

# Foundations for Inference

- case study: gender discrimination
- introduction to inference via simulation
- normal distribution
- confidence intervals

# **Case Study: Gender Discrimination**

# gender discrimination

- 48 male bank supervisors given the same personnel file, asked to judge whether the person should be promoted
- files were identical, except for gender of applicant
- random assignment
- 35 / 48 promoted.

The study is testing whether females are unfairly discriminated against.

Is this an observational study or an experiment?

*There is random assignment, hence this is an experiment.*

*B.Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions", J.Applied Psychology, 59:9-14.*

# Data

At a first glance, does there appear to be a relationship between promotion and gender?

		<i>Promotion</i>		Total
		Promoted	Not Promoted	
<i>Gender</i>	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

*% of males promoted:*  $21 / 24 = 0.875$

*% of females promoted:*  $14 / 24 = 0.583$

# Practice

We saw a difference of almost 30% (29.2% to be exact) between the proportion of male and female files that are promoted. Based on this information, which of the below is true?

- A. If we were to repeat the experiment we will definitely see that more female files get promoted. This was a fluke.
- B. Promotion is dependent on gender, males are more likely to be promoted, and hence there is gender discrimination against women in promotion decisions.
- C. The difference in the proportions of promoted male and female files is due to chance, this is not evidence of gender discrimination against women in promotion decisions.
- D. Women are less qualified than men, and this is why fewer females get promoted.

**B, C** : the difference might be indicative of discrimination against women in promotion decisions, or it might just be due to chance. We need further analysis to make a decision between these two competing claims.

## two competing claims

### 1. “There is nothing going on.”

Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.

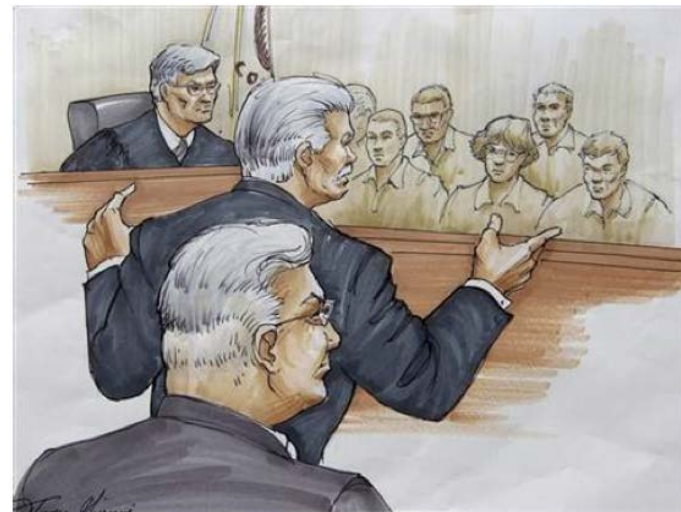
→ Null Hypothesis

### 2. “There is something going on.”

Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.

→ Alternative Hypothesis

# A trial as a hypothesis test



null hypothesis

$H_0$  : Defendant is innocent

alternative hypothesis

$H_a$  : Defendant is guilty

present the evidence

collect data

the burden of  
proof

judge the evidence

“Could these data plausibly have  
happened by chance if the null  
hypothesis were true?”

yes

Fail to reject  $H_0$

no

Reject  $H_0$

# A Trial as a Hypothesis Test

Hypothesis testing is very much like a court trial.

- $H_0$  : Defendant is innocent  
 $H_A$  : Defendant is guilty
- We then present the evidence - collect data.
- Then we judge the evidence - “Could these data plausibly have happened by chance if the null hypothesis were true?”
  - If they were very unlikely to have occurred, then the evidence raises more than a reasonable doubt in our minds about the null hypothesis.
- Ultimately we must make a decision. How unlikely is unlikely?





# A Trial as a Hypothesis Test (cont.)

- If the evidence is not strong enough to reject the assumption of innocence, the jury returns with a verdict of “not guilty”.
  - The jury does not say that the defendant is innocent, just that there is not enough evidence to convict.
  - The defendant may, in fact, be innocent, but the jury has no way of being sure.
- Said statistically, we fail to reject the null hypothesis.
  - We never declare the null hypothesis to be true, because we simply do not know whether it's true or not.
  - Therefore we never “accept the null hypothesis”.

# A Trial as a Hypothesis Test (cont.)

- In a trial, the burden of proof is on the prosecution.
- In a hypothesis test, the burden of proof is on the unusual claim.
- The null hypothesis is the ordinary state of affairs (the status quo), so it's the alternative hypothesis that we consider unusual and for which we must gather evidence.

## recap: hypothesis testing framework

- start with a *null hypothesis* ( $H_0$ ) that represents the present situation.
- set an *alternative hypothesis* ( $H_A$ ) that represents our research question, i.e. what we're testing for.
- conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation (today) or theoretical methods (later in the course).
  - If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis.
  - If they do, then we reject the null hypothesis in favor of the alternative.

# Simulating the experiment...

... under the assumption of independence, i.e. leave things up to chance.

- the results from the simulations based on the *chance model* look like the data  
→ determine that the difference between the proportions of promoted file between males and females was simply *due to chance*  
(promotion and gender are *independent*).
- the results from the simulations based on the chance model do not look like the data,  
→ determine that the difference between the proportions of promoted files between males and females was not due to chance, but  
*due to an actual effect of gender* (promotion and gender are *dependent*).

# simulation scheme

Use a deck of playing cards to simulate this experiment.

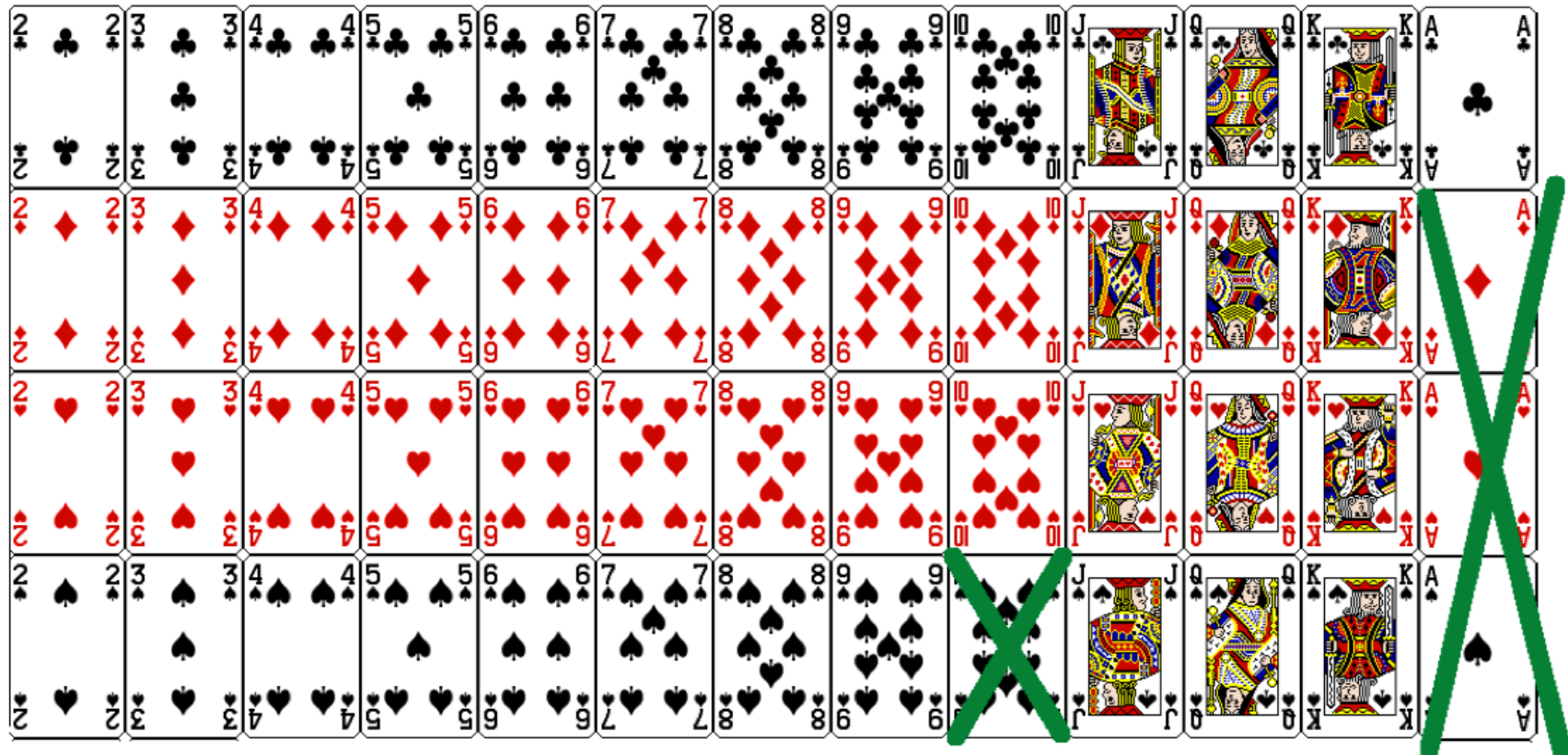
1. a face card : *not promoted* , a non-face card: a *promoted*.

- set aside the jokers, consider aces as face cards
- take out 3 aces → 13 face cards left in the deck (face cards: A, K, Q, J).
- take out a number card → 35 number (non-face) cards left in the deck (number cards: 2-10).

# Step 1

35 number (non-face) cards

13 face cards



# simulation scheme

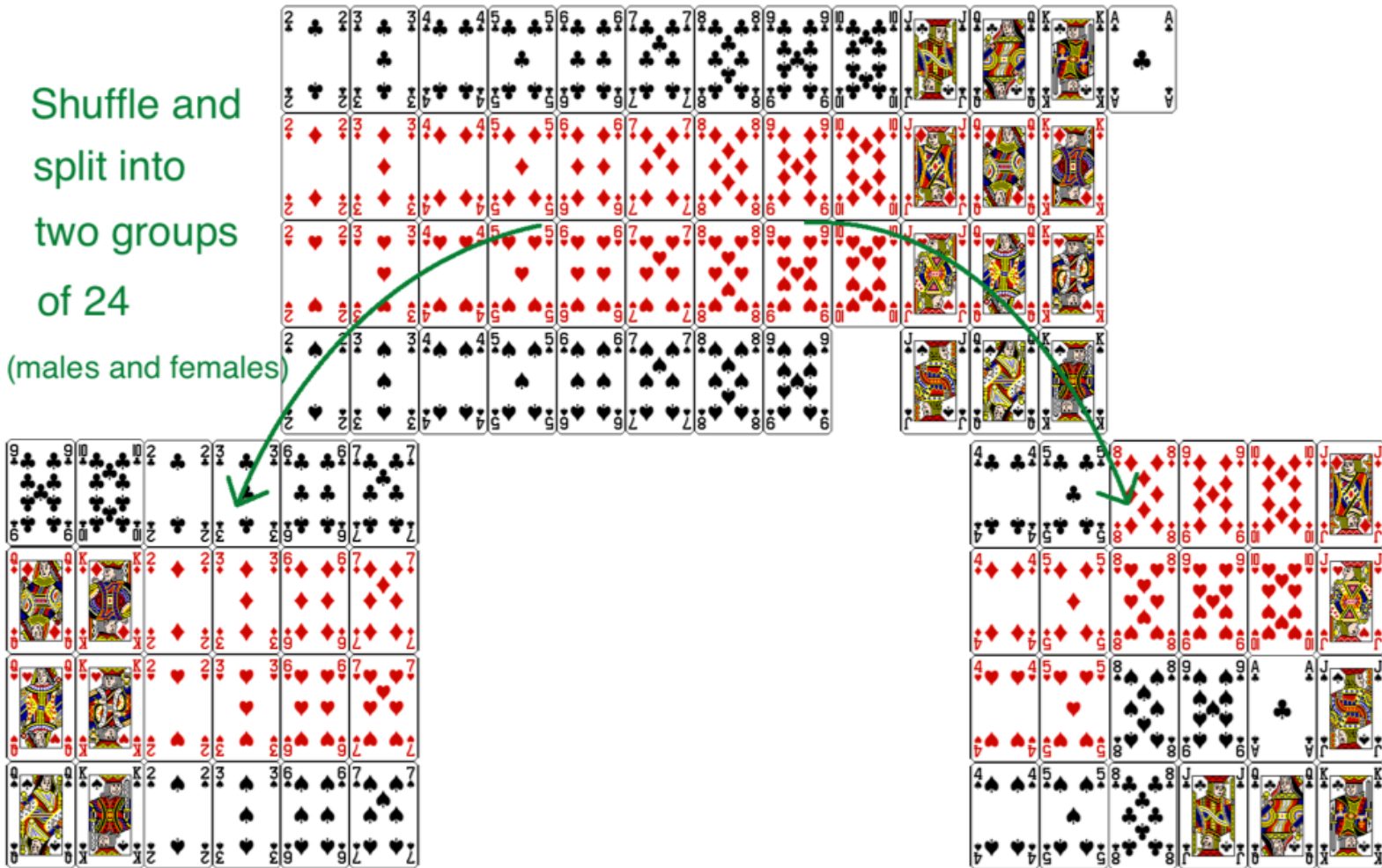
Use a deck of playing cards to simulate this experiment.

1. a face card : *not promoted* , a non-face card: a *promoted*.
  - set aside the jokers, consider aces as face cards
  - take out 3 aces → 13 face cards left in the deck (face cards: A, K, Q, J).
  - take out a number card → 35 number (non-face) cards left in the deck (number cards: 2-10).
2. shuffle the cards, deal into two groups of size 24, representing males and females.

## Step 2

Shuffle and  
split into  
two groups  
of 24

(males and females)





# simulation scheme

Use a deck of playing cards to simulate this experiment.

1. a face card : *not promoted* , a non-face card: a *promoted*.
  - set aside the jokers, consider aces as face cards
  - take out 3 aces → 13 face cards left in the deck (face cards: A, K, Q, J).
  - take out a number card → 35 number (non-face) cards left in the deck (number cards: 2-10).
2. shuffle the cards and deal them into two groups of size 24, representing males and females.
3. count how many number cards are in each group (representing promoted files).
4. calculate the proportion of promoted files in each group, take the difference (male - female), and record this value.

# Practice

Given that we randomly shuffled the cards into the two files to simulate randomly assigning male and female labels to the files, what would you expect to be the difference between the simulated proportions of promoted male and female files?

0?

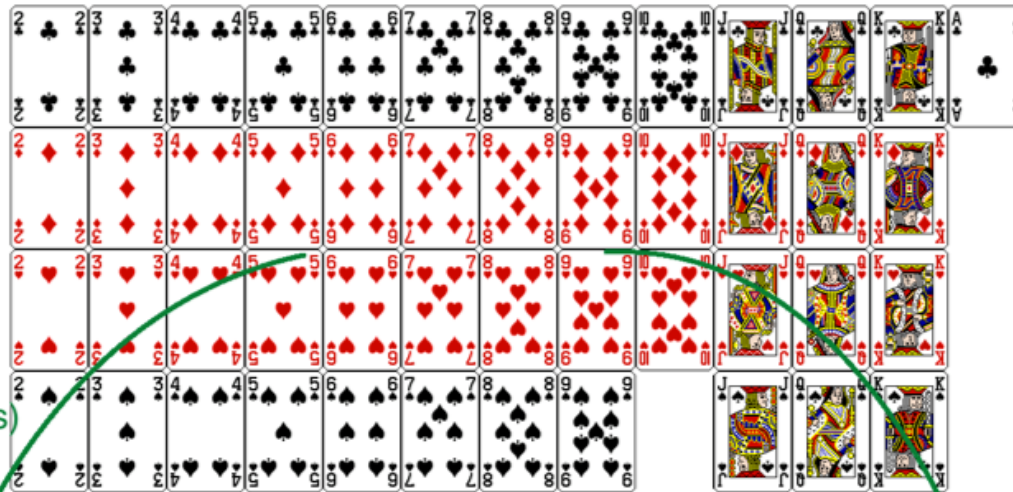
0.3?

*Since we randomly shuffled the cards we would expect equal proportions of "promoted" cards to end up in the male and female piles, hence resulting in 0 difference between the simulated proportions of promoted male and female files.*

# Step 3 & 4

Shuffle and  
split into  
two groups  
of 24

(males and females)



Males  
18 promoted  
 $18 / 24 = 0.75$

Females  
17 promoted  
 $17 / 24 = 0.708$

Difference =  $0.75 - 0.708 = 0.042$



# simulation scheme

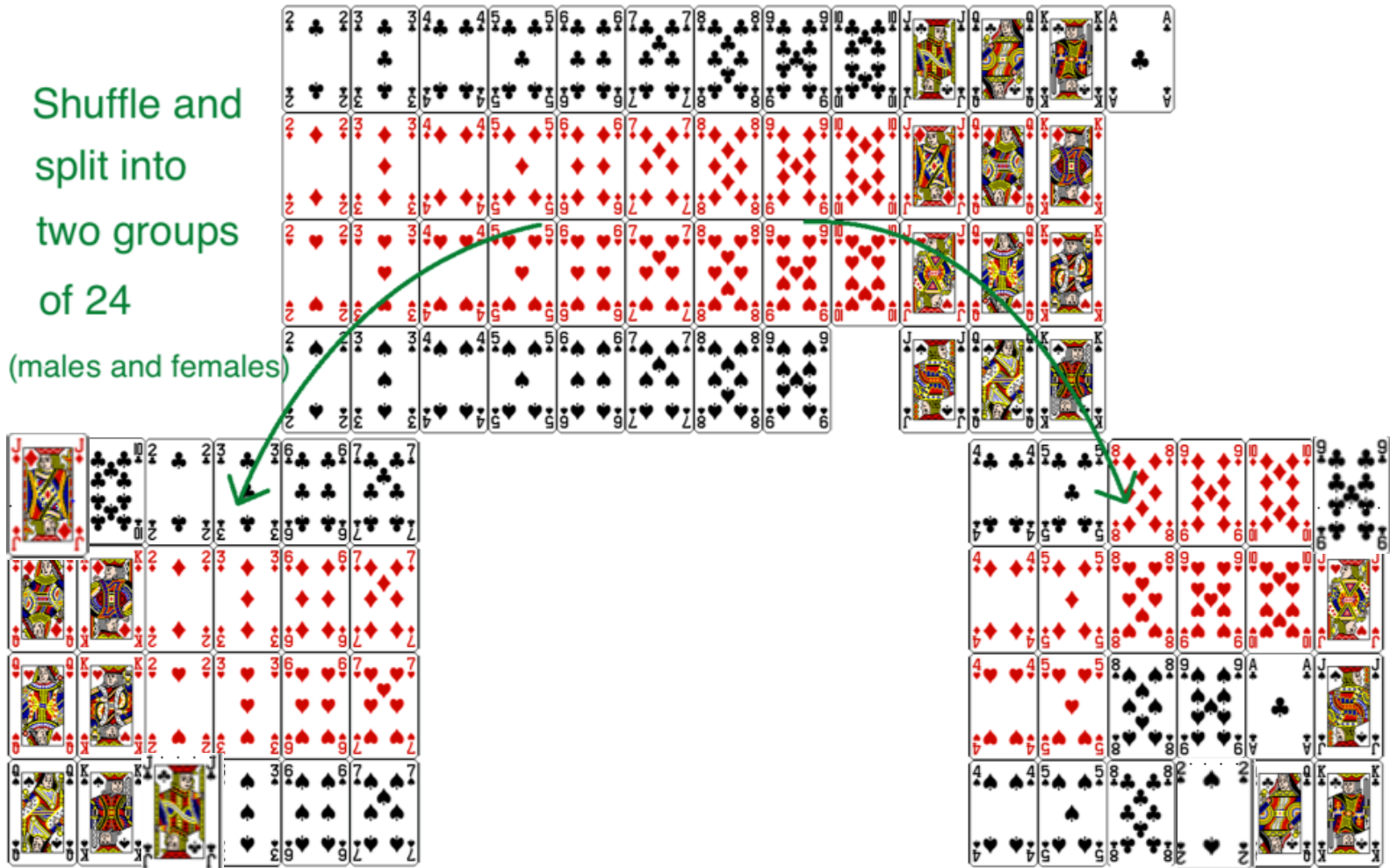
Use a deck of playing cards to simulate this experiment.

1. a face card : *not promoted* , a non-face card: a *promoted*.
  - set aside the jokers, consider aces as face cards
  - take out 3 aces → 13 face cards left in the deck (face cards: A, K, Q, J).
  - take out a number card → 35 number (non-face) cards left in the deck (number cards: 2-10).
2. shuffle the cards and deal them into two groups of size 24, representing males and females.
3. count how many number cards are in each group (representing promoted files).
4. calculate the proportion of promoted files in each group, take the difference (male - female), and record this value.
5. Repeat steps 2 - 4 many times.

## Step 2

Shuffle and  
split into  
two groups  
of 24

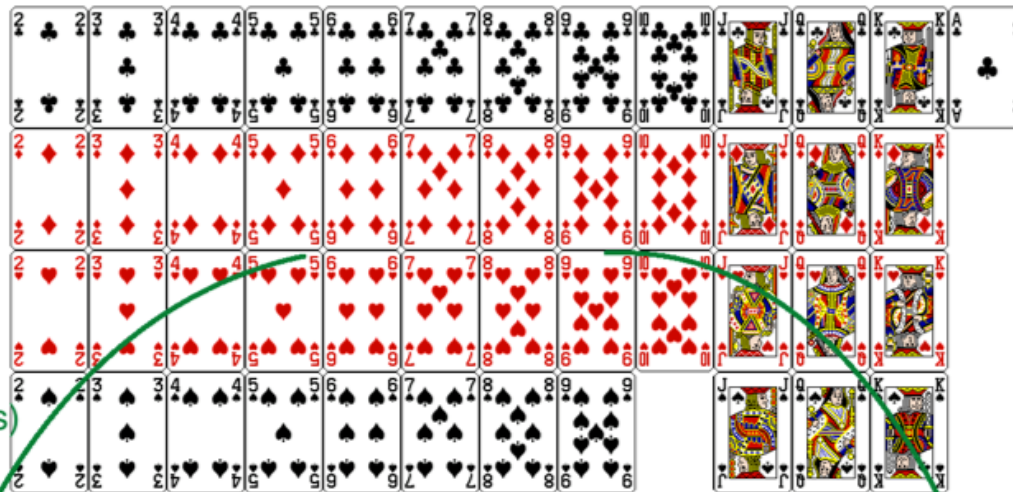
(males and females)



# Steps 3 & 4

Shuffle and  
split into  
two groups  
of 24

(males and females)



Males  
16 promoted  
 $16 / 24 = 0.667$

Females  
19 promoted  
 $19 / 24 = 0.791$

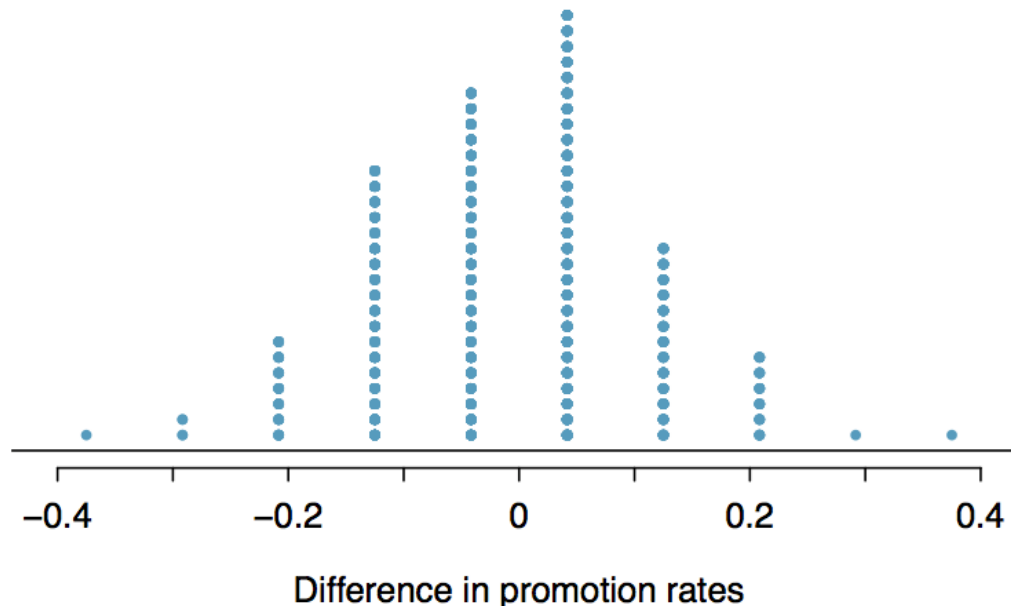
Difference =  $0.667 - 0.791 = -0.125$



## simulations using software

These simulations are tedious and slow to run using the method described earlier. In reality, we use software to generate the simulations.

The dot plot below shows the distribution of simulated differences in promotion rates based on 100 simulations.





# Practice

Do the results of the simulation you just ran provide convincing evidence of gender discrimination against women, i.e. dependence between gender and promotion decisions?

- A. No, the data do not provide convincing evidence for the alternative hypothesis, therefore we can't reject the null hypothesis of independence between gender and promotion decisions. The observed difference between the two proportions was due to chance.
- B. Yes, the data provide convincing evidence for the alternative hypothesis of gender discrimination against women in promotion decisions. The observed difference between the two proportions was due to a real effect of gender.

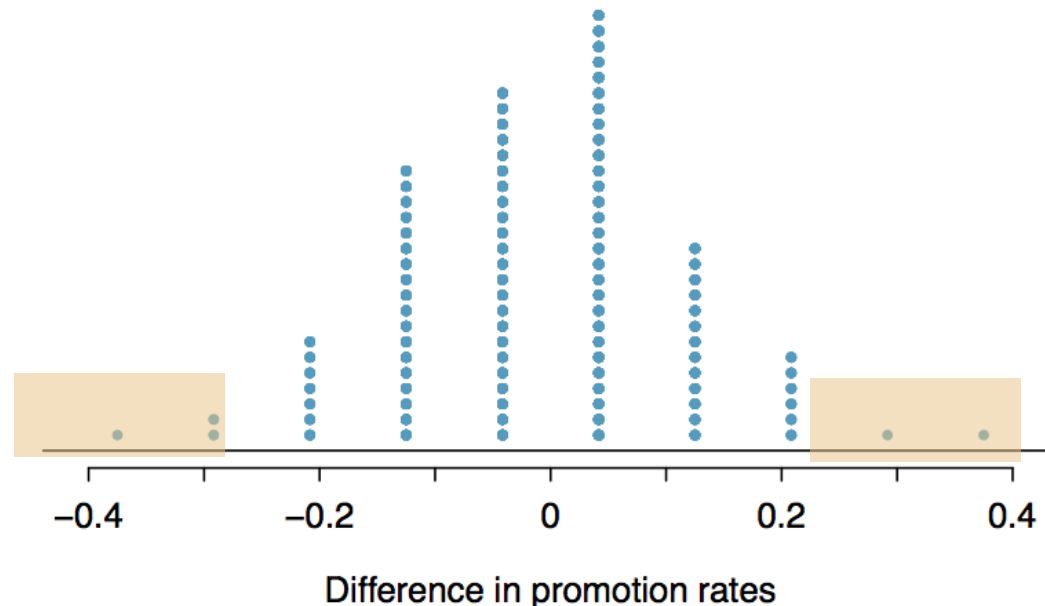


## summary

- set a null and an alternative hypothesis
- simulate the experiment assuming that the null hypothesis is true
- evaluate the probability of observing an outcome at least as extreme as the one observed in the original data

## p-value

- if this probability is low, reject the null hypothesis in favor of the alternative



# Homework Week 6

**Problem 1:** State the appropriate null hypothesis  $H_0$  and alternative hypothesis  $H_A$ . The mean area of the several thousand apartment in a new development is advertised to be 1250 square feet. A tenant group thinks that the apartments are smaller than advertised. They hire an engineer to measure a sample of apartment to test their suspicion.

**Problem 2:** Find a news article or research report that conducts and describes a hypothesis test