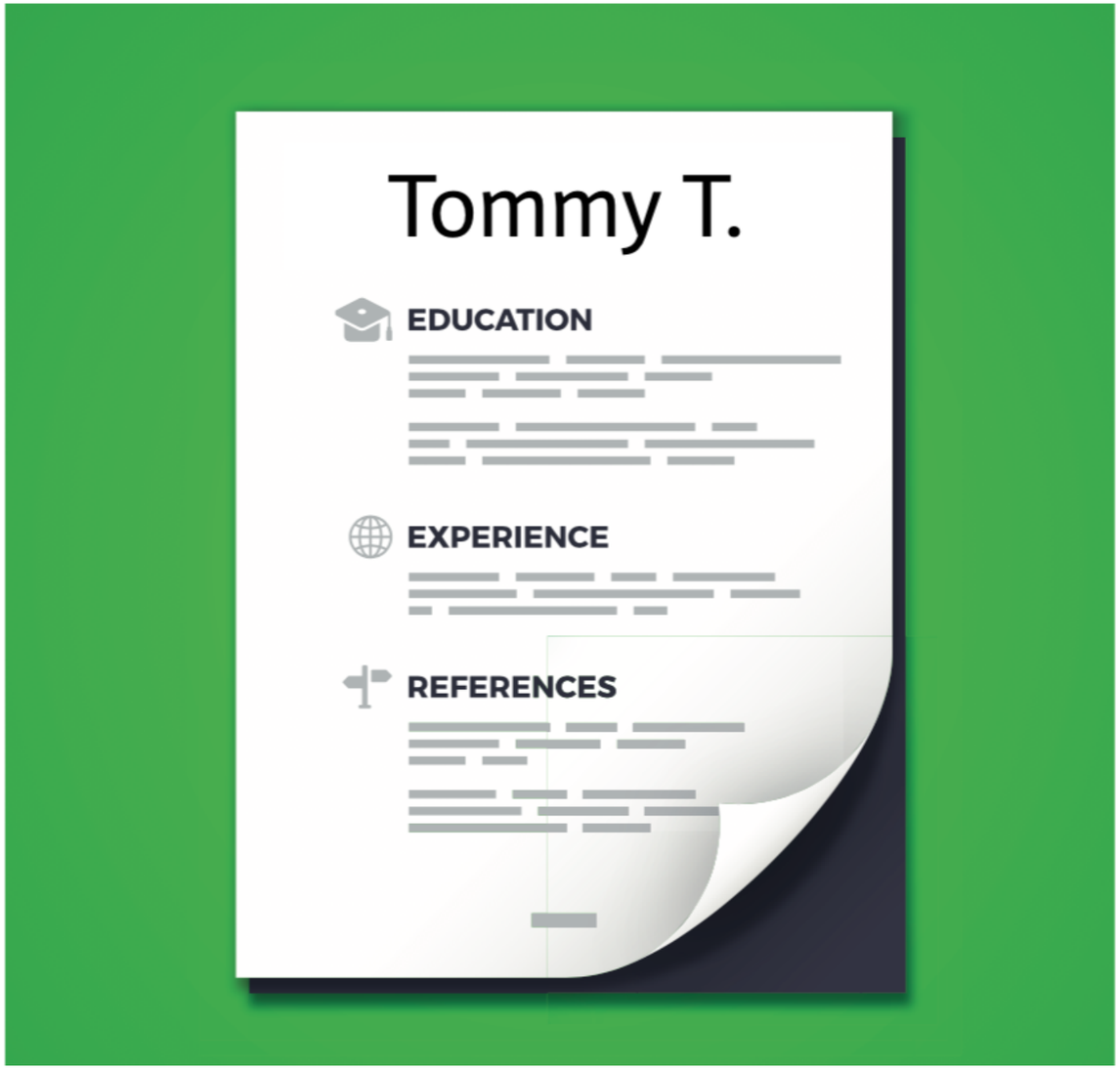
Example: gender discrimination

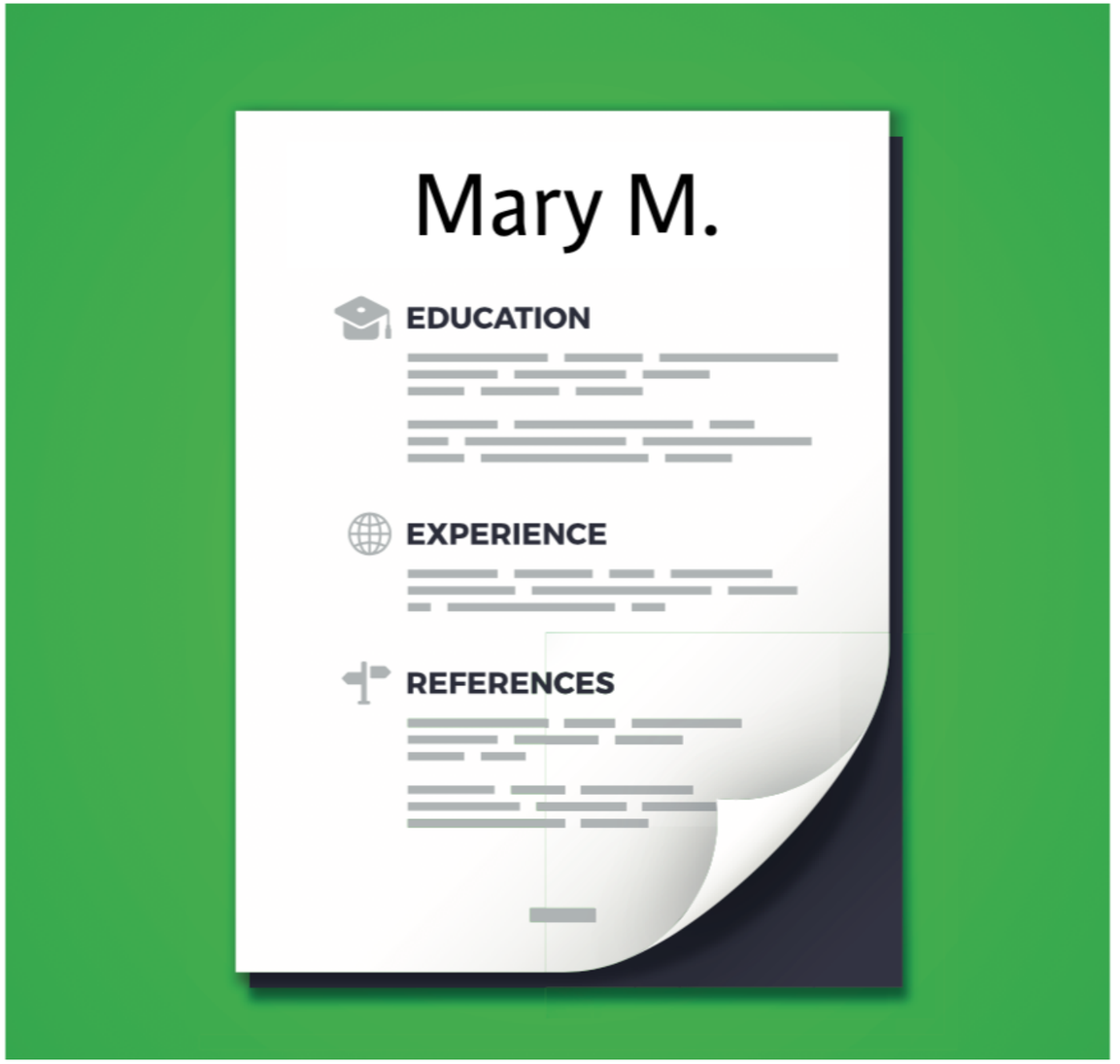
FOUNDATIONS OF INFERENCE

Jo Hardin
Instructor

datacamp

[1] Rosen B and Jerdee T. 1974. Influence of sex role stereotypes on personnel decisions. Journal of Applied Psychology





[1] Rosen B and Jerdee T. 1974. Influence of sex role stereotypes on personnel decisions. Journal of Applied Psychology

The data

	Promoted	Not promoted	Total
Male	21	3	24
Female	14	10	24
Total	35	13	48

Fewer women were promoted

```
disc <- data.frame(
    promote = c(rep("promoted", 35), rep("not_promoted", 13)),
    sex = c(rep("male", 21), rep("female", 14),
    rep("male", 3), rep("female", 10)))

disc %>%
    group_by(sex) %>%
    summarize(promoted_prop = mean(promote == "promoted"))
```

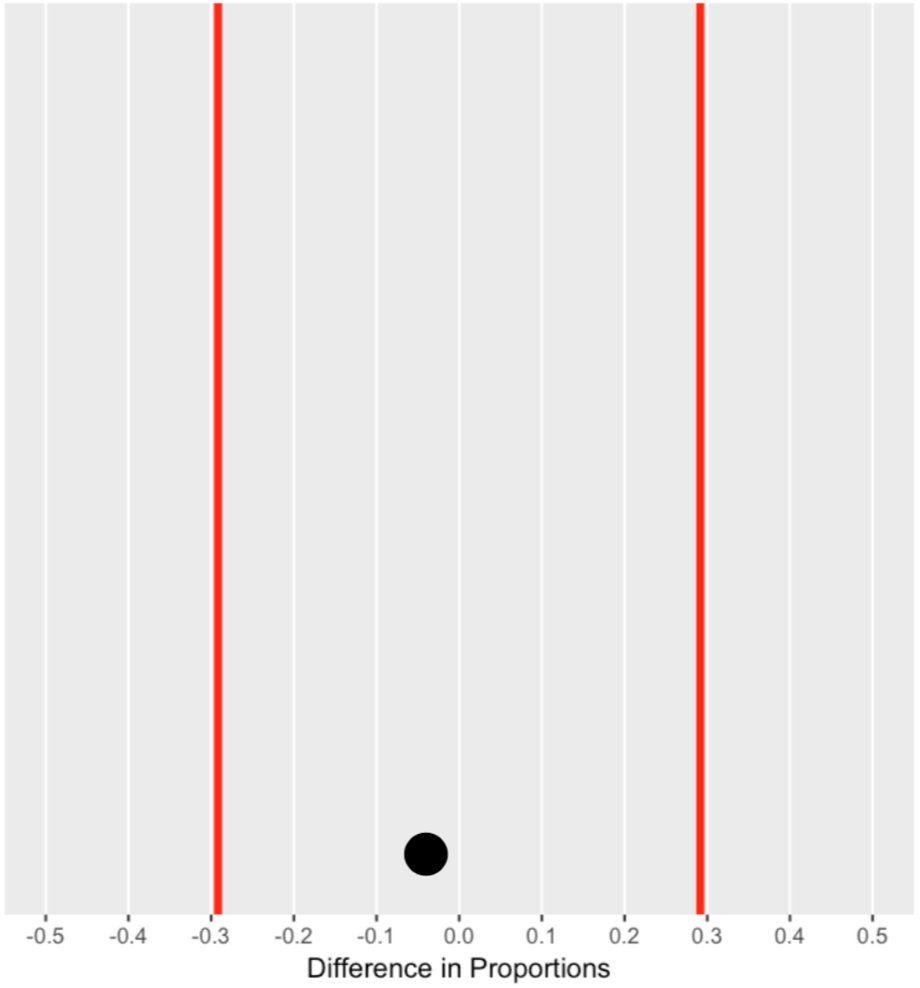
```
# A tibble: 2 × 2
     sex promoted_prop
  <fctr>         <dbl>
1 female     0.5833333
2   male     0.8750000
```

Random chance?

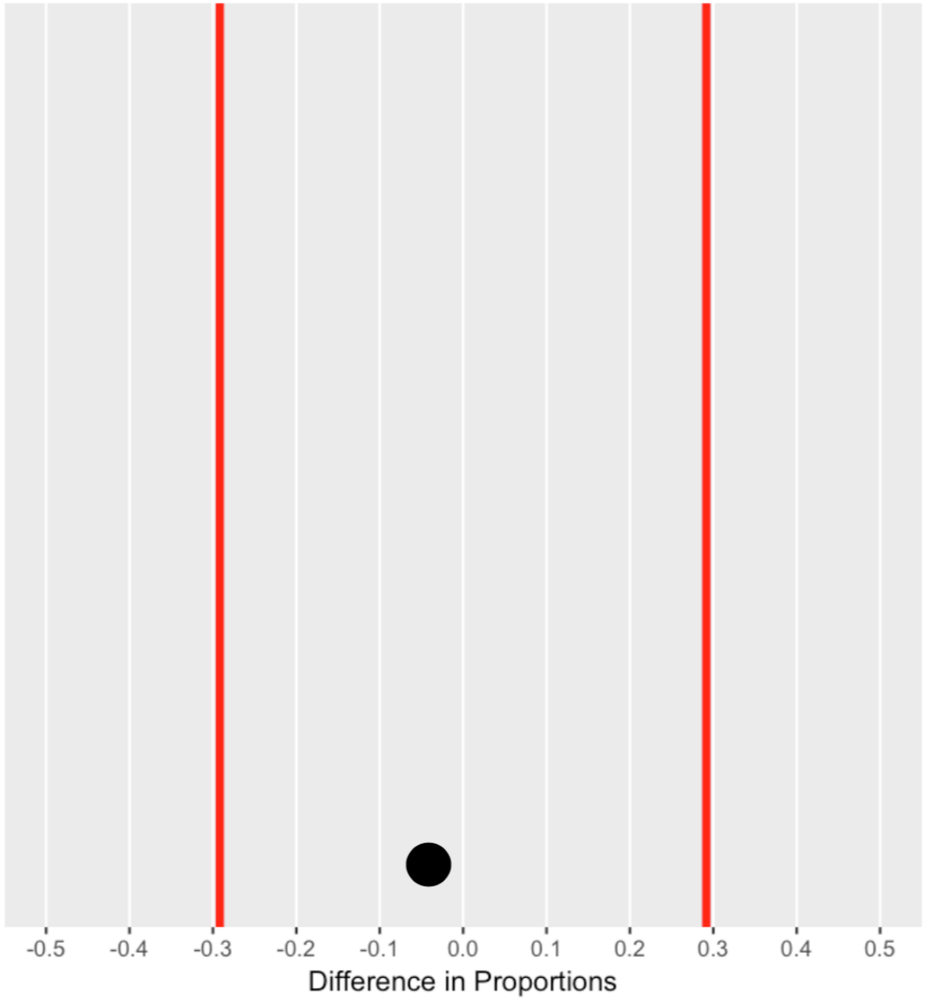
First shuffle

	Promoted	Not promoted
Male	18	6
Female	17	7

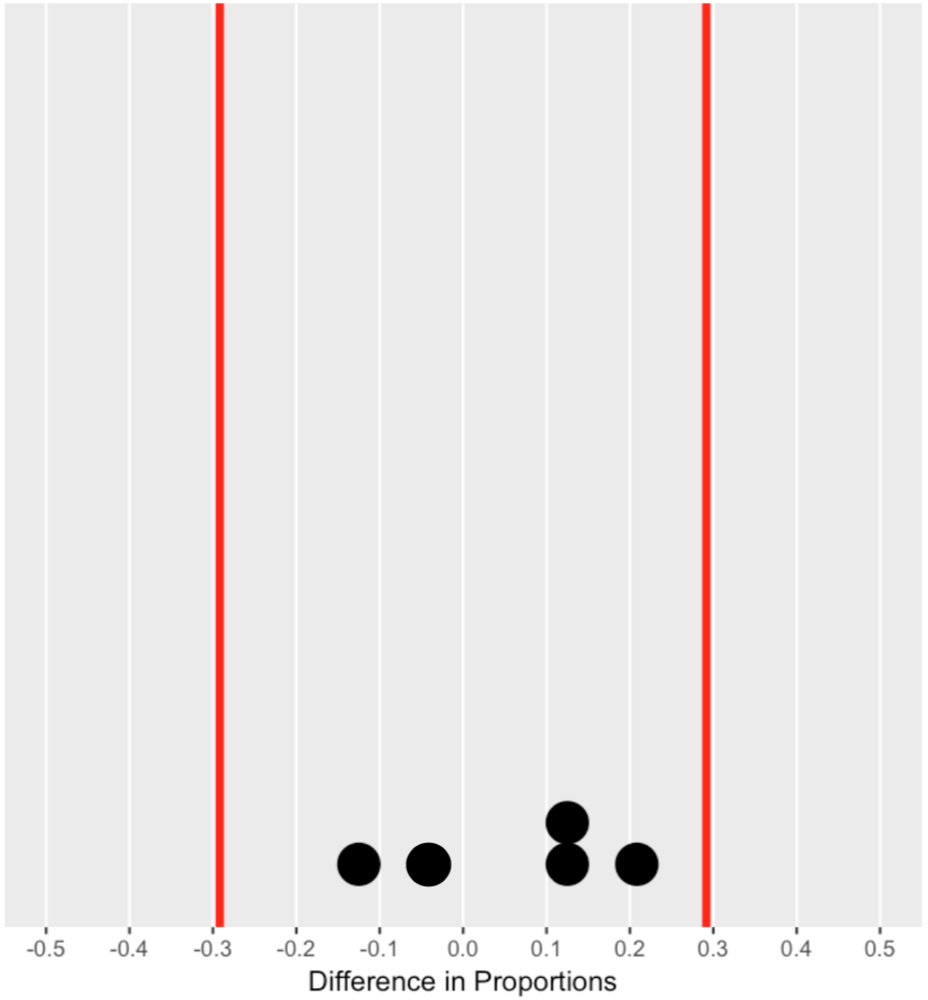
- 24 resumes given for male/females
- 35 allowed promotions



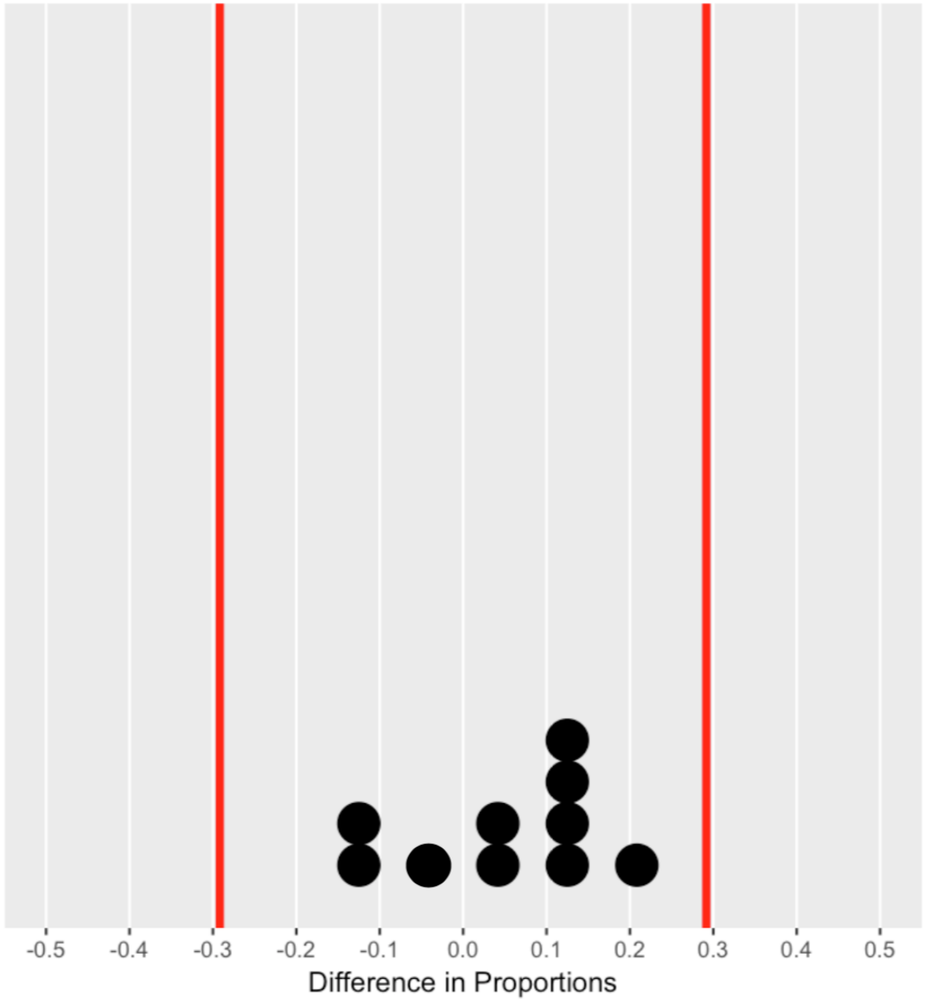
Random chance?



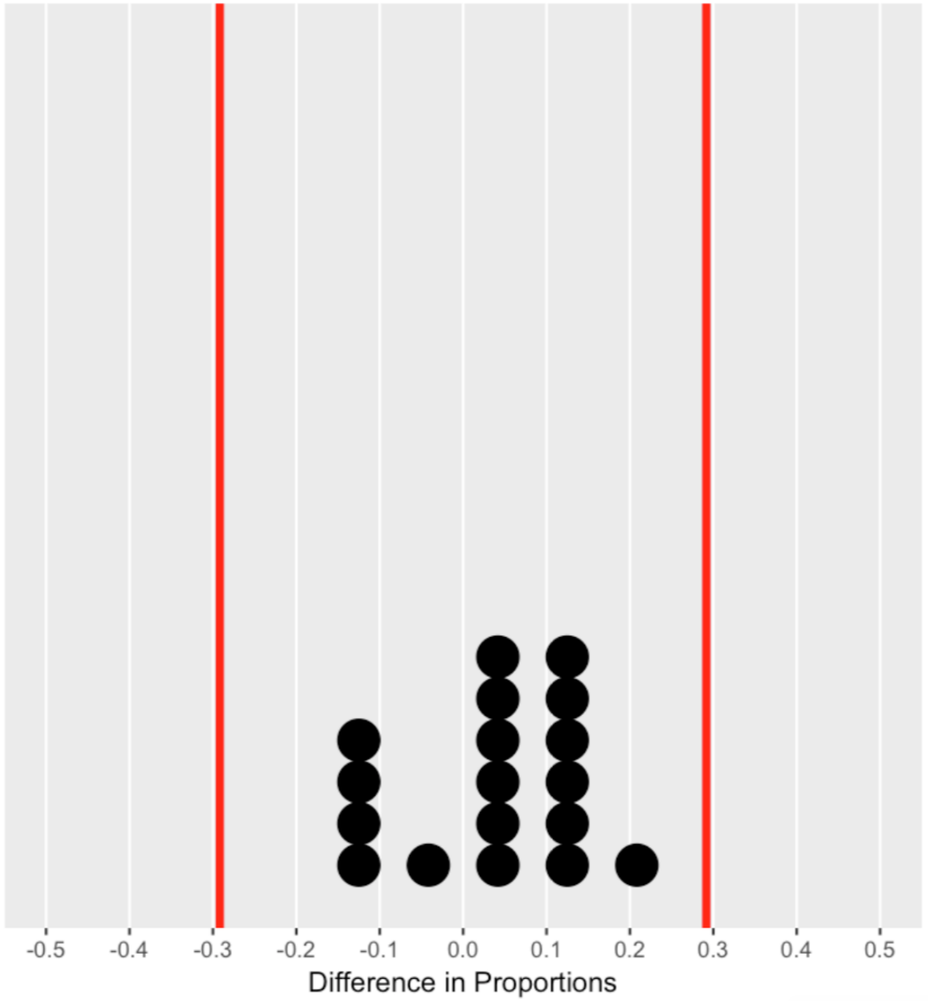
Random chance?



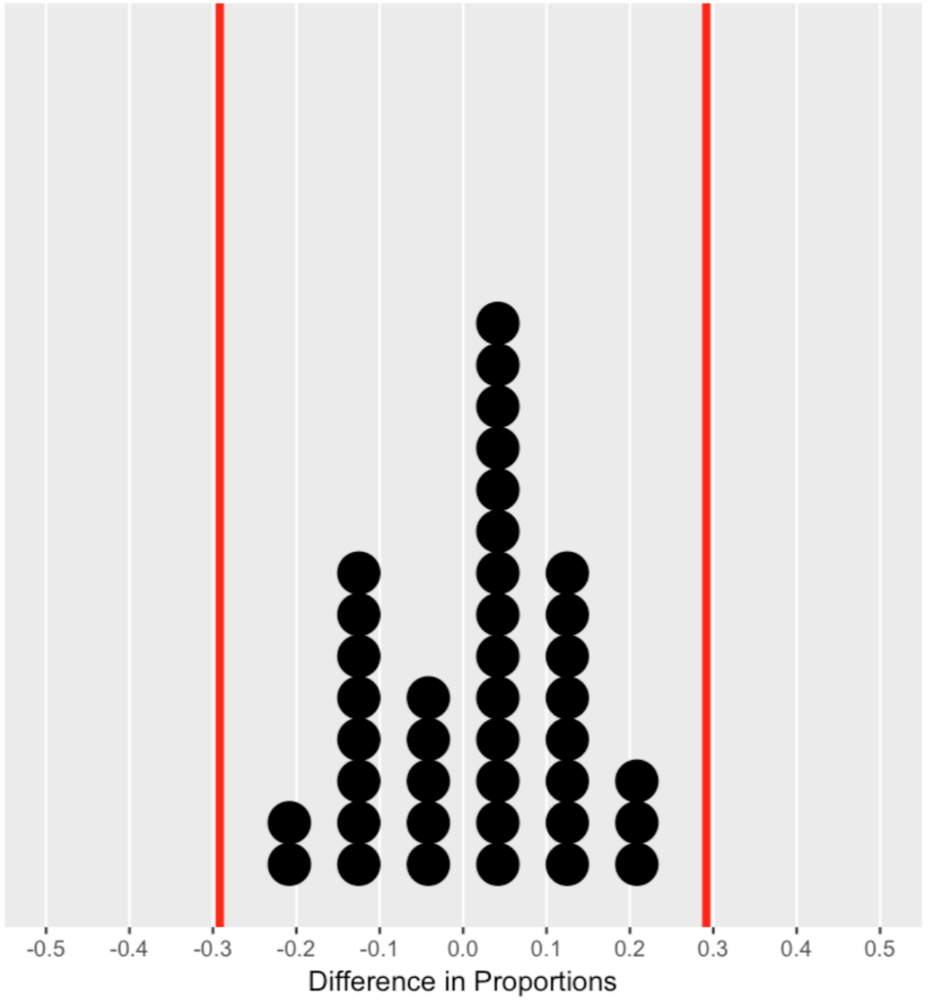
Random chance?



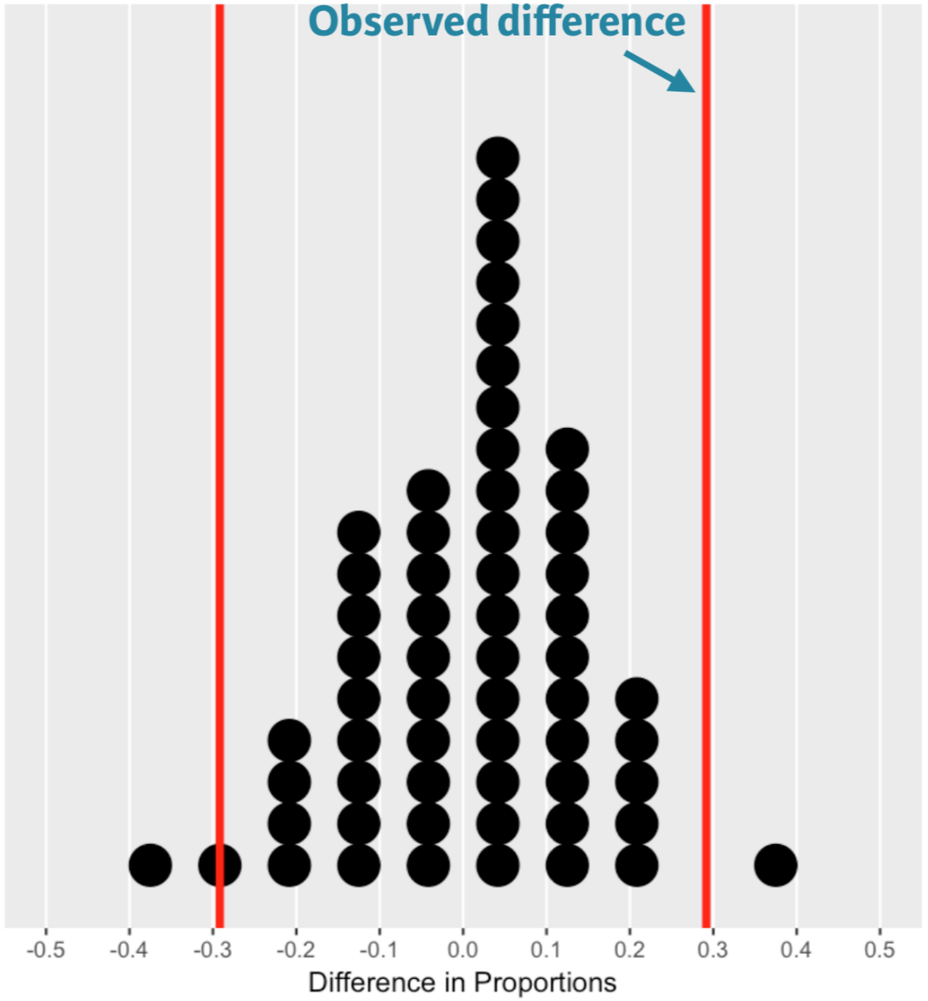
Random chance?



Random chance?



Random chance?



Let's practice!

FOUNDATIONS OF INFERENCE

Distribution of statistics

FOUNDATIONS OF INFERENCE

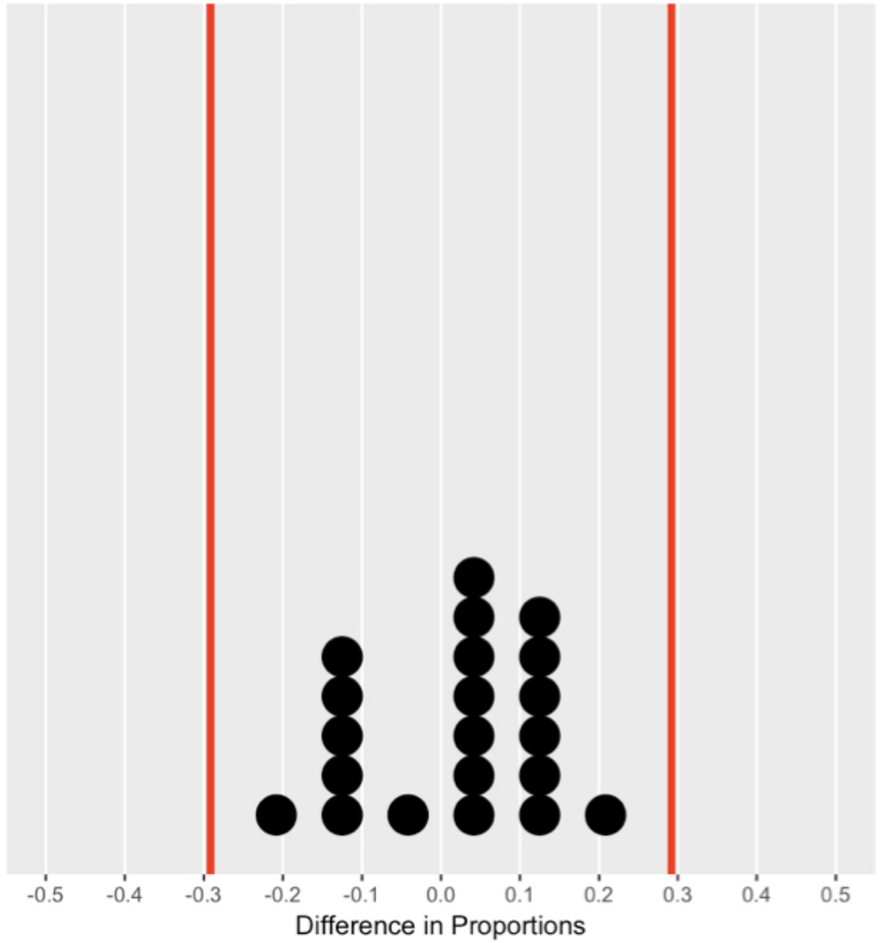
Jo Hardin
Instructor

datacamp

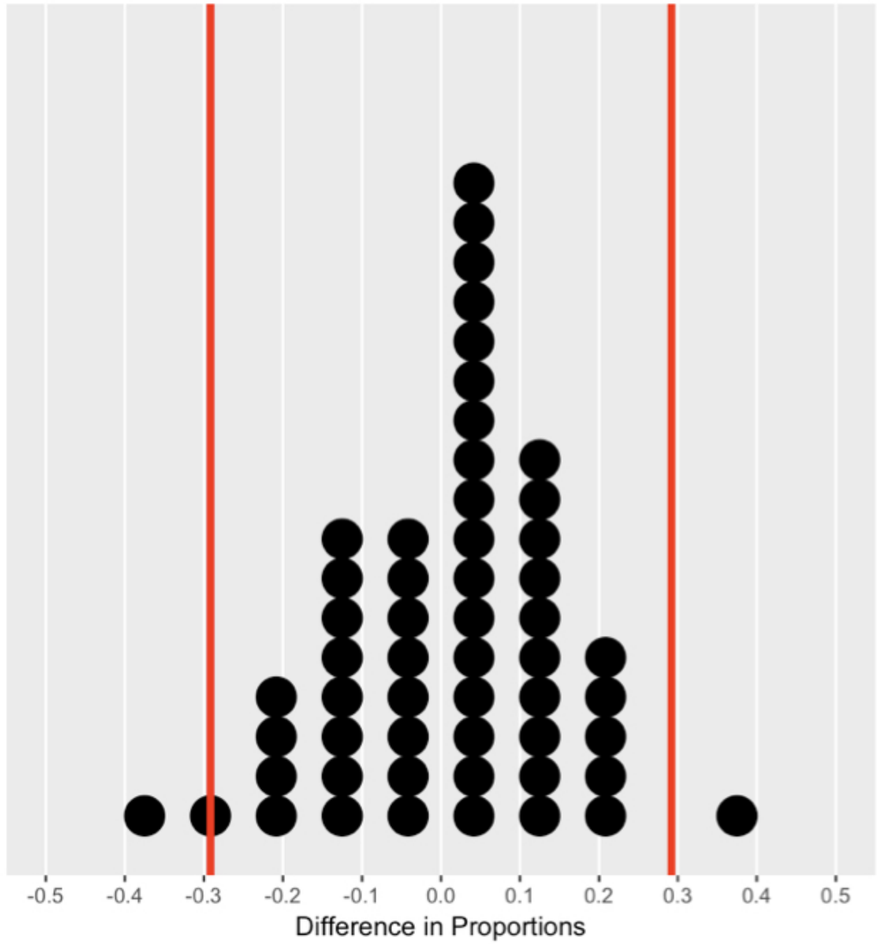
Null statistic

- Difference in proportions: $\hat{p} - p$
- Ratio: $\frac{\hat{p}}{p}$
- Interested in whether observed statistic is different from values obtained by shuffling

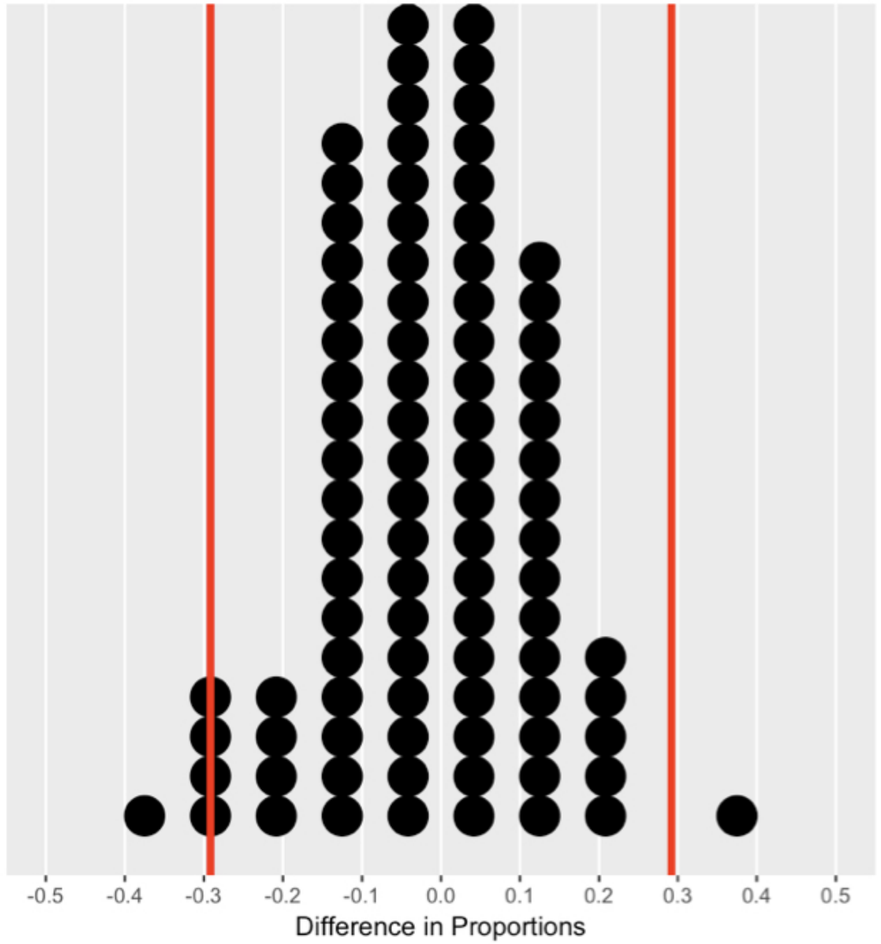
Calculating quantiles



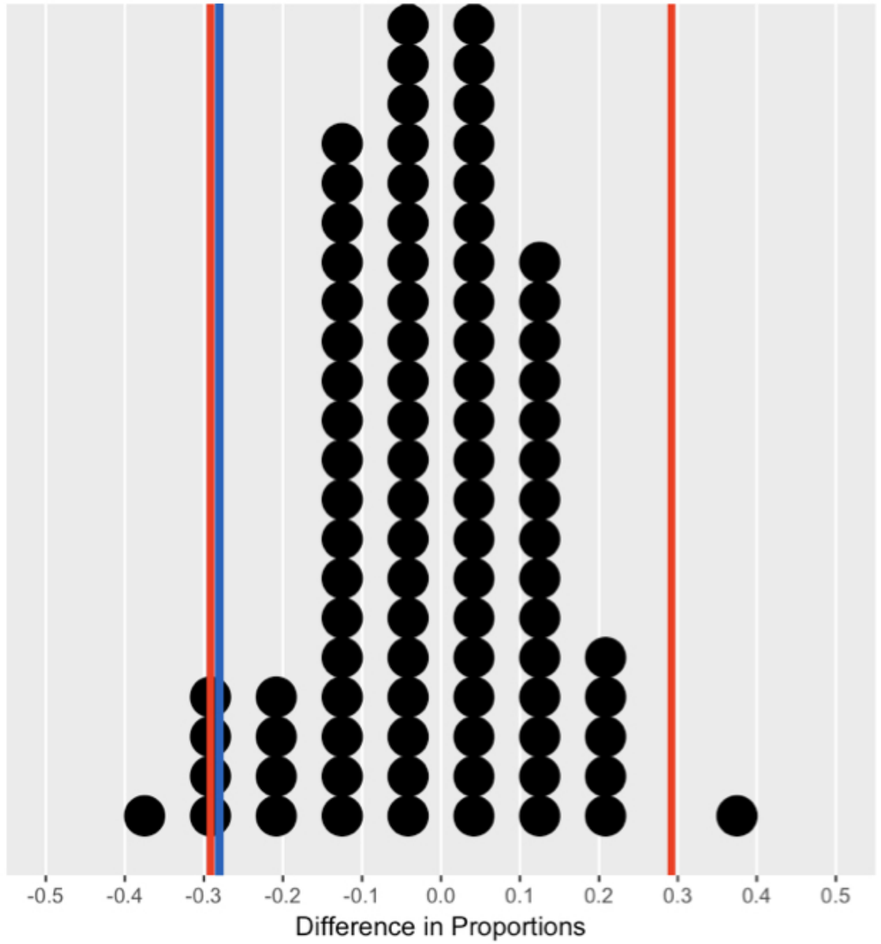
Calculating quantiles



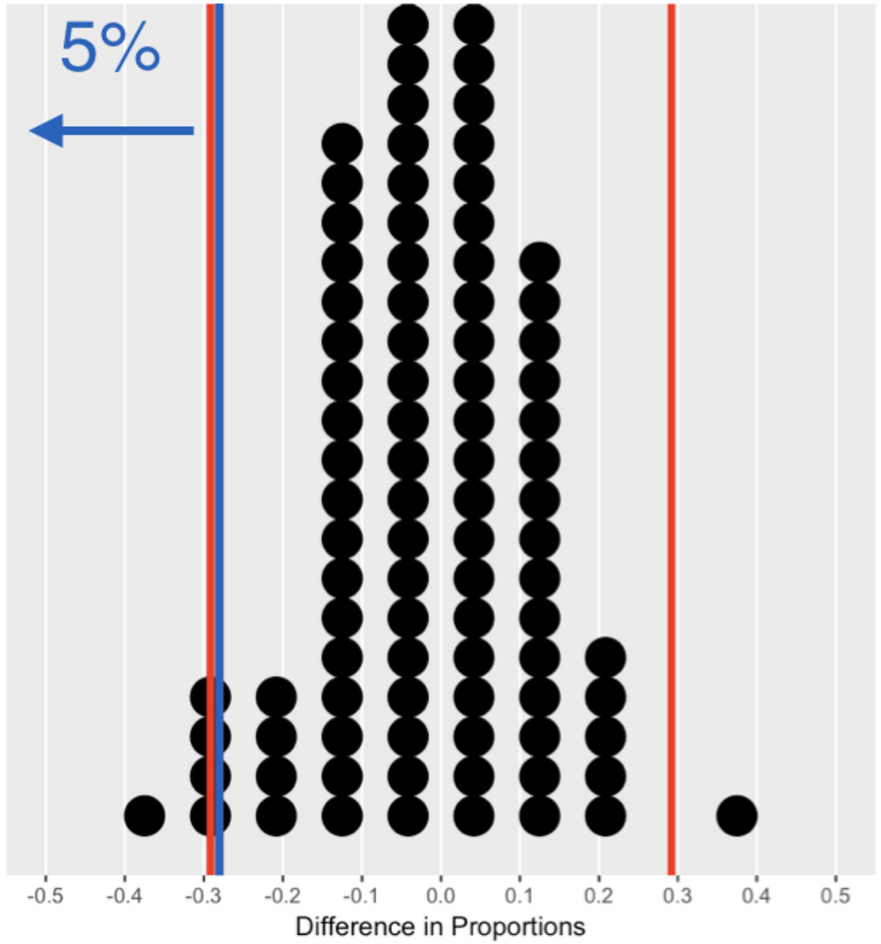
Calculating quantiles



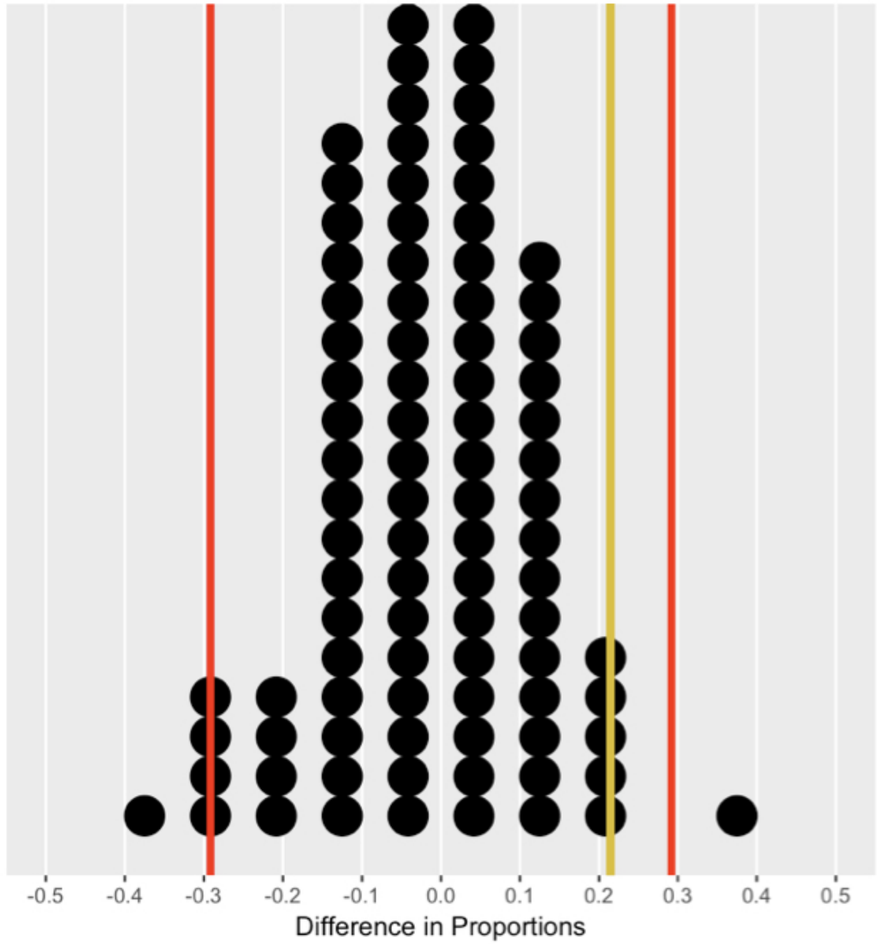
Calculating quantiles



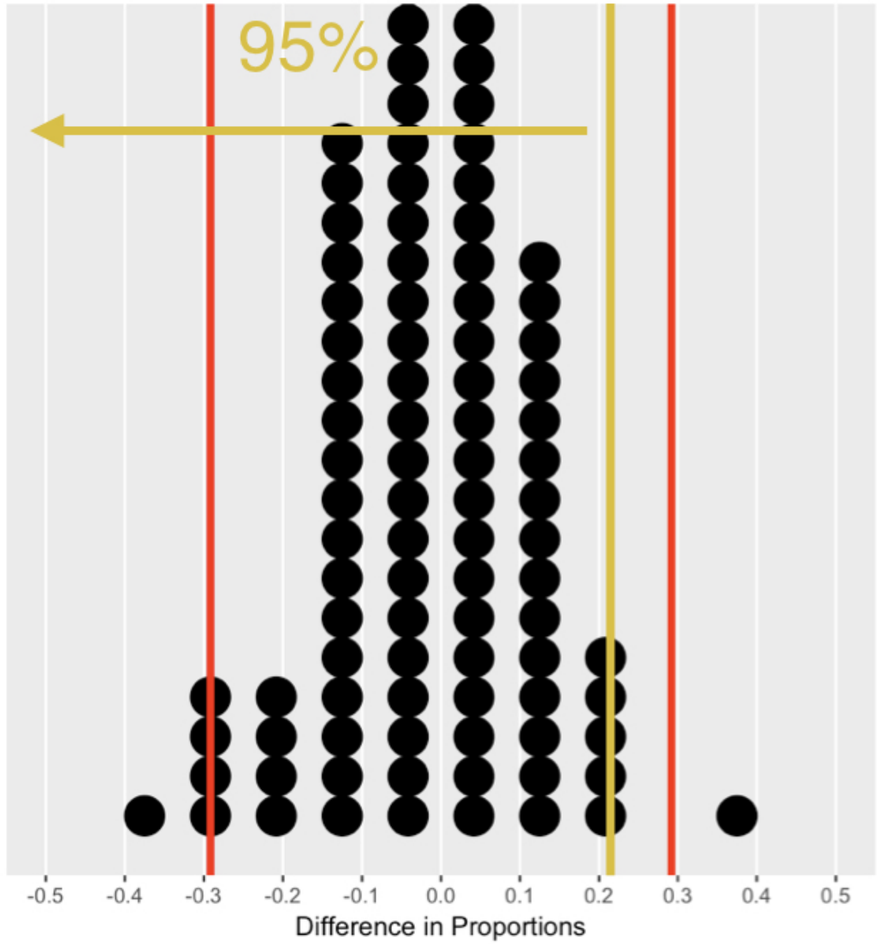
Calculating quantiles



Calculating quantiles



Calculating quantiles

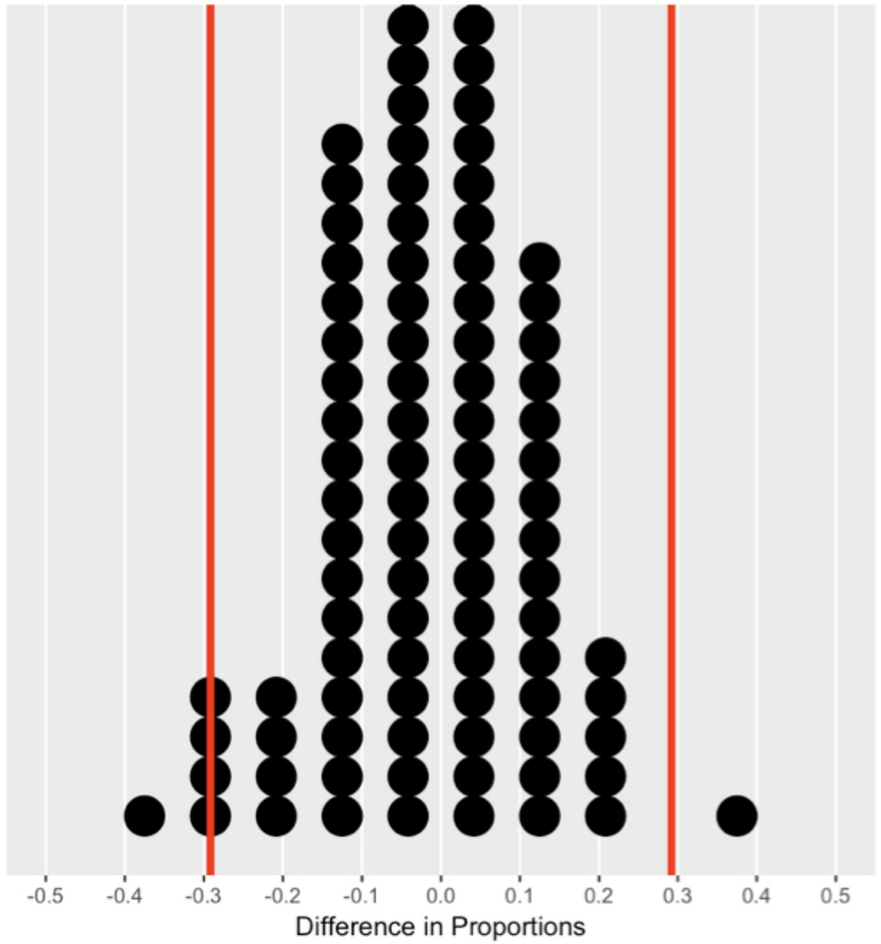


Quantile measurement

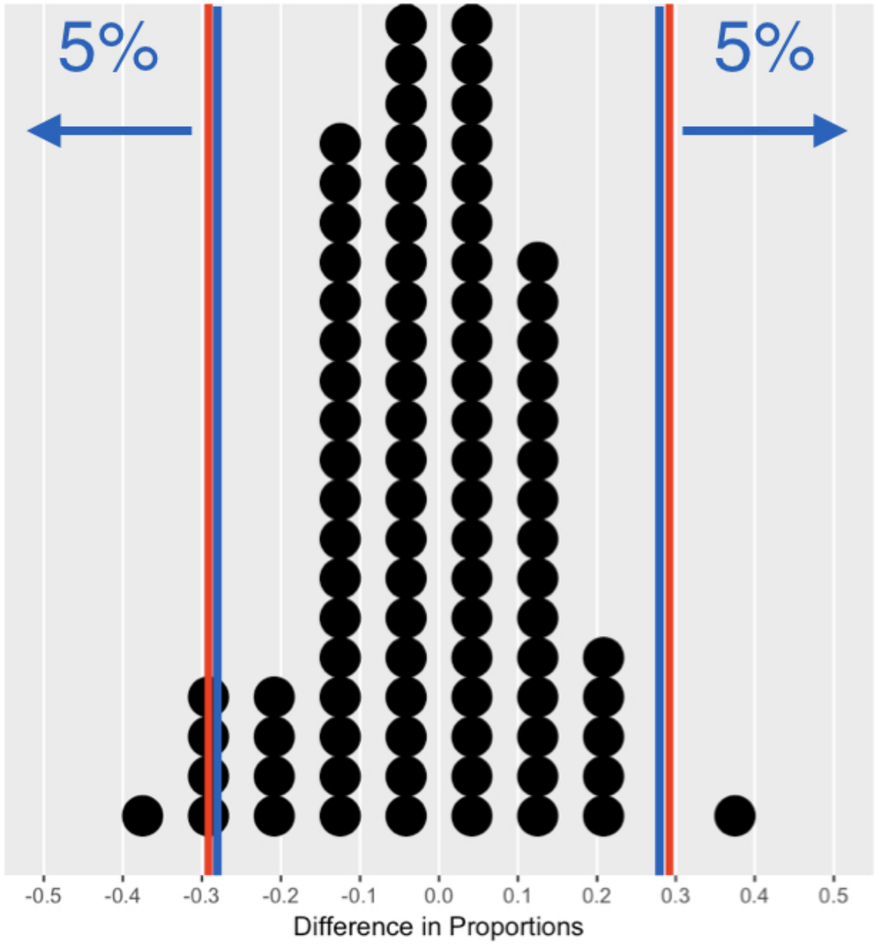
```
disc_perm %>%
    summarize(q.05 = quantile(diff_perm, p = 0.05),
              q.95 = quantile(diff_perm, p = 0.95))
```

```
# A tibble: 1 × 2
       q.05      q.95
      <dbl>     <dbl>
1 -0.2083333 0.2083333
```

Critical region



Critical region



Let's practice!

FOUNDATIONS OF INFERENCE

Why 0.05?

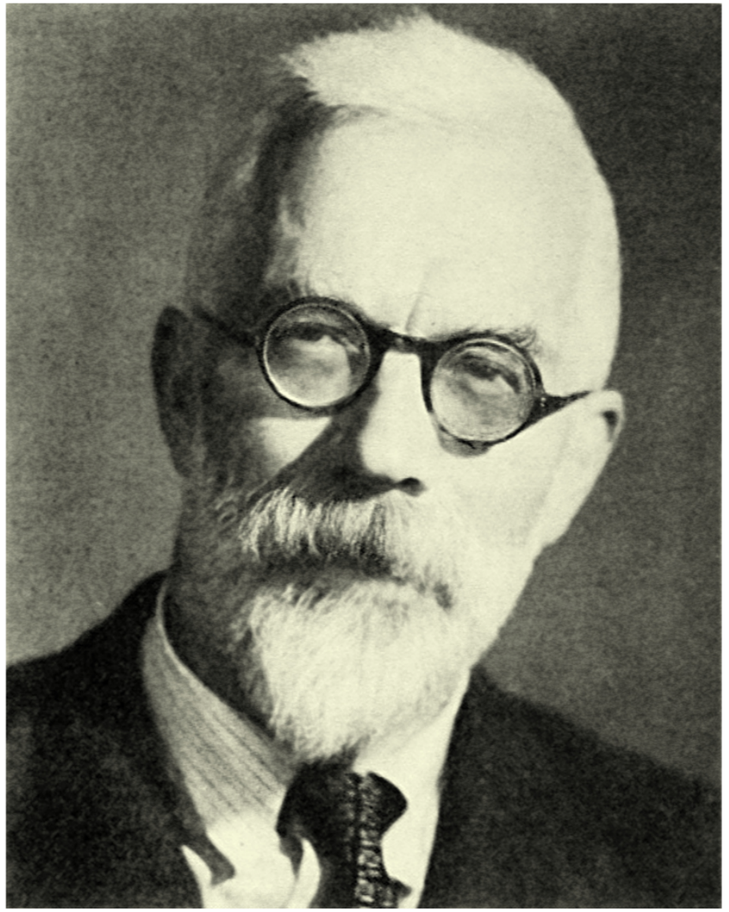
FOUNDATIONS OF INFERENCE

Jo Hardin
Instructor

datacamp

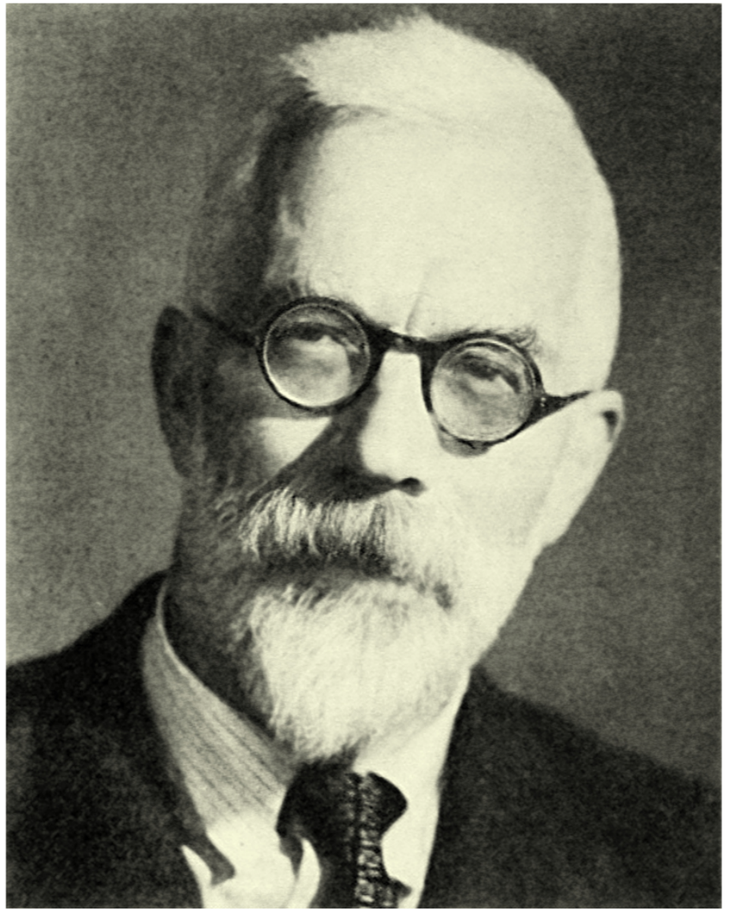
"... It is a common practice to judge a result significant, if it is of such a magnitude that it would have been **produced by chance not more frequently than once in twenty trials**. This is an arbitrary, but convenient, level of significance for the practical investigator, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation."

- RA Fisher (1929)



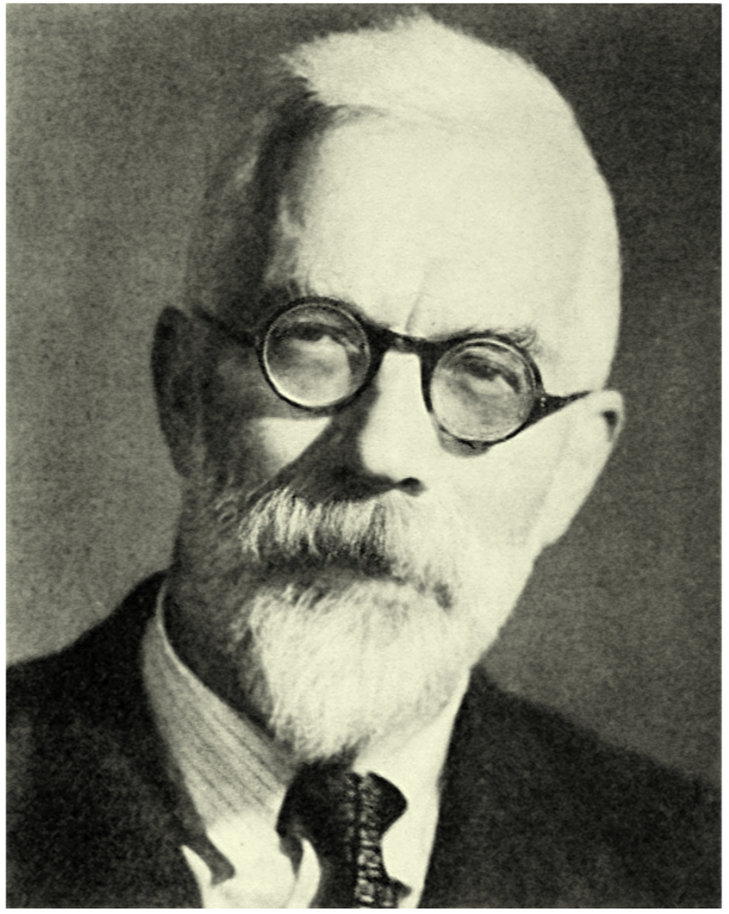
"... It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once in twenty trials. **This is an arbitrary, but convenient, level of significance for the practical investigator**, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation."

- RA Fisher (1929)



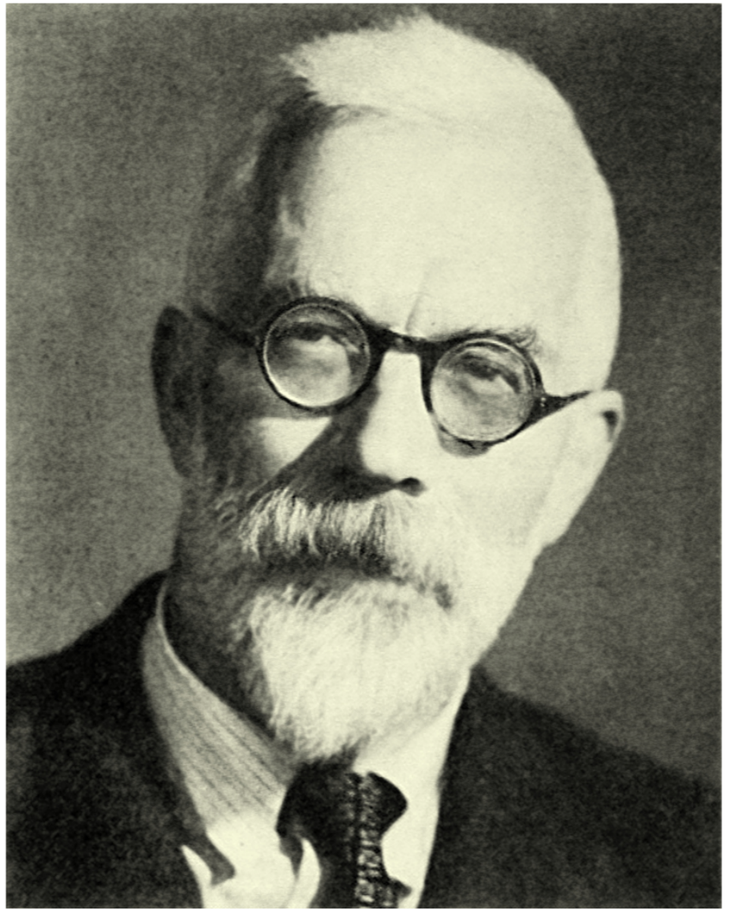
"... It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once in twenty trials. This is an arbitrary, but convenient, level of significance for the practical investigator, but **it does not mean that he allows himself to be deceived once in every twenty experiments**. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation."

- RA Fisher (1929)



"... It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once in twenty trials. This is an arbitrary, but convenient, level of significance for the practical investigator, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him **what to ignore**, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation."

- RA Fisher (1929)



Personal level of significance

$$P(H) = 1/2$$

Personal level of significance

$$P(HH) = (1/2)^2 = 1/4$$

Personal level of significance

$$P(HHH) = (1/2)^3 = 0.125$$

Personal level of significance

$$P(HHHH) = (1/2)^4 = 0.0625$$

Personal level of significance

$$P(HHHHH) = (1/2)^5 = 0.03125$$

Degree of skepticism

- Cutoff of 0.01 instead of 0.05 is more skeptical of observed results
- 0.05 is subjective
- Only significant results from well-designed studies should lead to further investigation

Let's practice!

FOUNDATIONS OF INFERENCE

What is a p-value?

FOUNDATIONS OF INFERENCE

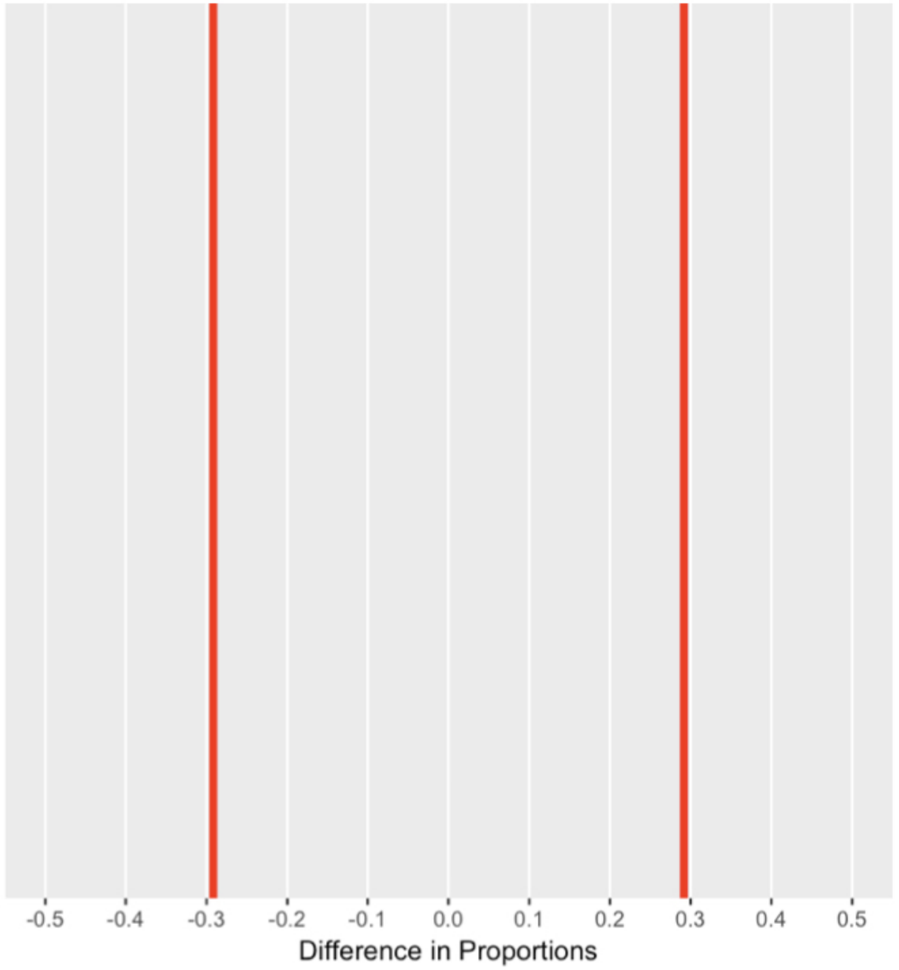
Jo Hardin
Instructor

datacamp

Understanding the null distribution

Observed data

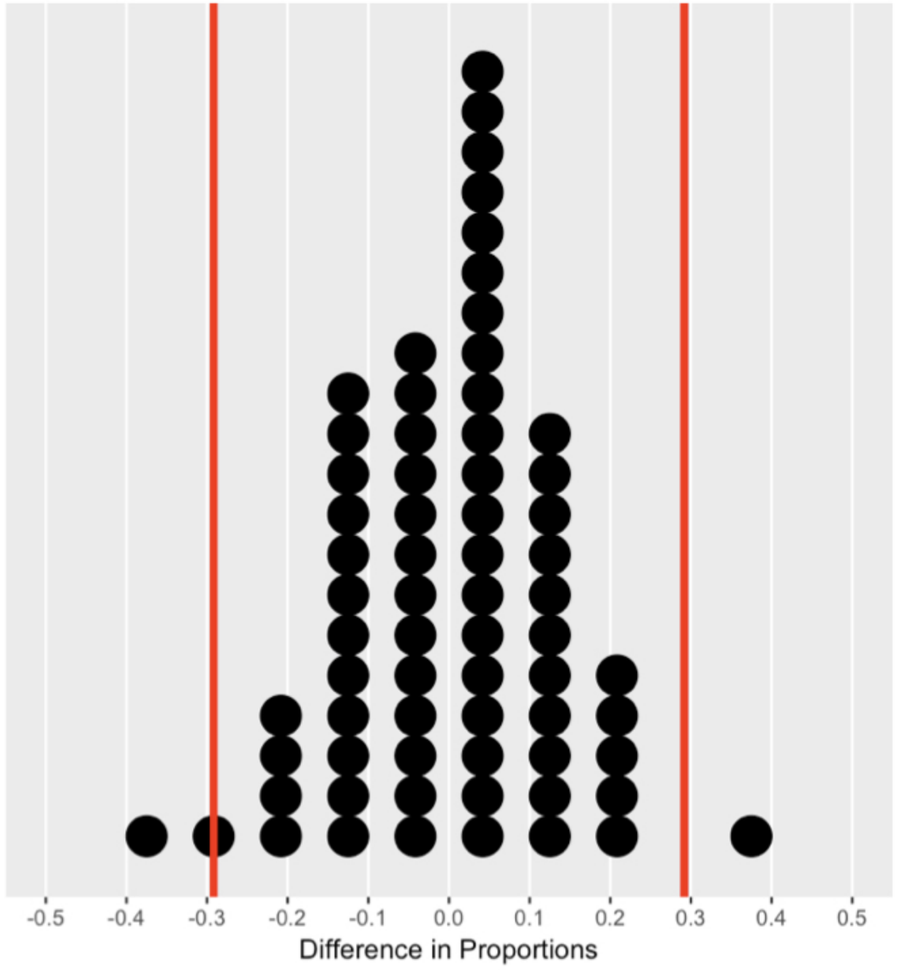
	Promoted	Not promoted
Male	21	3
Female	14	10



Understanding the null distribution

Observed data

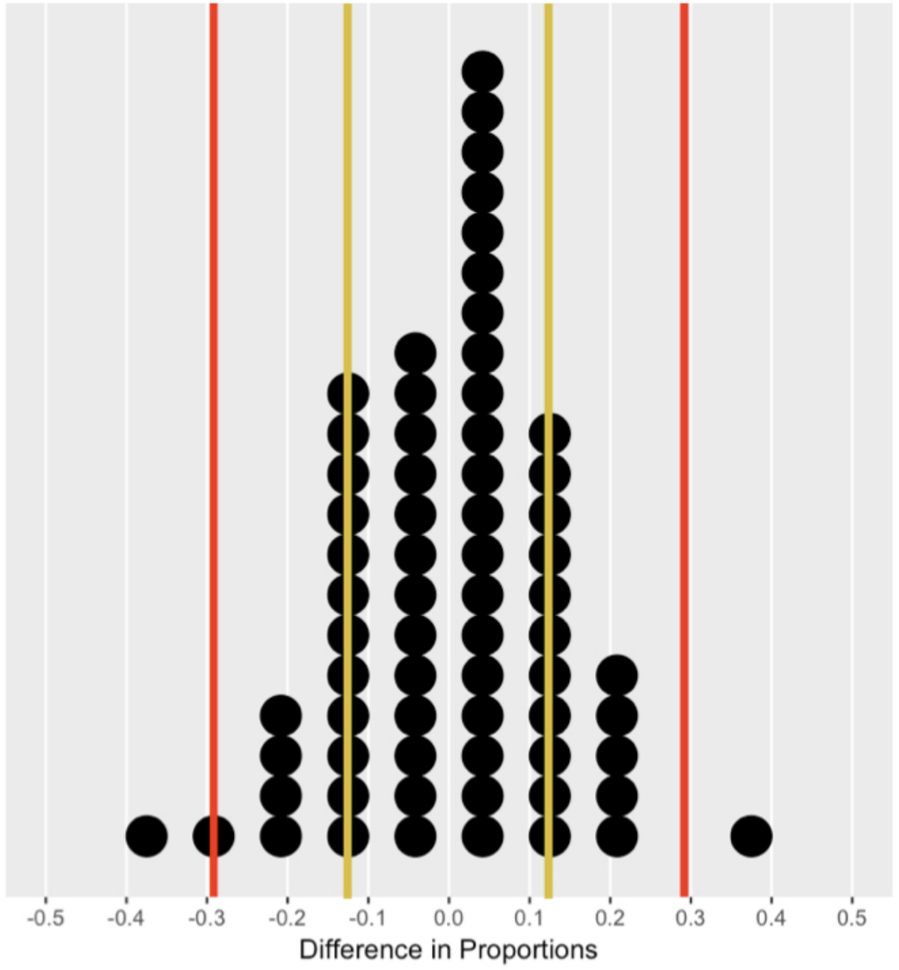
	Promoted	Not promoted
Male	21	3
Female	14	10



Understanding the null distribution

Observed data

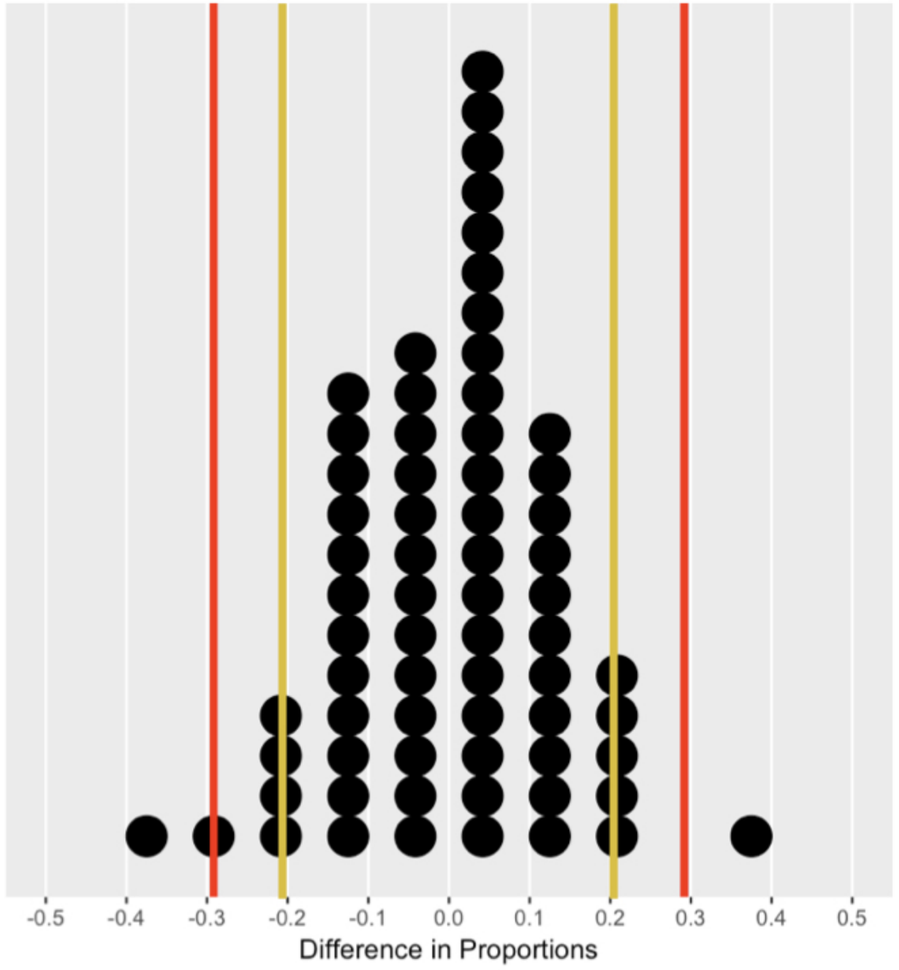
	Promoted	Not promoted
Male	21	3
Female	14	10



Understanding the null distribution

Observed data

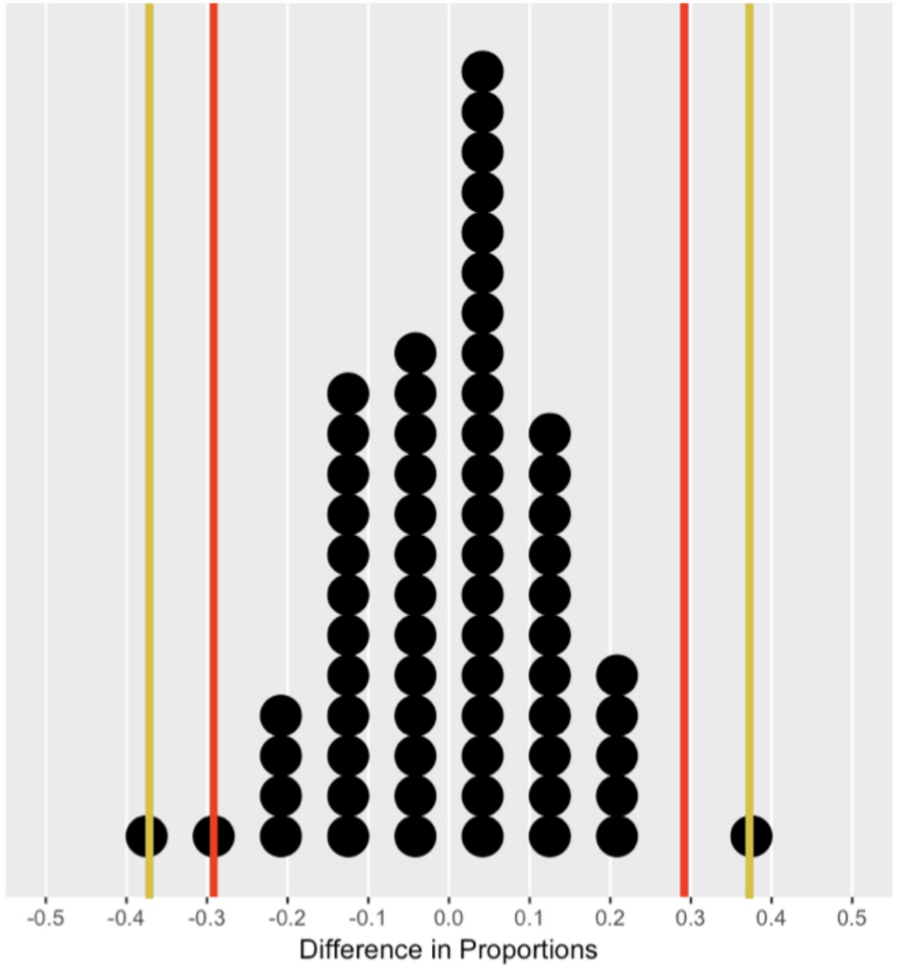
	Promoted	Not promoted
Male	21	3
Female	14	10



Understanding the null distribution

Observed data

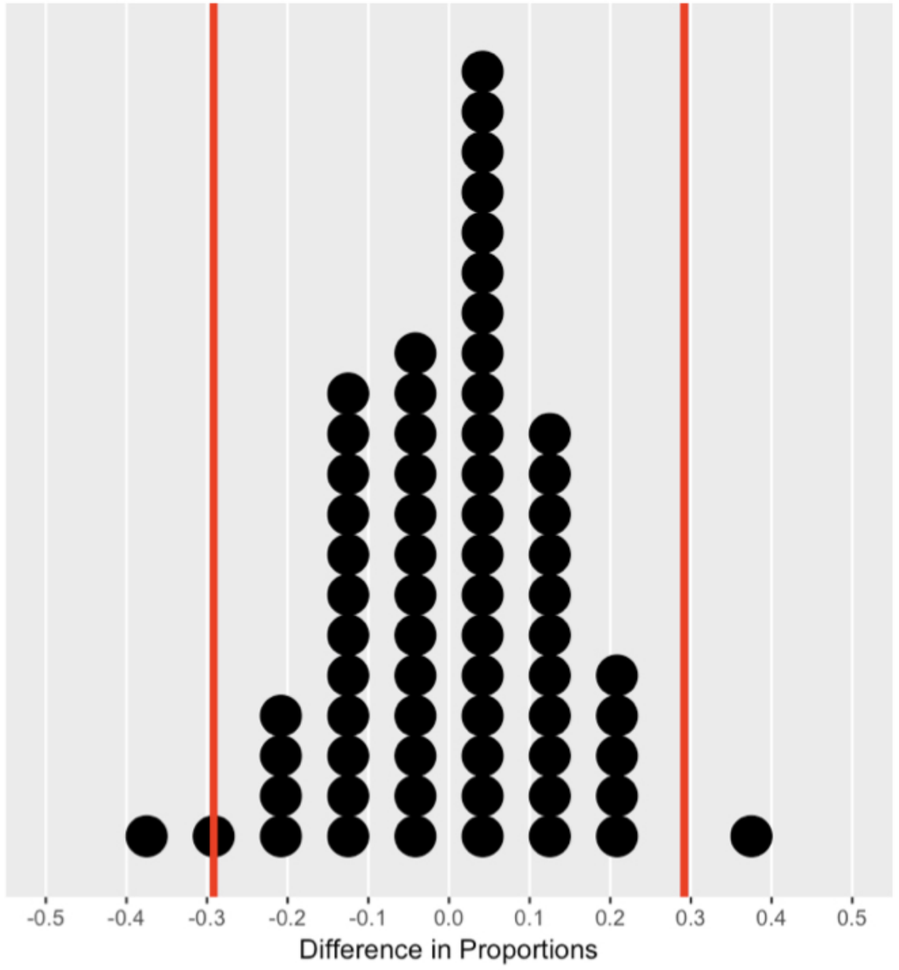
	Promoted	Not promoted
Male	21	3
Female	14	10



Understanding the null distribution

Observed data

	Promoted	Not promoted
Male	21	3
Female	14	10



Definition of p-value

> Probability of observing data as or more extreme than what we actually got given that the null hypothesis is true

Gender discrimination p-value

Probability of a observing a difference of 0.2917 or greater when promotion rates do not vary across gender = 0.03

Let's practice!

FOUNDATIONS OF INFERENCE

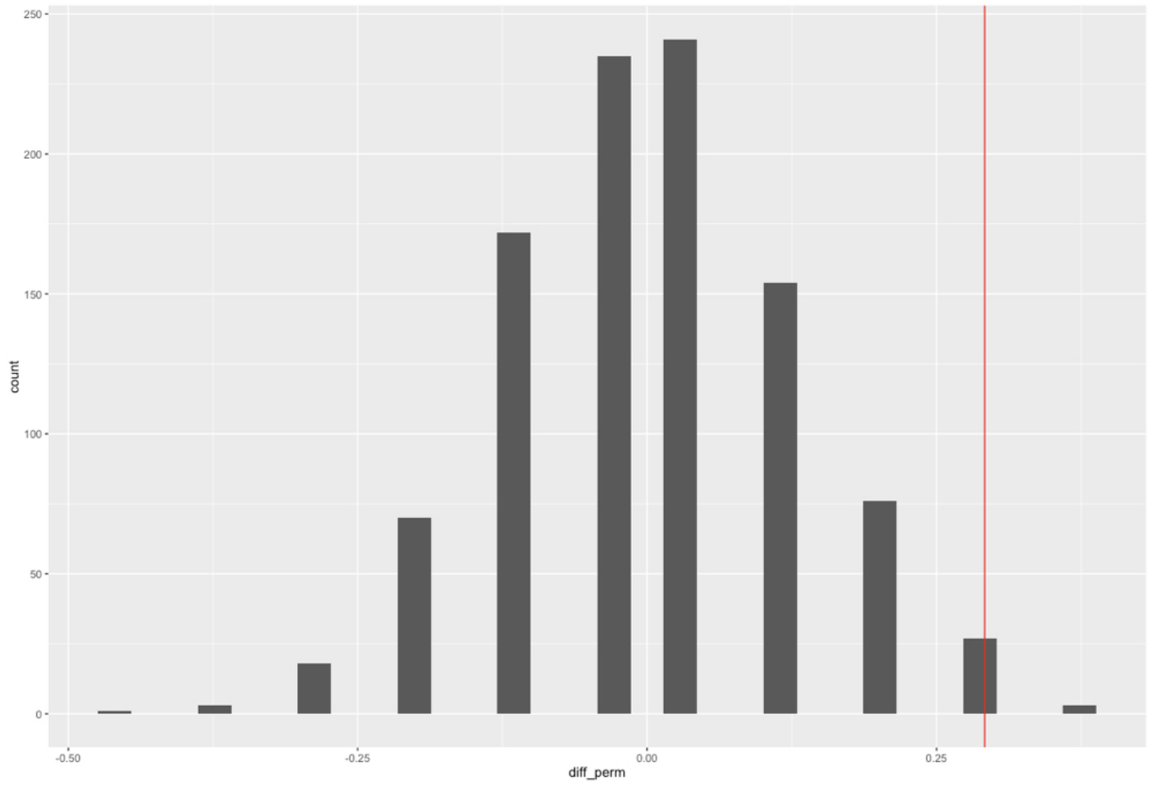
Summary of gender discrimination

FOUNDATIONS OF INFERENCE

Jo Hardin
Instructor

datacamp

Significance




```
disc_perm %>%
  summarize(pvalue =
      mean(diff_orig <= stat))
```

```
# A tibble: 1 × 1
  pvalue
   <dbl>
1   0.03
```

< 0.05, so reject H_0 in favor of H_A

Causation

- Study was randomized
- Nothing systematically different about two groups of participants other than which resumes they evaluated
- Any difference in promotion rates **is due to the gender of the applicant**

Random sample

- 35 individuals in sample were not randomly sampled from all managers
- They were at a management training session
- In order to generalize, we need:
  - More information about the study
  - Careful thinking about who the participants represent

Let's practice!

FOUNDATIONS OF INFERENCE