DATA 8
Spring 2021

Discussion 9

(Optional) AB-Testing

Materials: tinyurl.com/d8-disc09
or access through kevin-miao.com under teaching

Today

- Announcements
- Midterm Feedback
- Worksheet
  - Link: www.tinyurl.com/d8-disc09

Announcements

- **No vitamin discussion question**
- **Cancelled this week**
  - Office Hours
    - Instructor OH are not cancelled
  - Labs
    - No lab this week
- Midterm grades will be released together with mid-semester report

Midterm Feedback 👍👎

Please complete the poll!

Any other thoughts?

- Were there any questions that were too difficult in particular?
- Did the material prepare you well enough?
- Did we focus too much on one topic?
- Were our instructions clear enough?
- Any other comments?

A/B Testing 🎭

Data 8 Spring 2021
Discussion 09: A/B Testing

0. What has been your favorite topic/assignment/lecture/anything so far with half of the class done?

One special kind of hypothesis test we do in this class is called an A/B test. The steps used to run an A/B test are the same as a general hypothesis test, but A/B tests have a specific null hypothesis (that two samples were drawn from the same distribution), which we test by performing a *permutation*.

1. When should you use an A/B test vs another kind of hypothesis test?

2. Caddisflies
Natalia, a museum curator, has recently been brought specimens of caddisflies collected from various parts of Northern California. The scientists who collected the caddisflies think that caddisflies collected at higher altitudes tend to be bigger. They tell her that the average length of the 560 caddisflies collected at high elevation is 14mm, while the average length of the 450 caddisflies collected slightly lower down is 12mm. She's not sure that this difference really matters, and thinks that this could just be the result of chance in sampling.

a) What's the null hypothesis Natalia can test?

b) How could you test the null hypothesis in the A/B test from above? What assumption would you make to test the hypothesis, and how would you simulate under that assumption?

c) What would be a useful test statistic for the A/B test?

d) Assume `flies` refers to the following table:

Elevation	**Specimen length**
High elevation	12.3
Low elevation	13.1
High elevation	12.0

...
(1007 rows omitted)

Fill in the blanks in this code to generate <u>one value of the test statistic under the null hypothesis.</u>

```
def permutation_test():
      shuffled_labels = flies._________________.column('Elevation')

      shuffled_flies =
      flies.drop('Elevation').with_columns(___________, ___________)

      grouped = shuffled_flies._________(_______,_______)

      means = grouped.column('Specimen length mean')
      statistic = ___________________________________
      return statistic
```

e) Fill in the code below to simulate 10000 trials of our permutation test

```
test_stats = ___________________________________
repetitions = ___________

for i in np.arange(___________):

      one_stat = _____________________
      test_stats = ____________________________________________

test_stats
```

f) Given that the p-value of the test was 0.1, draw a possible histogram of the test statistics

g) If the p value cutoff we use is 7.5%, what is the conclusion of our test? With a p value cutoff of 7.5%, if we ran this same hypothesis test 1000 times, assuming the null hypothesis is true, how many times would we expect to incorrectly reject the null hypothesis?

To the worksheet! ✍️

tinyurl.com/d8-disc09

Happy Spring Break! ☀️

Psssst, leave some feedback behind
https://tinyurl.com/kevind8feedback