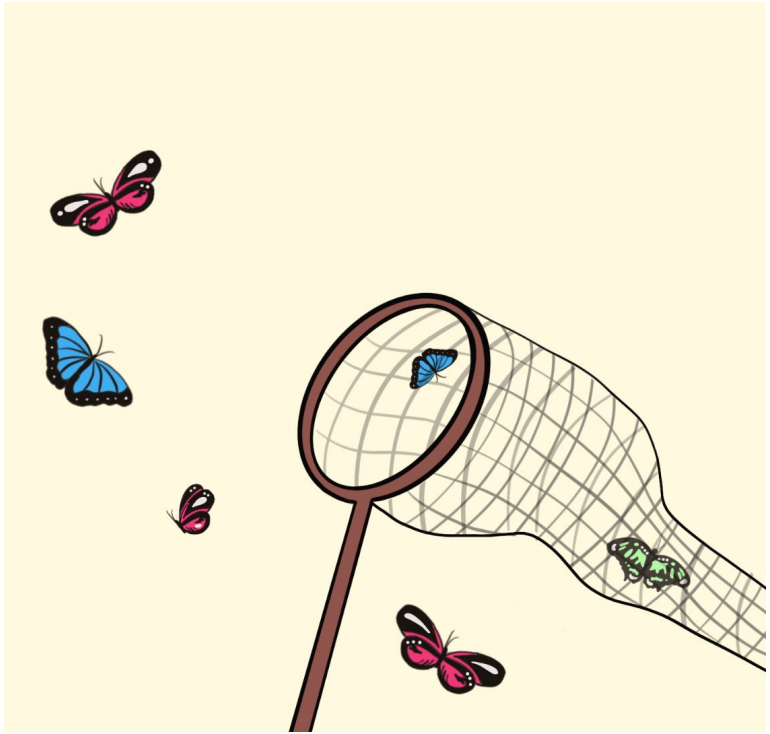
Photo: Hakai Institute



Artwork: Kat Beheshti

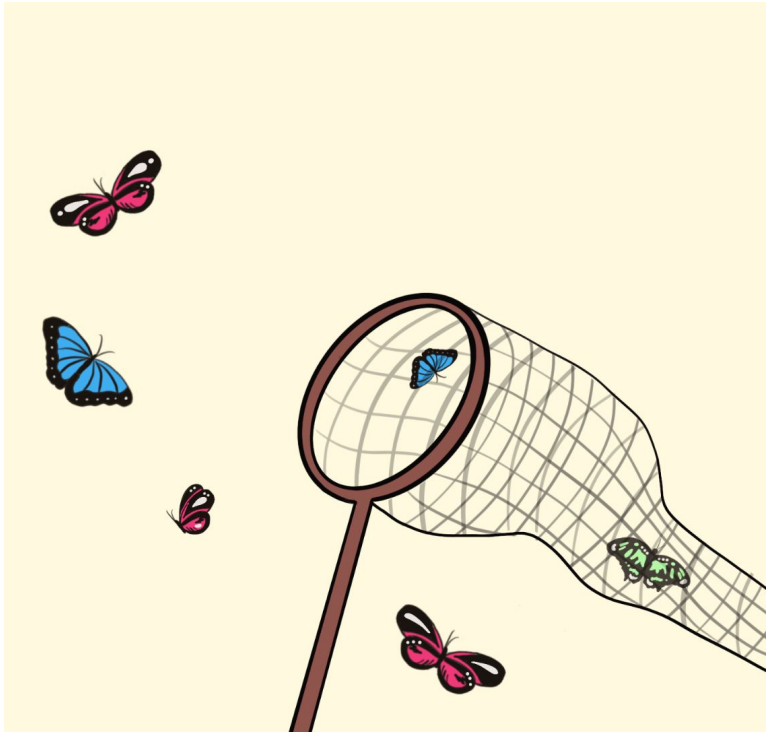
Today’s agenda

- ➔ Normal approximations of sampling distributions
- ➔ Building confidence intervals
- ➔ Estimating p-values

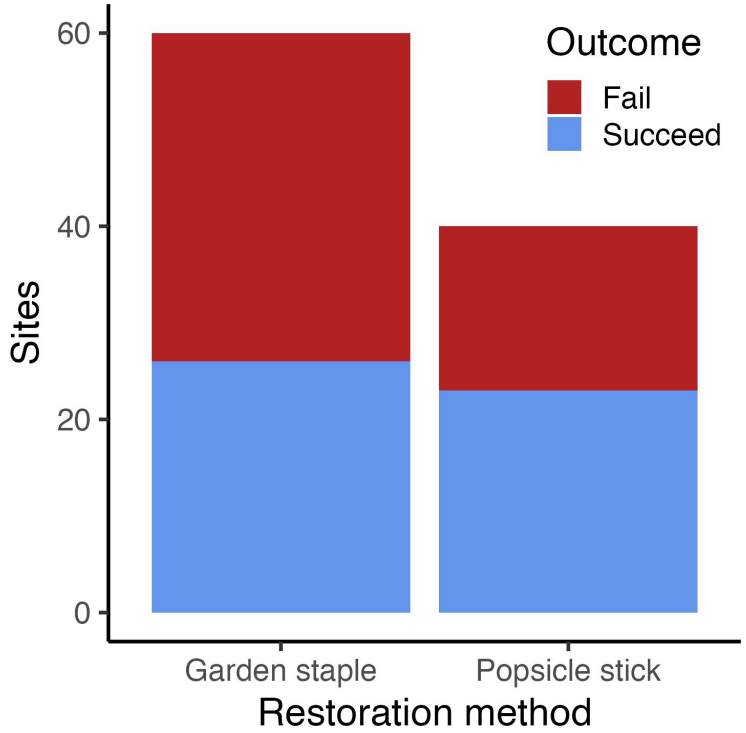


Today’s agenda

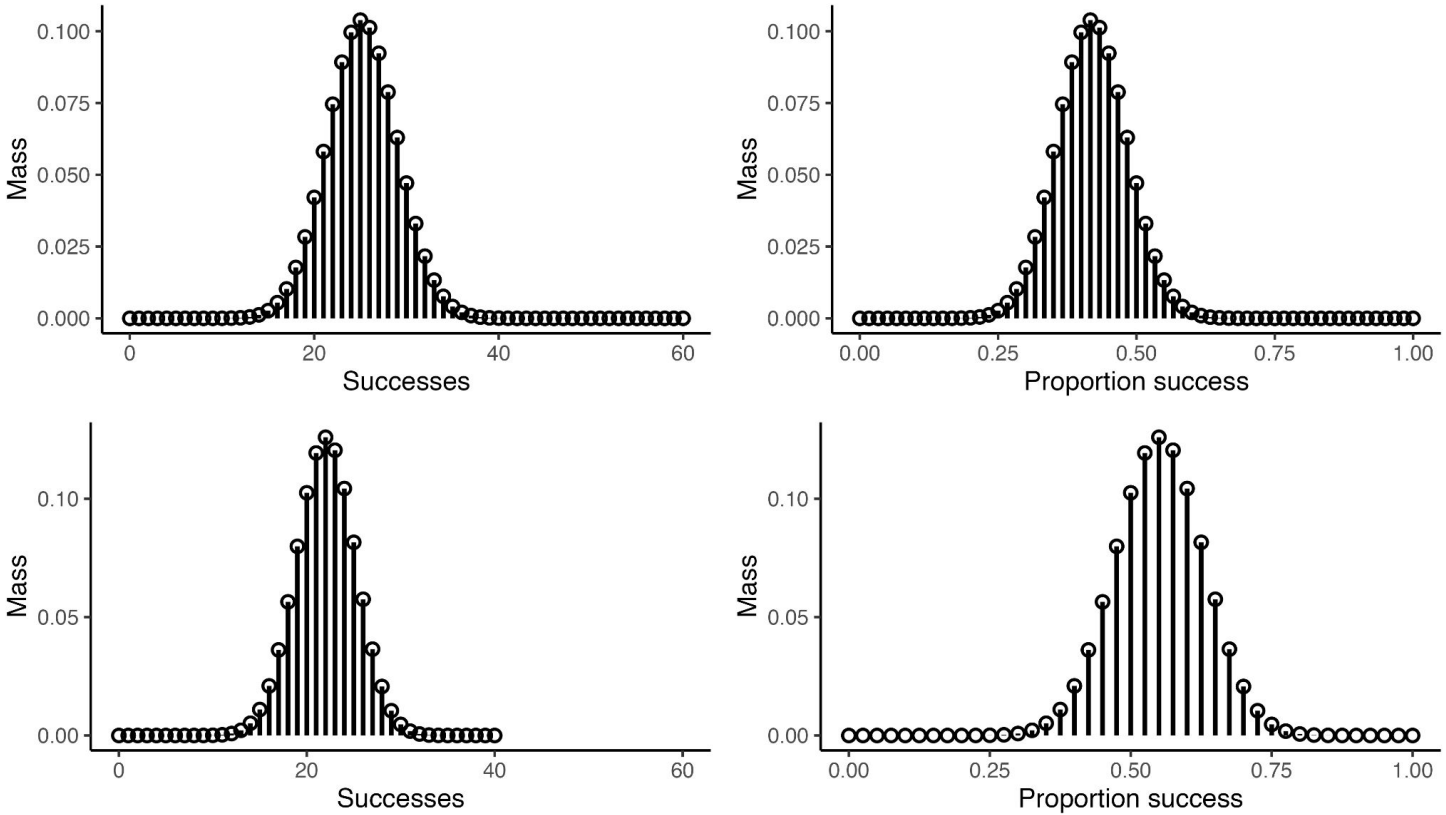
- ➔ **Normal approximations of sampling distributions**
- ➔ Building confidence intervals
- ➔ Estimating p-values



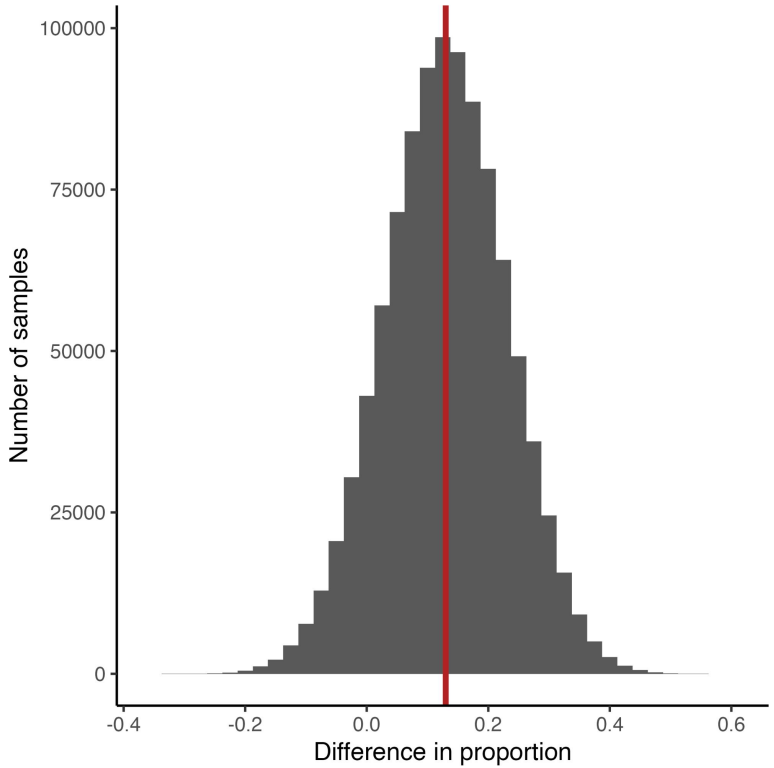
Data distribution



Population distribution



Sampling distribution

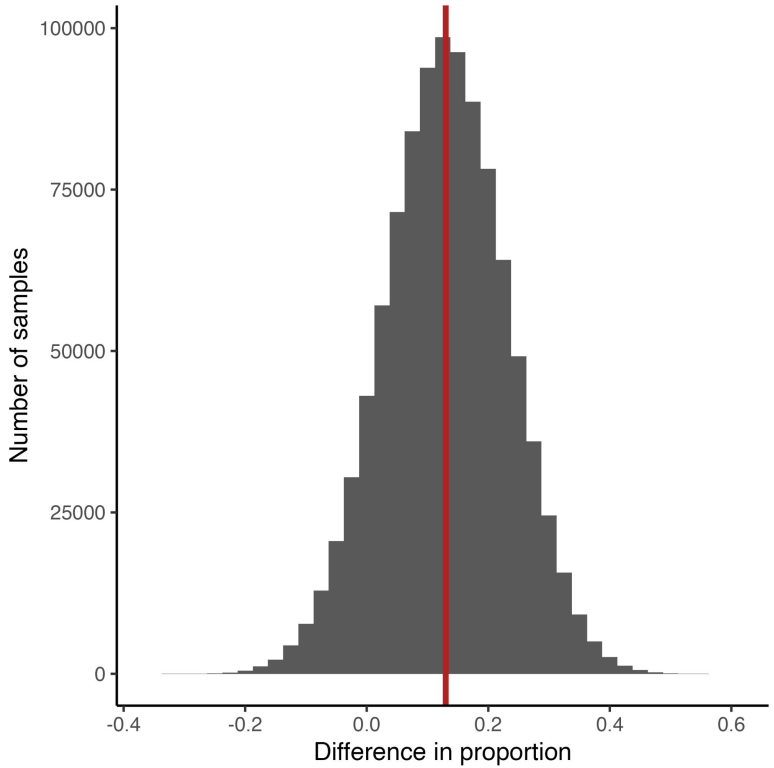


A few more samples

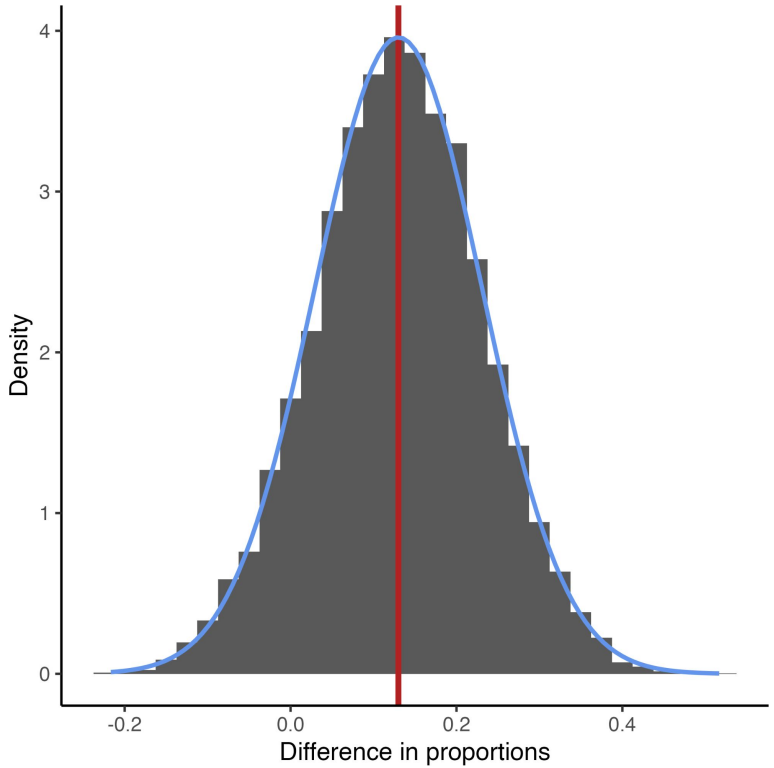
```
# Repeat 10 times
sim_spl <- function(i) {
  # Simulate samples
  sample1 <- rbinom(40, size = 1, prob = p1)
  sample2 <- rbinom(60, size = 1, prob = p2)

  # Calculate test statistic
  delta_p_spl <- mean(sample2) - mean(sample1)

  delta_p_spl
}
spl_10 <- map_dbl(1:10, sim_spl)
round(spl_10, 2)
```

```
[1]  0.28  0.24 -0.05
0.17  0.28  0.08  0.25
0.07  0.20  0.07
```

How do we estimate this?



How do we estimate this?

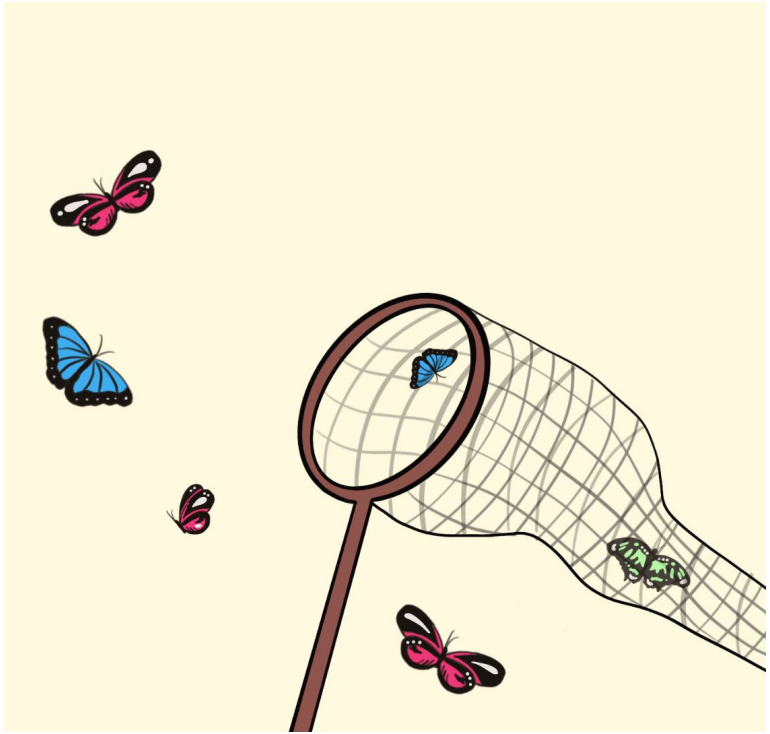


When does this work?

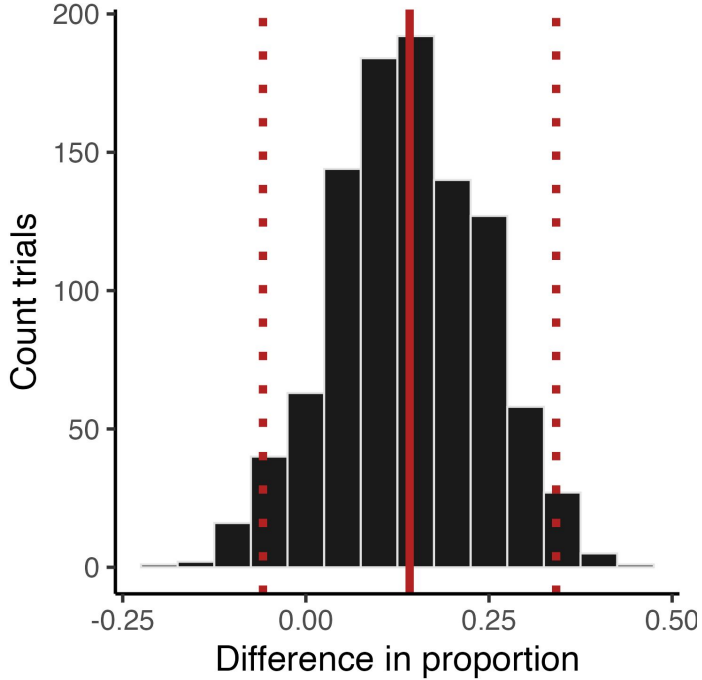
Normal approximations of sampling distributions

Today’s agenda

- ➔ Normal approximations of sampling distributions
- ➔ **Building confidence intervals**
- ➔ Estimating p-values



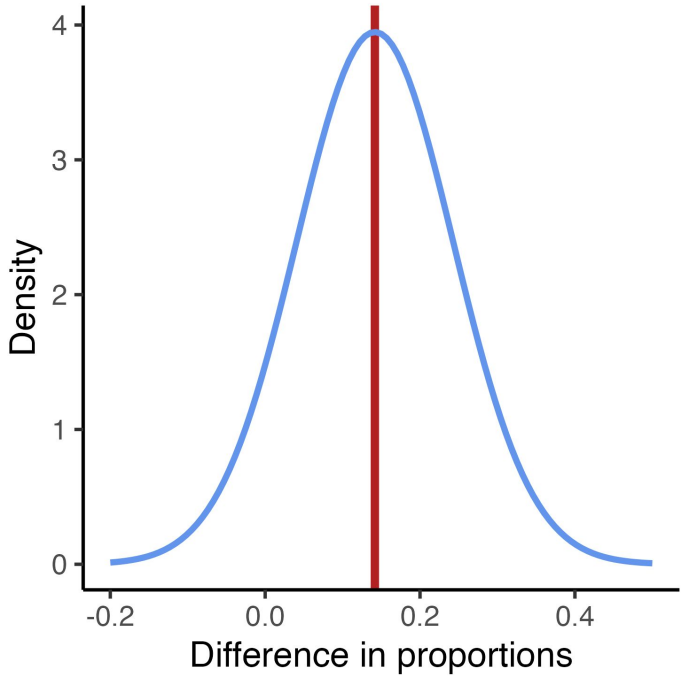
Recall bootstrapping



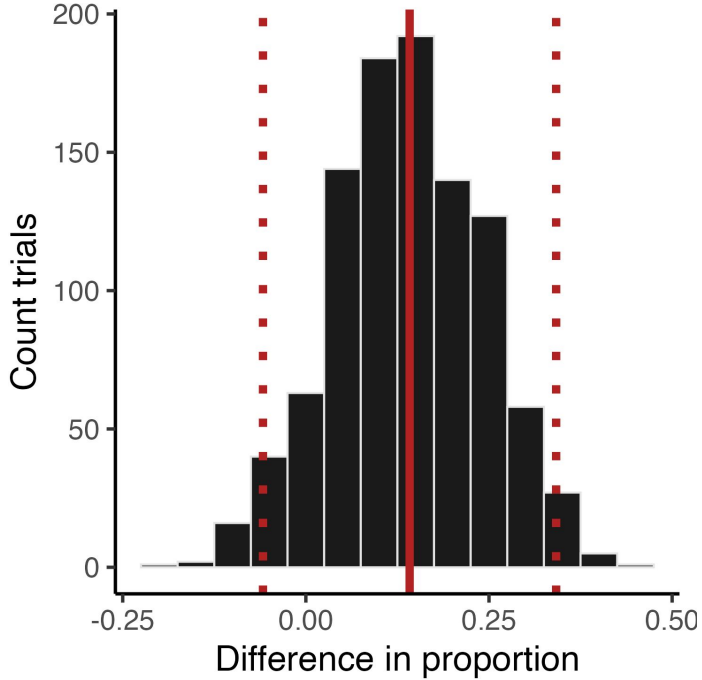
SE(difference in proportions)

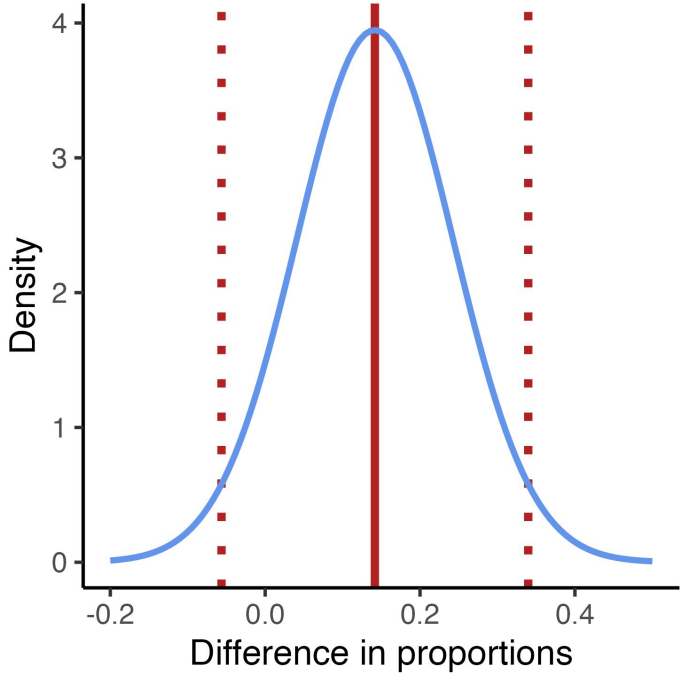
$$\hat{p}_2 - \hat{p}_1 \sim Normal\left(\mu, \sigma\right)$$
$$\mu = p_2 - p_1$$
$$\sigma = SE\left(\hat{p}_2 - \hat{p}_1\right)$$

$$SE\left(\hat{p}_2 - \hat{p}_1\right) = \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}$$



Confidence interval

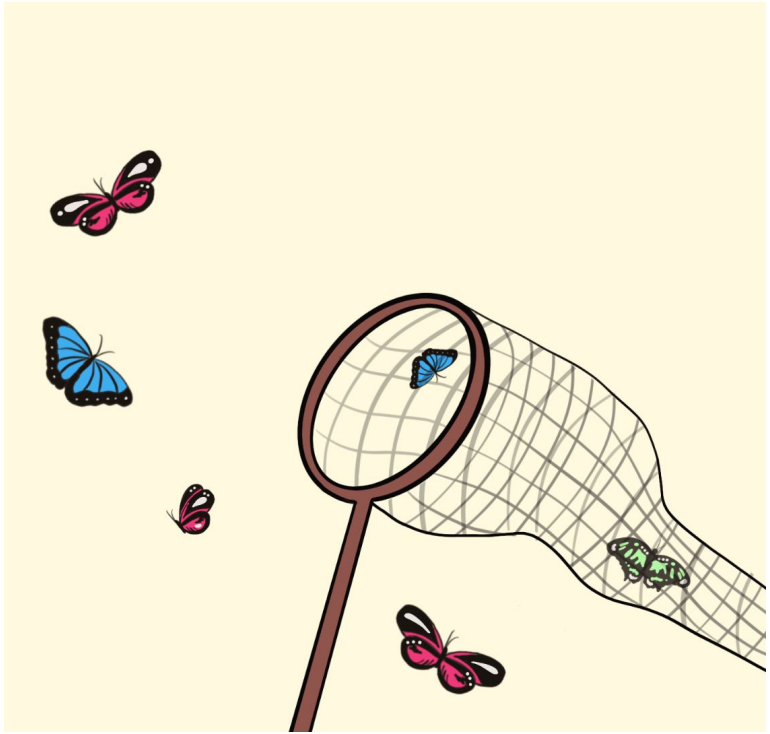




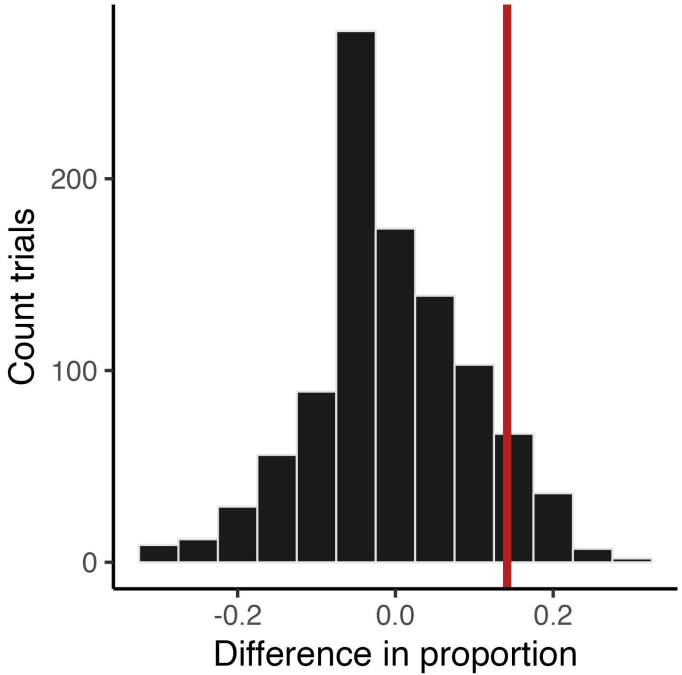
Building confidence intervals

Today’s agenda

- ➔ Normal approximations of sampling distributions
- ➔ Building confidence intervals
- ➔ **Estimating p-values**



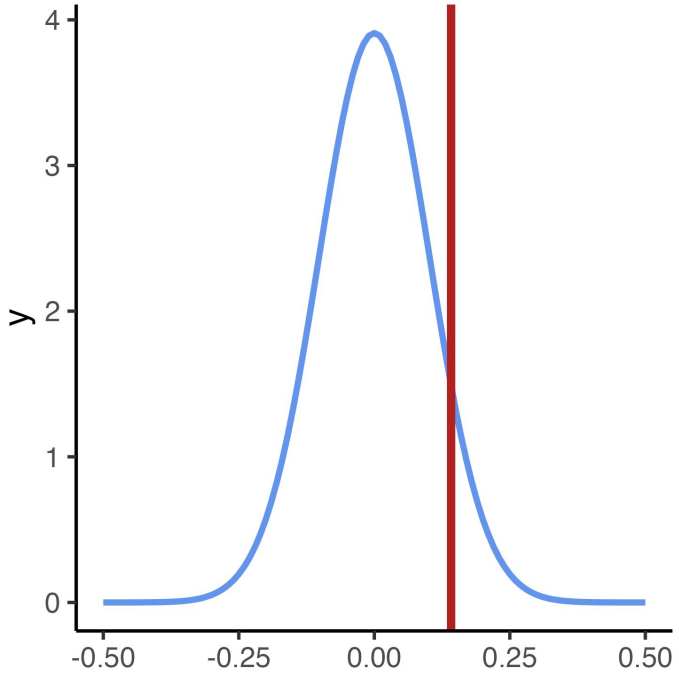
Recall permutation



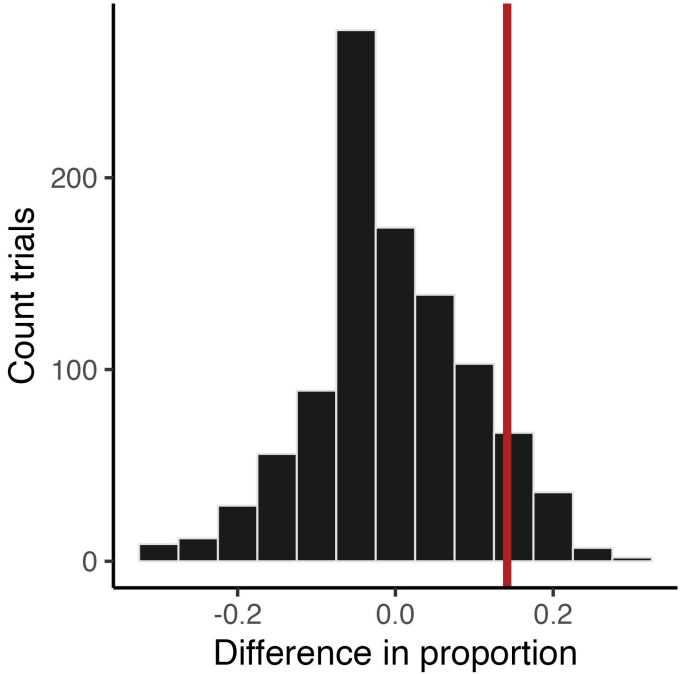
Null hypothesis

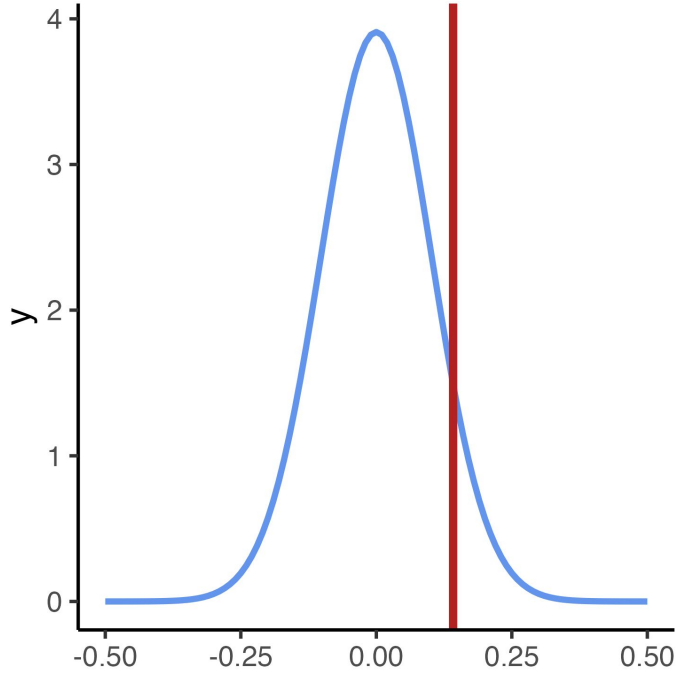
$$
\begin{aligned}
\hat{p}_2 - \hat{p}_1 &\sim Normal\left(\mu, \sigma\right) \\
\mu &= 0 \\
\sigma &= SE\left(\hat{p}\right) \\
\hat{p} &= \frac{x_1 + x_2}{n_1 + n_2}
\end{aligned}
$$

$$
SE\left(\hat{p}\right) = \sqrt{\hat{p}\left(1 - \hat{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}
$$



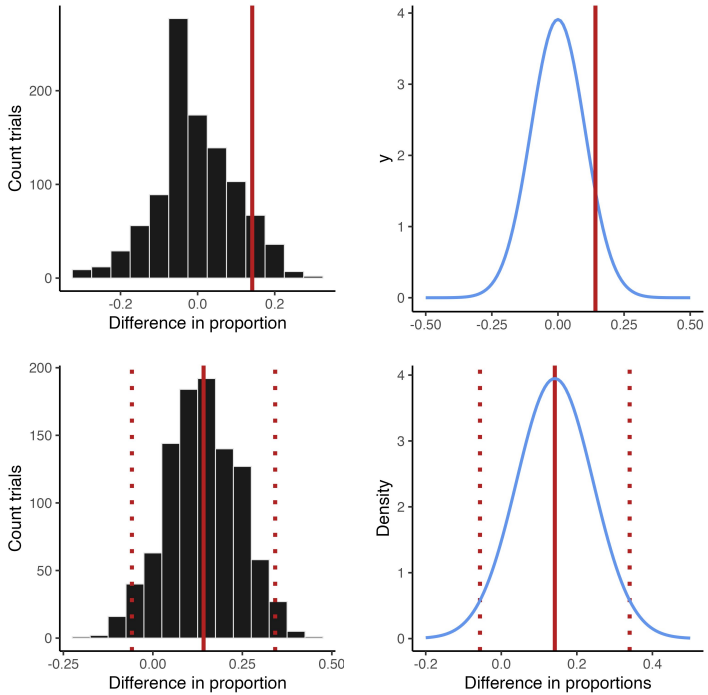
P-value


Count trials
200
100
0
-0.2
0.0
0.2
Difference in proportion


y
4
3
2
1
0
-0.50
-0.25
0.00
0.25
0.50

Estimating p-values

Recap


Count trials
200
100
0
-0.2
0.0
0.2
Difference in proportion
y
4
3
2
1
0
-0.50
-0.25
0.00
0.25
0.50
Count trials
200
150
100
50
0
-0.25
0.00
0.25
0.50
Difference in proportion
Density
4
3
2
1
0
-0.2
0.0
0.2
0.4
Difference in proportions