Unit 1: Introduction to data
Lecture 3: EDA (cont.) and Introduction to statistical inference via simulation

Statistics 101

Mine Çetinkaya-Rundel

September 5, 2013

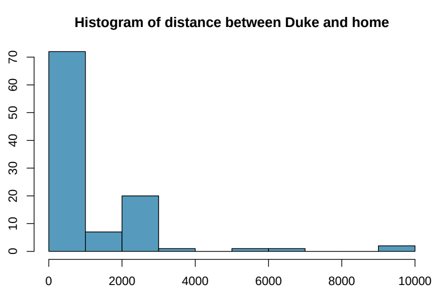
Announcements

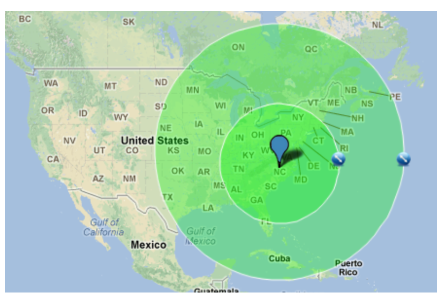
- Performance assessment (PA) 1 will be available at 5pm tonight and is due by 5pm tomorrow evening.
- Readiness assessment (RA) 2 next Tuesday - Unit 2 resources available on the course website.
- Problem set (PS) 1 due next Thursday - must submit electronically on Sakai by course start time.
- Make sure to complete your clicker registration by tomorrow. If you're still waiting on a clicker, register before class on Tuesday at the latest - we'll start recording clicker data on Tuesday.

Typical observation

How far is the *typical* student's home from Duke?

mean = 1250 miles median = 600 miles





http://www.freemaptools.com/radius-around-point.htm

Robust statistics

Since the median and IQR are more robust to skewness and outliers than mean and SD:

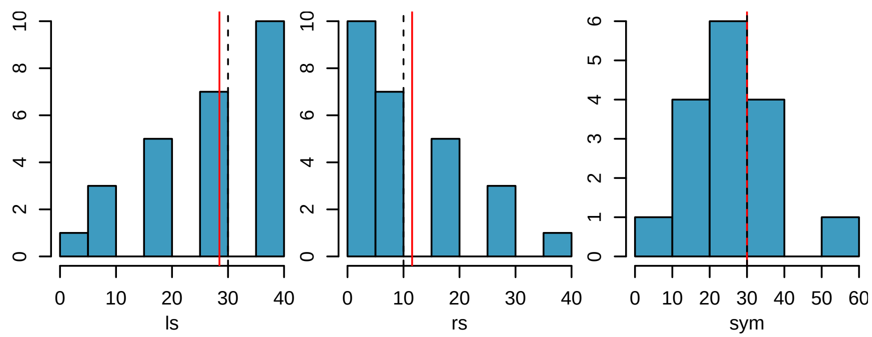
- skewed → median and IQR
- symmetric → mean and SD

If you were searching for a car, and you are price conscious, would you be more interested in the mean or median vehicle price when considering a car?

Mean vs. median

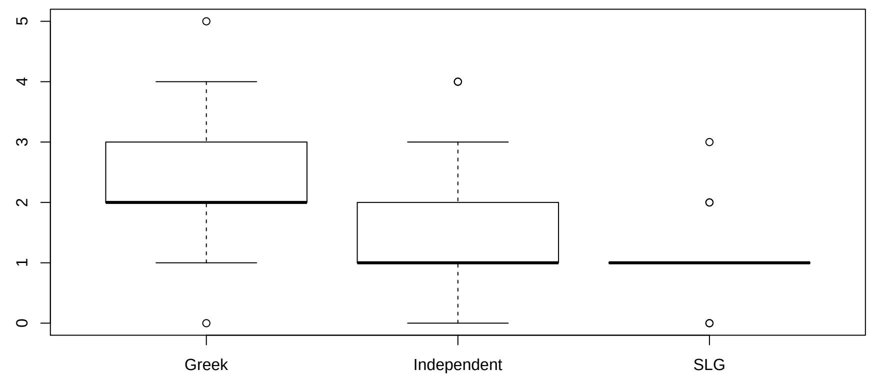
- If the distribution is symmetric, center is the mean
  - Symmetric: mean is roughly equal to the median
- If the distribution is skewed or has outliers center is the median
  - Right-skewed: mean is likely greater than the median
  - Left-skewed: mean is likely less than the median

red solid - mean, black dashed - median



Side-by-side box plot

How does the number of the average number of times students go out per week vary by involvement? Do the two variables appear to be associated or independent?



Gender discrimination

- In 1972, as a part of a study on gender discrimination, 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as “routine”.
- The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female.
- It was randomly determined which supervisors got “male” applications and which got “female” applications.
- Of the 48 files reviewed, 35 were promoted.
- The study is testing whether females are unfairly discriminated against.

Is this an observational study or 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 relatonship between promotion and gender?

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

% of males promoted: $21/24 = 0.875$
% of females promoted: $14/24 = 0.583$

Clicker question

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.

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

Hypothesis testing is very much like a court trial.

- 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?"

- Evidence not strong enough to reject the assumption of innocence → verdict: 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.

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We 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.
  - 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".

Simulating the experiment...

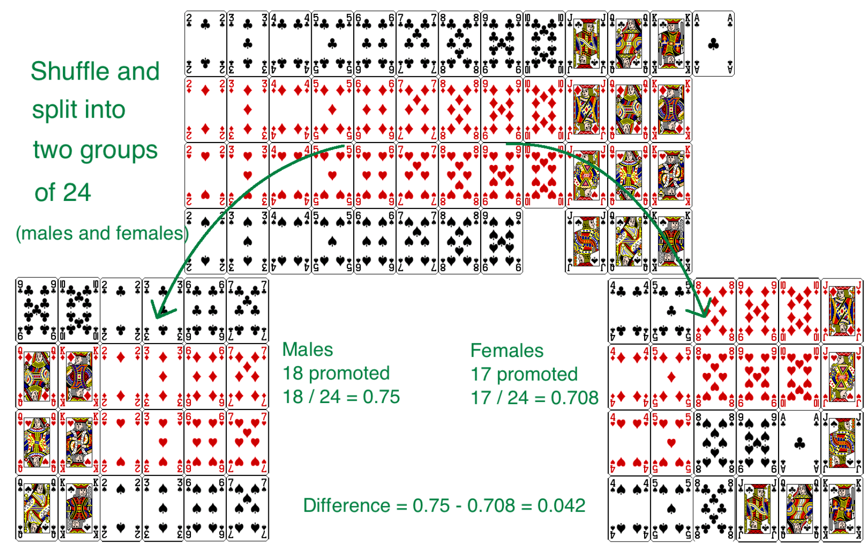
- Simulate the experiment in a way that satisfies the null hypothesis (in this case, in a way that there is no discrimination against females)
- Determine if the observed outcome from the original experiment (roughly 30% more males being promoted) is a likely outcome when things are left up to *chance*.
- If the results from the simulations based on the chance model do not look like the data, determine that the observed difference between males and females was *due to an actual effect of gender* (promotion and gender are dependent).

Simulation setup

1. We'll let a face card represent *not promoted* and a non-face card represent a *promoted*. Consider aces as face cards.
   - Set aside the jokers.
   - Take out 3 aces → there are exactly 13 face cards left in the deck (face cards: A, K, Q, J): NOT PROMOTED
   - Take out a number card → there are exactly 35 number (non-face) cards left in the deck (number cards: 2-10): PROMOTED
2. Shuffle the cards and deal them into two groups of size 24, representing males and females.
3. Count and record how many files in each group are promoted (number cards).
4. Calculate the proportion of promoted files in each group and take the difference (male - female).
5. Report the difference using your clicker (up to 2 decimal places, only 1 submission for each simulation).

Step 1

Step 2 - 4



Clicker question

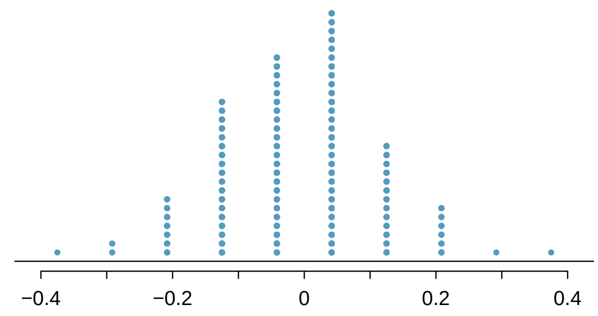
Do the data provide convincing evidence of gender discrimination against women, i.e. dependence between gender and promotion decisions?

(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.

Making a decision

- The probability of observing a difference at least as favorable to the alternative hypothesis as the one observed in the original data (a difference of 29.2%) if H_0 is true is called the *p-value*.
- The *significance level* is the threshold against which we compare the p-value to determine if it's small enough to reject the null hypothesis (this is usually 5%).



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).

Data

	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

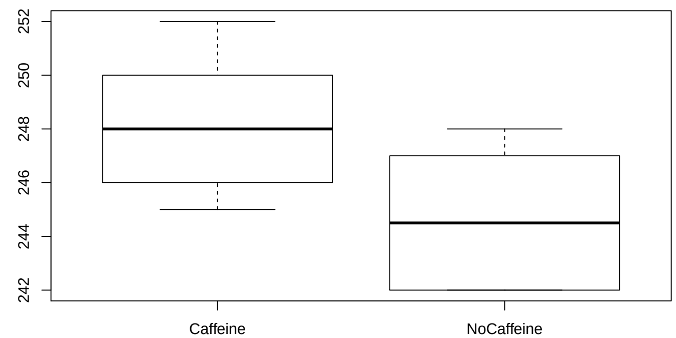
What type of plot would be useful to visualize the distributions of tapping rate in the caffeine and no caffeine groups.

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

Exploratory data analysis

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

	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



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

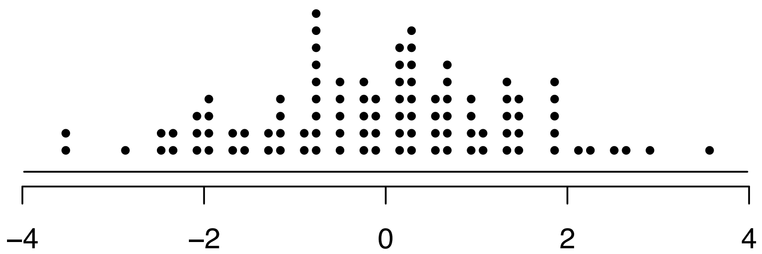
(e) $H_0 : \mu_{caff} = \mu_{no\ caff}$
$H_A : \mu_{caff} \neq \mu_{no\ caff}$

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*.

Making a decision

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



Testing for the median

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

Testing for the median (cont.)

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