Lab 07: A/B Testing

Data 8 Discussion Worksheet

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). We carry out this test by performing a *permutation* of our data.

# Mid-semester Check In

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

*If you have any concerns about your performance in the class so far, feel free to bring it up to your lab TA.*

# A/B Testing and Error Probabilities

1. **Warmup:** When should you use an A/B test versus another kind of hypothesis test?

Two distribution use A/B test. 1 Distribution use hypothesis test.

You should use A/B testing when determining whether two samples also known A group and B group were Sampled from the same underlying population/distribution.

Kevin, a museum curator, has recently been given 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 him that the average length of the 560 caddisflies collected at high elevation is 14mm, while the average length of the 450 caddisflies collected from a slightly lower elevation is 12mm. He’s not sure that this difference really matters, and thinks that this could just be the result of chance in sampling.

1. What’s an appropriate null hypothesis that Kevin can simulate under?

The distribution of speciman lengths is the same for coaddleflies sampled from highelevation as those sampled from low elevation.

1. 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?

Take the sample from from high elevation and low elevation. Then shuffle. If the data is around the same, then it is null hypothesis.

**TEACHER RESPONSE:**

If null hypothesis is true the flies did not come from different distributions – them it shouldn’t matter how the sampels were labels(high elevation or low elevation). Under this assumptuin yuou could shuffle the labels of the flies and calculate your test statistic from this relabeled data.

Shuffle is too prove it is null hypothesis.

1. What would be a useful test statistic for the A/B test? Remember that the *direction* of your test statistic should come from the initial setting.

Find the average.

1. 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 one value of the test statistic under the null hypothesis.

def one\_simulation():

shuffled\_labels = flies.shuffle().column(‘Elevation’)

shuffled\_flies = flies.drop(‘Elevation’).with\_columns(“Shuffled Elevation”),

shuffled\_labels))

grouped = shuffled\_flies.group(“Shuffled Elevation”, np.mean)

means = grouped.column(‘Shuffled Elevation”)

statistic = means[1]-means[0]

return statistic

1. Fill in the code below to simulate 10000 trials of our permutation test.

test\_stats = make\_array

repetitions = 10000

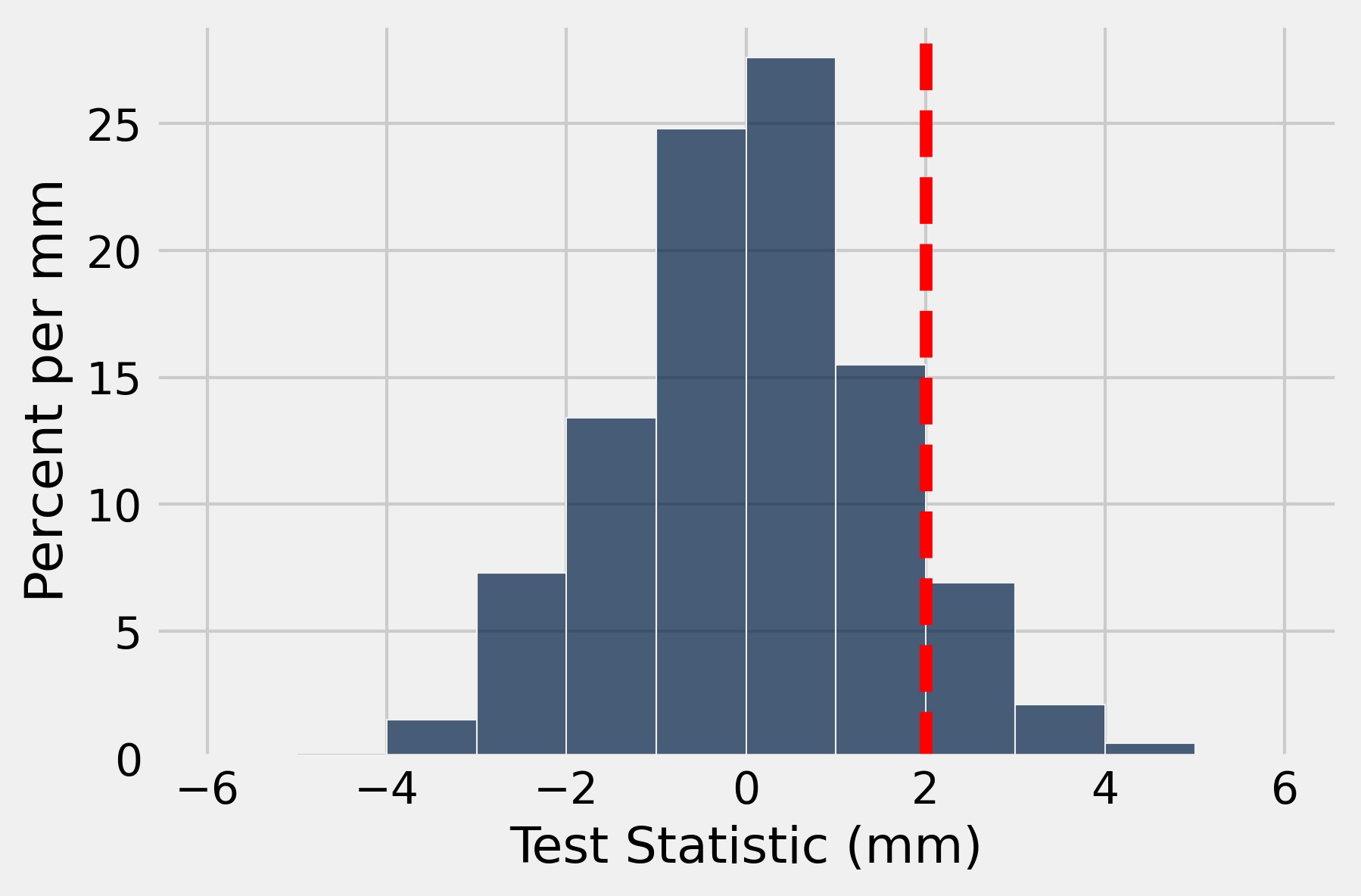
for i in np.arange(10000):

one\_stat = one\_simulation()

test\_stats = np.append(test\_stats, one\_simulation())

test\_stats

1. The histogram of test\_stats is plotted below with a vertical red line indicating the observed value of our test statistic. If the p-value cutoff we use is 5%, what is the conclusion of our test?



1. Suppose that the null hypothesis is true. If we ran this same hypothesis test 1000 times, each time drawing a new random sample from the population and with a p-value cutoff of 5%, how many times would we expect to *incorrectly* reject the null hypothesis?

If right is greater than 5 % then we failed to reject null hypothesis.

We failed reject the null hypothesis because right side is greater than 5 %.

**We can inspect the histogram, to see theat the area to the right of the observed value is greater than 5%. Since our p\_value is greater than our p value cutoff(5%) we faile to reject the null hypothesis and conclude that the data tend to favor null hypothesis.**

1. What effect does *decreasing* our p-value cutoff have on the number of times we expect to *incorrectly reject* the null hypothesis?

If we decrease our p\_value cut off we are reducing the expected number of times we will incorrectly reject the null.