Unit 1: Introduction to data

3. Introduction to statistical inference

Sta 101 - Spring 2015

Duke University, Department of Statistical Science

January 21, 2015

Dr. Çetinkaya-Rundel

Slides posted at http://bitly.com/sta101sp15

1. Housekeeping

2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence

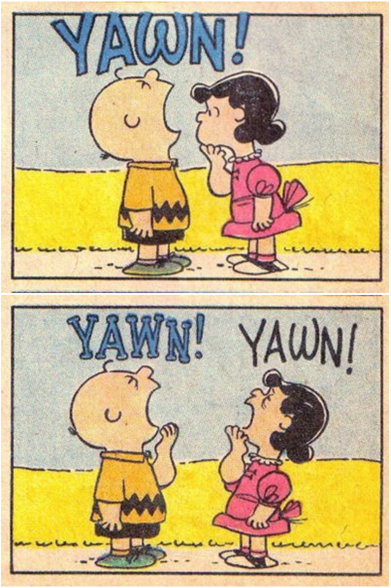
3. Case study: Tapping on caffeine

- PA1 opens after class today, due by 11:59pm on Friday
- RA2 on Monday - review Unit 2 videos + learning objectives
- Feedback from PS1:
  - Random sampling → generalizability, random assignment → causality – lack of a random sample doesn't mean we can't infer causation
  - We covered 3 (good) methods of sampling (SRS, stratified, cluster), there are other methods as well (convenience, volunteer, etc.), they just happen to be not good methods for obtaining representative samples
- Finish up remaining material from last class...

1. Housekeeping

2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence

3. Case study: Tapping on caffeine



Clicker question

Do you think yawning is contagious?

(a) Yes
(b) No
(c) Don't know

An experiment conducted by the MythBusters tested if a person can be subconsciously influenced into yawning if another person near them yawns.

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

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			

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	$\frac{10}{34} = 0.29$		

50 people were randomly assigned to two groups:

- treatment: see someone yawn, $\mathrm{n} = 34$
- control: don't see someone yawn, $\mathrm{n} = 16$

	Treatment	Control	Total
Yawn	10	4	14
Not Yawn	24	12	36
Total	34	16	50
% Yawners	$\frac{10}{34} = 0.29$	$\frac{4}{16} = 0.25$	

50 people were randomly assigned to two groups:

- treatment: see someone yawn, $\mathrm{n} = 34$
- control: don't see someone yawn, $\mathrm{n} = 16$

	Treatment	Control	Total
Yawn	10	4	14
Not Yawn	24	12	36
Total	34	16	50
% Yawners	$\frac{10}{34} = 0.29$	$\frac{4}{16} = 0.25$	

Based on the proportions we calculated, do you think yawning is really contagious, i.e. are seeing someone yawn and yawning dependent?

- The observed differences might suggest that yawning is contagious, i.e. seeing someone yawn and yawning are dependent

- The observed differences might suggest that yawning is contagious, i.e. seeing someone yawn and yawning are dependent
- But the differences are small enough that we might wonder if they might simple be *due to chance*

- The observed differences might suggest that yawning is contagious, i.e. seeing someone yawn and yawning are dependent
- But the differences are small enough that we might wonder if they might simple be *due to chance*
- Perhaps if we were to repeat the experiment, we would see slightly different results

- The observed differences might suggest that yawning is contagious, i.e. seeing someone yawn and yawning are dependent
- But the differences are small enough that we might wonder if they might simple be *due to chance*
- Perhaps if we were to repeat the experiment, we would see slightly different results
- So we will do just that - well, somewhat - and see what happens

- The observed differences might suggest that yawning is contagious, i.e. seeing someone yawn and yawning are dependent
- But the differences are small enough that we might wonder if they might simple be *due to chance*
- Perhaps if we were to repeat the experiment, we would see slightly different results
- So we will do just that - well, somewhat - and see what happens
- Instead of actually conducting the experiment many times, we will *simulate* our results

1. Housekeeping

2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence

3. Case study: Tapping on caffeine

1. "There is nothing going on."
   Seeing someone yawn and yawning are *independent*, observed difference in proportions of yawners in the treatment and control is simply due to chance. → *Null hypothesis*

1. "There is nothing going on."
   Seeing someone yawn and yawning are *independent*, observed difference in proportions of yawners in the treatment and control is simply due to chance. → *Null hypothesis*
2. "There is something going on."
   Seeing someone yawn and yawning are *dependent*, observed difference in proportions of yawners in the treatment and control is not due to chance. → *Alternative hypothesis*

- 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?”

- Start with a *null hypothesis (H_0)* (the status quo)

- Start with a *null hypothesis* (H_0) (the status quo)
- Set an *alternative hypothesis* (H_A) (the research question, i.e. what we're testing for)

- Start with a *null hypothesis (H_0)* (the status quo)
- Set an *alternative hypothesis (H_A)* (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 (today) or theoretical methods (later in the course)

- Start with a *null hypothesis (H_0)* (the status quo)
- Set an *alternative hypothesis (H_A)* (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 (today) or theoretical methods (later in the course)
- Test results suggest the observed data are
  - reasonably likely to have occurred even if the null is true → stick with the null hypothesis (fail to reject H_0)
  - unlikely to have occurred even if the null is true → reject H_0 in favor of H_A

- Start with a *null hypothesis (H_0)* (the status quo)
- Set an *alternative hypothesis (H_A)* (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 (today) or theoretical methods (later in the course)
- Test results suggest the observed data are
  - reasonably likely to have occurred even if the null is true → stick with the null hypothesis (fail to reject H_0)
  - unlikely to have occurred even if the null is true → reject H_0 in favor of H_A
- Never declare the null hypothesis to be true, because we simply do not know whether it's true or not → never "accept the null hypothesis"

1. Housekeeping
2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence
3. Case study: Tapping on caffeine

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

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

- Results from the simulations based on the *null hypothesis* look like the data → difference between the proportions of yawners in the treatment and control groups was simply *due to chance* (yawning is not contagious)

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

- Results from the simulations based on the *null hypothesis* look like the data → difference between the proportions of yawners in the treatment and control groups was simply *due to chance* (yawning is not contagious)
- Results from the simulations based on the chance model do not look like the data → difference between the proportions of yawners in the treatment and control groups was not due to chance, but *due to an actual effect of seeding* (yawning is contagious)

- 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.

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

Why shuffle 7 times: *http://www.dartmouth.edu/~chance/course/topics/winning_number.html*

1. Housekeeping

2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence

3. Case study: Tapping on caffeine

Clicker question

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

1. Housekeeping

2. Case study: is yawning contagious?
    1. Competing claims
    2. Testing via simulation
    3. Checking for independence

3. Case study: 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

Clicker question

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) Mosaic plot
(c) Pie chart
(d) Side-by-side box plots
(e) Single box plot

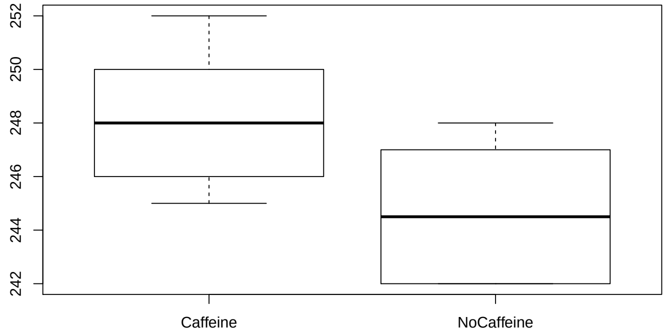
Clicker question

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) Mosaic plot
(c) Pie chart
(d) *Side-by-side box plots*
(e) Single box plot

Compare the distributions of tapping rates in the caffeine and no caffeine groups.

	Caffeine	No Caffeine	Differen
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



Clicker question

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

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

Clicker question

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

(a) $H_0 : \mu_{caff} = \mu_{no\ caff}$
$H_A : \mu_{caff} < \mu_{no\ caff}$

(b) $H_0 : \mu_{caff} = \mu_{no\ caff}$
$H_A : \mu_{caff} > \mu_{no\ caff}$

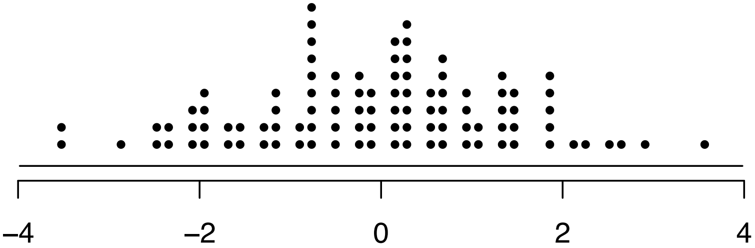
(c) $H_0 : \bar{x}_{caff} = \bar{x}_{no\ caff}$
$H_A : \bar{x}_{caff} > \bar{x}_{no\ caff}$

(d) $H_0 : \mu_{caff} > \mu_{no\ caff}$
$H_A : \mu_{caff} = \mu_{no\ caff}$

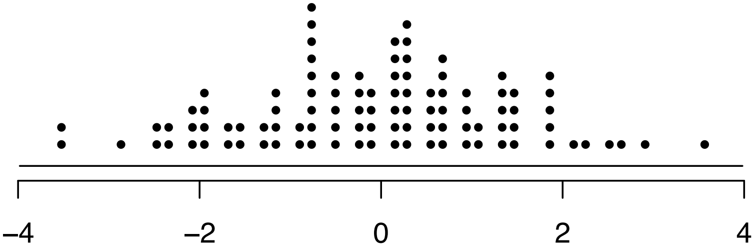
(e) $H_0 : \mu_{caff} = \mu_{no\ caff}$
$H_A : \mu_{caff} \neq \mu_{no\ caff}$

- 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*.

Calculate the p-value based on the randomization distribution below and determine the conclusion of the hypothesis test. (100 simulations)



Calculate the p-value based on the randomization distribution below and determine the conclusion of the hypothesis test. (100 simulations)



1/100 = 0.01

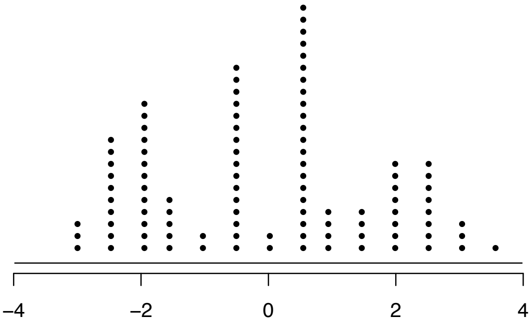
Describe how could we use the same approach to test whether the median tapping rate is higher for the caffeine group?

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, and calculate the p-value as the proportion of simulations where the simulated difference in medians is at least 3.

Using the randomization distribution below of simulated differences in means, determine whether the data provide convincing evidence that caffeine increases median tapping rate.

	Caffeine	No Caffeine	Difference
median	248.0	245.5	3.5



Application exercise: 1.4 Randomization testing

See the course website for instructions.