DATA 8
Spring 2021

Discussion 8

Hypothesis Testing and MT Review

Materials: tinyurl.com/d8-disc08
or access through kevin-miao.com under teaching

Today

- Announcements
- (30 min) Worksheet
  - Link: www.tinyurl.com/d8-disc08
- (15 min) Midterm Q&A

Announcements

- **Midterm** on **Friday (3/12)** from **7-9PM**
  - Create a piazza post if you haven't received an email and …
    - you are in DSP
    - you signed up to take the alternate exam (only int'l students)
- Assignments
  - **Vitamin 7** is due tonight
  - **Homework 7** is due Thursday
- **No vitamin question today** but **attendance will be taken**
- **Regrades** are **due Friday**
  - Gradescope: Submit regrade via button
  - OkPy: Email me

To the worksheet! ✍️

tinyurl.com/d8-disc08

Hypothesis Testing – 7, 5, 6

Question 7. List out the steps used in the process of hypothesis testing.

1	Null and alternative Hypothesis
2	Test Statistic
3	simulate the data many times
4	calculate observed
5	Calculate the p-value; conclude test (compare against p-val cutoff)

Question 5. What is the difference between an AB test and a hypothesis test?

AB testing is a specific kind of hypothesis testing; we want to see whether two distributions come from the same underlying distribution.

Question 6. How do you simulate the null hypothesis for an AB test?

permutation test: shuffling the labels.

↳ sample without replacement.

Ex:

Smoker / non smoker	Baby birth weight
Smoker	..
non smoker	..

Procedure
(data shuffling)

rolling a dice

1	2	3	4	5	6
1/10	2/10	3/10	1/10	2/10	1/10

Some kind of ~~test statistic~~ to test categorical distribution against a theoretical distribution.

AB Testing != TVD

- we see two distributions and see if they are from the same underlying one
- we see one distribution

Test Statistic

- We design a test statistic based on the context that has been given.
  - **Test whether a coin is biased**

    $|\#\text{ of heads} - 5|$
  - **Test whether a coin is biased towards heads**

    $\#\text{ of heads}$
- Generally, we want to pick a statistic that
  - is **small** when it is close to the **null**
  - is **large** when it is close to the **alternative**

Designing Statistics – 8

Question 8. For each of the following parts, you will be given a specific scenario, how would you each calculate the test statistic for each of the following. You don't need to write out the full code, just explain which main randomization tools you would need.

a) Drawing 3 marbles from a bag with ½ red marbles, ¼ blue marbles, and ¼ green

One category

↳ sample proportions

given a model - proportions!

b) A table where each row is a Data 8 student with either prior coding experience or no prior coding experience

↳ AB testing → shuffle

c) The number of times you choose the number 3 in a 7-digit phone number (not including the area code digits)

↳ Check the count and use np.random.choice

P-values

- **P-value**
  - the probability that under the null hypothesis we see values that are equal to or more in the direction of the alternative than the observed statistic
- **P-value cutoff**
  - We can decide which p-value is too extreme and determine whether to **reject** or **support** either **hypotheses**
  - **Note:** we never say a hypothesis is correct or incorrect
- **What does that mean if we use a 5% p-value cutoff?**
  - 5% of the times, we support the alternative wrongly / reject the null

Question 1. On last week's worksheet we proposed two models for how Francie's coin works, the first model was that the coin is fair and the second model was that the coin is not fair. Which model corresponds to the null hypothesis and which model corresponds to the alternative hypothesis? Formally state the hypotheses. As a reminder, Francie flipped a coin 10 times and observed 9 heads. Francie is trying to determine if the coin is fair.

Null : coin is fair. 50/50 heads tails. any deviation is due to chance.

alternative : coin is unfair.

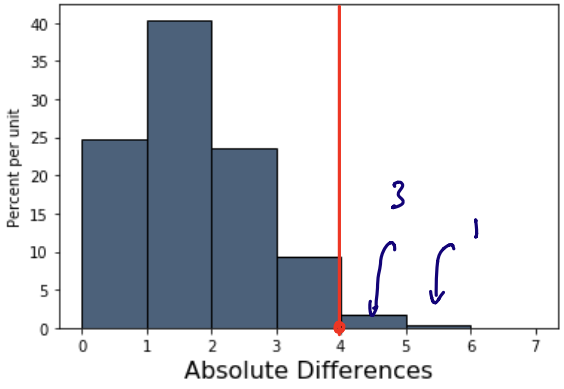
Question 2. We presented the following histogram of simulated values of the test statistic last week. Calculate the empirical p value for the test, and write a line of code to calculate that same value. The simulated test statistics are stored in an array called `abs_differences`.

The heights of the bins are 24, 40, 23, 9 and 3, and 1 percent per unit from left to right.

Observed =

|# of heads - 5| =

|9 - 5| = 4



↳ 4% = (3% + 1%)

Question 3. If we use a p value cutoff of 5% what is the conclusion of our test? What if we use a p value cutoff of 1%?

4% = obs < 5% cutoff → Reject Null

Question 4. Instead of using absolute difference as the test statistic, the test statistic is now the number of heads flipped in 10 flips. What alternative hypothesis would be associated with each of the p value calculations?

```
np.count_nonzero(num_heads >= 9) / len(num_heads)
```

Biased towards head

```
np.count_nonzero(num_heads <= 9) / len(num_heads)
```

Biased towards tails

num. heads <= 1

↳ would be better.

Q&A: Midterm

You got this!

End of Section
How did I do?

https://tinyurl.com/kevind8feedback