



Statistical Significance

From Data to Insight

Dr. Çetinkaya-Rundel
July 25, 2016

Is yawning
contagious?

Do you think yawning is contagious?

(A) Yes

(B) No

Is yawning contagious?

[http://www.discovery.com/tv-shows/mythbusters/
videos/is-yawning-contagious-minimyth.htm](http://www.discovery.com/tv-shows/mythbusters/videos/is-yawning-contagious-minimyth.htm)

MythBusters experiment

- ▶ 50 people were randomly assigned to two groups:
 - ▶ Treatment: see someone yawn, $n = 34$
 - ▶ Control: don't see someone yawn, $n = 16$

	Treatment	Control	Total
Yawn	10	4	14
Not yawn	24	12	36
Total	34	16	50
% yawners	0.29	0.25	

Two competing claims

- ▶ **Null hypothesis:** “There is nothing going on” - Yawning and seeing someone yawn are **independent**
- ▶ **Alternative hypothesis:** “There is something going on” - Yawning and seeing someone yawn are **dependent**

A trial as a hypothesis test

- ▶ Two competing claims:
 - ▶ H_0 : Defendant is innocent
 - ▶ H_A : Defendant is guilty
- ▶ Present the evidence: collect data
- ▶ Judge the evidence: “Could these data plausibly have happened by chance if the null hypothesis were true?”
- ▶ Make a decision: “How unlikely is unlikely?”

Hypothesis testing framework

- ▶ Start with a null hypothesis (H_0) that represents the status quo
- ▶ Set an alternative hypothesis (H_A) that represents the 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 or theoretical methods
- ▶ If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, stick with the null hypothesis
- ▶ If they do, then reject the null hypothesis in favor of the alternative

Simulation scheme

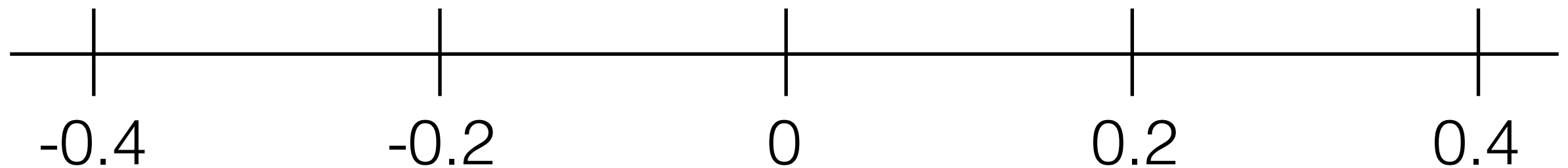
- ▶ A regular deck of cards is comprised of 52 cards: 4 aces, 4 of numbers 2-10, 4 jacks, 4 queens, and 4 kings.
- ▶ Take out two aces from the deck of cards and set them aside.
- ▶ The remaining 50 playing cards to represent each participant in the study:
 - ▶ 14 face cards (including the 2 aces) represent the people who yawn.
 - ▶ 36 non-face cards represent the people who don't yawn.

DEMO: Watch me go through the activity before you start it in your teams

Activity: running the simulation

- ▶ Shuffle the 50 cards at least 7 times to ensure that the cards counted out are from a random process
- ▶ Divide the cards into two decks:
 - ▶ deck 1: 16 cards → control
 - ▶ deck 2: 34 cards → treatment
- ▶ Count the number of face cards (yawners) in each deck
- ▶ Calculate the difference in proportions of yawners (treatment - control), and submit this value using your clicker (value must be between 0 and 1) - only one submission per team per simulation
- ▶ Repeat steps (1) - (4) many times

Activity: results



Making a decision

- ▶ Results from the simulations look like the data → the difference between the proportions of yawners in the treatment and control groups was due to chance (yawning and seeing someone yawn are independent)
- ▶ Results from the simulations do not look like the data → the difference between the proportions of yawners in the treatment and control groups was not due to chance (yawning and seeing someone yawn are dependent)

Do the simulation results suggest that yawning is contagious, i.e. does seeing someone yawn and yawning appear to be dependent?

(**Hint:** In the actual data the difference was 0.04, does this appear to be an unusual observation for the chance model?)

(A) Yes

(B) No

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:
 - ▶ Conclude that the data provide convincing evidence for the alternative hypothesis
- ▶ If this probability is high, fail to reject the null hypothesis in favor of the alternative:
 - ▶ Conclude that the data do not provide convincing evidence for the alternative hypothesis

Tapping on caffeine

Tapping on caffeine

- ▶ In a double-blind experiment a sample of male college students were asked to tap their fingers at a rapid rate.
- ▶ The sample was then divided at random into two groups of 10 students each.
- ▶ Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group.
- ▶ After a two hour period, each student was tested to measure finger tapping rate (taps per minute).

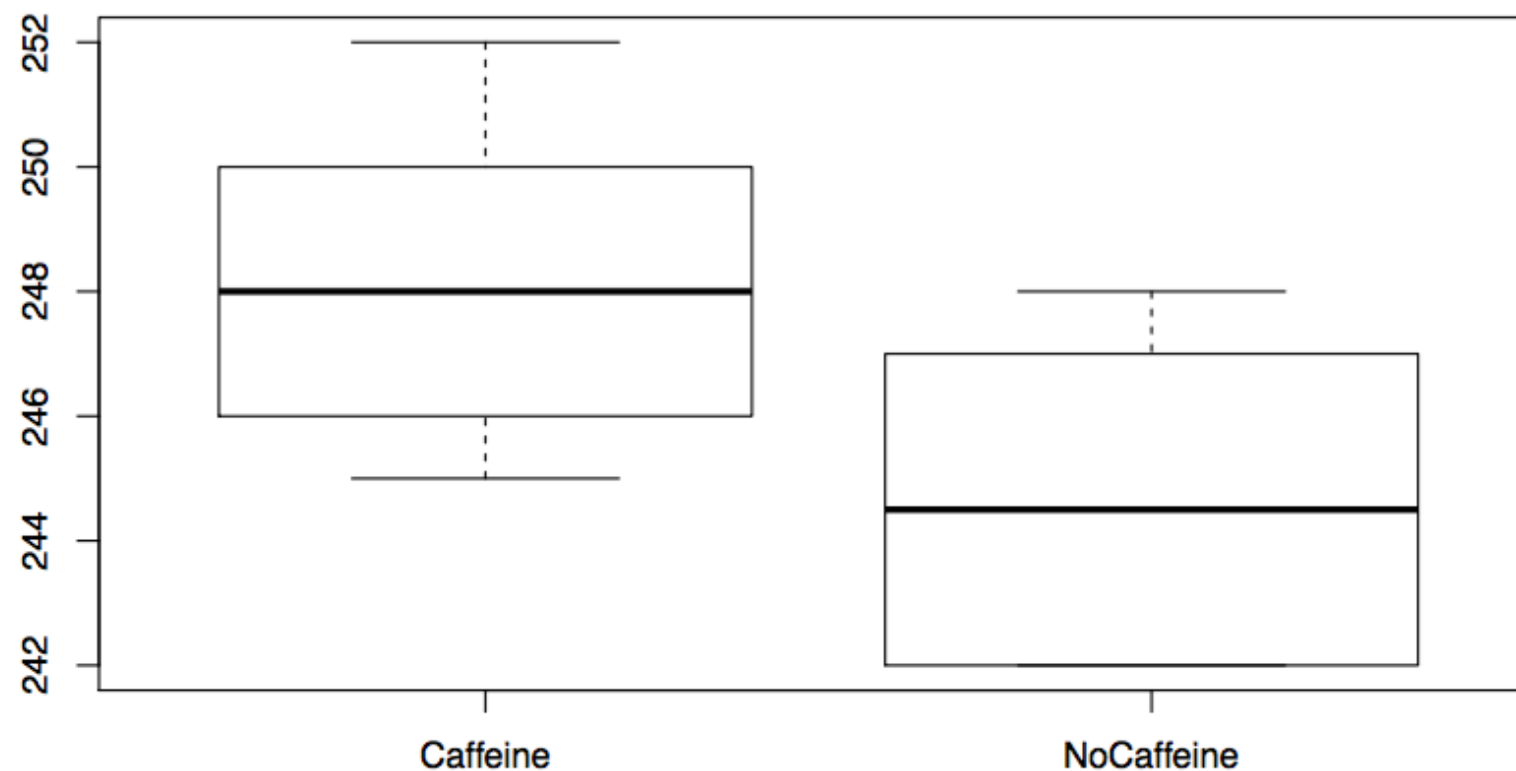
	Taps	Group
1	246	Caffeine
2	248	Caffeine
3	250	Caffeine
4	252	Caffeine
5	248	Caffeine
6	250	Caffeine
...		
16	248	NoCaffeine
17	242	NoCaffeine
18	244	NoCaffeine
19	246	NoCaffeine
20	242	NoCaffeine

What type of plot would be useful to visualize the distributions of tapping rate in the caffeine and no caffeine groups?

- (A) Bar plot
- (B) Scatterplot
- (C) Pie chart
- (D) Side-by-side box plots
- (E) Single box plot

Exploratory data analysis

	Caffeine	No Caffeine	Difference
mean	248.3	244.8	3.5
SD	2.21	2.39	-0.18
median	248	245	3
IQR	3.5	4.25	-0.75



We are interested in finding out if caffeine increases tapping rate. Which of the following are the correct set of hypotheses?

Note: μ = population mean, \bar{x} = sample mean

(A) $H_0 : \mu_{\text{caff}} = \mu_{\text{no caff}}; H_A : \mu_{\text{caff}} < \mu_{\text{no caff}}$

(B) $H_0 : \mu_{\text{caff}} = \mu_{\text{no caff}}; H_A : \mu_{\text{caff}} > \mu_{\text{no caff}}$

(C) $H_0 : \bar{x}_{\text{caff}} = \bar{x}_{\text{no caff}}; H_A : \bar{x}_{\text{caff}} > \bar{x}_{\text{no caff}}$

(D) $H_0 : \mu_{\text{caff}} > \mu_{\text{no caff}}; H_A : \mu_{\text{caff}} = \mu_{\text{no caff}}$

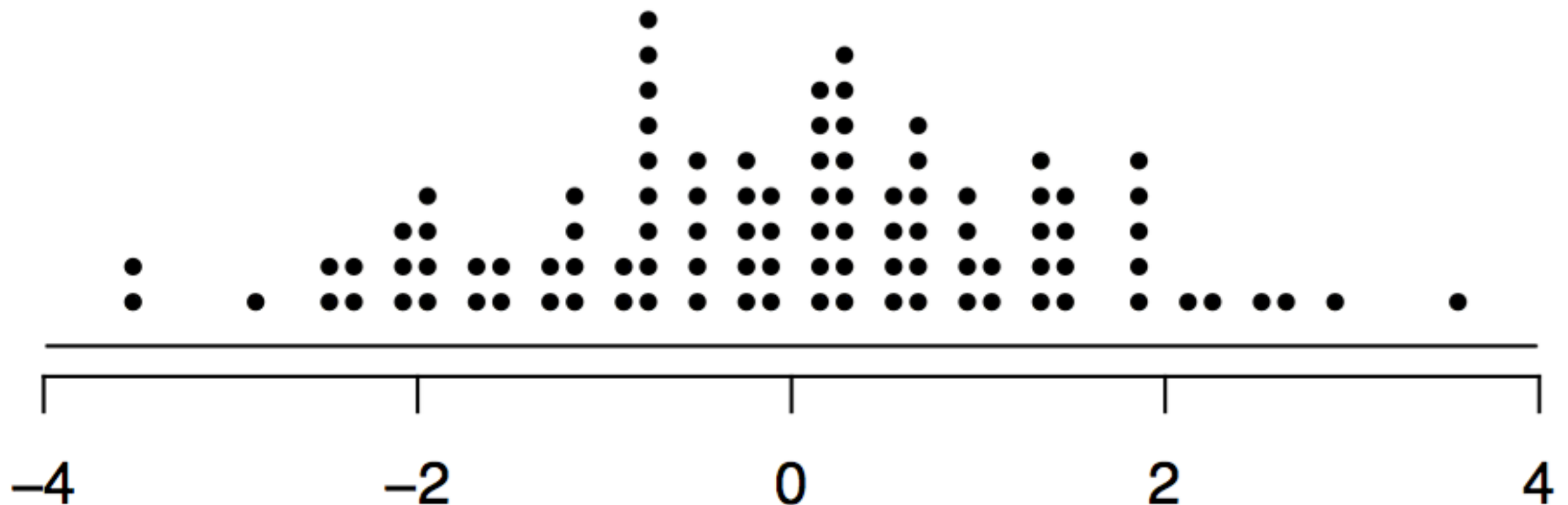
(E) $H_0 : \mu_{\text{caff}} = \mu_{\text{no caff}}; H_A : \mu_{\text{caff}} \neq \mu_{\text{no caff}}$

Simulation scheme

- ▶ On 20 index cards write the tapping rate of each subject in the study.
- ▶ Shuffle the cards and divide them into two stacks of 10 cards each, label one stack “caffeine” and the other stack “no caffeine”.
- ▶ Calculate the average tapping rates in the two simulated groups, and record the difference on a dot plot.
- ▶ Repeat steps (2) and (3) many times to build a randomization distribution.

Below is a randomization distribution of 100 simulated differences in means ($\bar{x}_{\text{caff}} - \bar{x}_{\text{no caff}}$). Calculate the p-value for the hypothesis test evaluating whether caffeine increases average tapping rate.

	Caffeine	No Caffeine	Difference
mean	248.3	244.8	3.5
SD	2.21	2.39	-0.18
median	248	245	3
IQR	3.5	4.25	-0.75



Describe how could we use the same approach to test whether the median tapping rate is higher for the caffeine group?

- ▶ Use the same simulation scheme but record the difference between the medians instead of the means
- ▶ Calculate the p-value as the proportion of simulations where the simulated difference in medians is at least 3.

Below is a randomization distribution of 100 simulated differences in medians ($\text{median}_{\text{caff}} - \text{median}_{\text{no caff}}$). Calculate the p-value for the hypothesis test evaluating whether caffeine increases median tapping rate.

	Caffeine	No Caffeine	Difference
mean	248.3	244.8	3.5
SD	2.21	2.39	-0.18
median	248	245	3
IQR	3.5	4.25	-0.75

