lab07

January 25, 2024

```
[40]:  # Initialize Otter
       import otter
       grader = otter.Notebook("lab07.ipynb")
```

1 Lab 7: Great British Bake Off (A/B Test)

Welcome to Data 8 Lab 7! This week's lab will focus on A/B Testing using data from the ever-popular British television show, *The Great British Bakeoff*.

Submission: Once you're finished, run all cells besides the last one, select File > Save Notebook, and then execute the final cell. Then submit the downloaded zip file, that includes your notebook, according to your instructor's directions.

First, set up the notebook by running the cell below.

```
[41]:  # Run this cell to set up the notebook, but please don't change it.

       # These lines import the Numpy and Datascience modules.
       import numpy as np
       from datascience import *

       # These lines do some fancy plotting magic.
       import matplotlib
       %matplotlib inline
       import matplotlib.pyplot as plots
       plots.style.use('fivethirtyeight')
       import warnings
       warnings.simplefilter('ignore', (FutureWarning, np.VisibleDeprecationWarning))

       import d8error
```

1.1 1. A/B Testing

A/B testing is a form of hypothesis testing that allows you to make comparisons between two distributions. We may also refer to an A/B test as a permutation test.

You'll almost never be explicitly asked to perform an A/B test. Make sure you can identify situations where the test is appropriate and know how to correctly implement each step. Oftentimes,

we use an A/B test to determine whether or not two samples came from the same underlying distribution.

Question 1.1. The following statements are the steps of an A/B hypothesis test presented in a *random order*:

1. Choose a test statistic (typically the difference in means between two categories)

2. Shuffle the labels of the original sample, find your simulated test statistic, and repeat many times

3. Find the value of the observed test statistic

4. Calculate the p-value based off your observed and simulated test statistics

5. Define a null and alternate model

6. Use the p-value and p-value cutoff to draw a conclusion about the null hypothesis

Assign `ab_test_order` to an array of integers that contains the correct order of an A/B test, where the first item of the array is the first step of an A/B test and the last item of the array is the last step of an A/B test.

```
[42]: ab_test_order = make_array(5, 1, 3, 2, 4, 6)
```

```
[43]: grader.check("q1_1")
```

```
[43]: q1_1 results: All test cases passed!
```

Question 1.2. If the null hypothesis of an A/B test is correct, should the order of labels affect the differences in means between each group? Why do we shuffle labels in an A/B test? If you are in a lab section, confirm your answer with a neighbor or staff member before moving on.

Type your answer here, replacing this text.

1.2 2. The Great British Bake Off

"The Great British Bake Off (often abbreviated to Bake Off or GBBO) is a British television baking competition, produced by Love Productions, in which a group of amateur bakers compete against each other in a series of rounds, attempting to impress a group of judges with their baking skills" Wikipedia

For every week of the competition, the judges assign one contestant the title "Star Baker". Ultimately, one winner is crowned every season. Using this information, we would like to investigate how winning Star Baker awards affects the odds of winning a season of the show.

Question 2.1. We want to know whether winning more Star Baker awards *causes* a change in likelihood of winning the season. Why is it not sufficient to compare star baker rates for winners and losers?

Type your answer here, replacing this text.

1.2.1 Running an Experiment

We are going to run the following hypothesis test to determine the association between winning and number of Star Baker awards. The population we are examining is every contestant from seasons 2 through 11 of GBBO. We are going to use the following null and alternative hypotheses:

Null hypothesis: The distribution of Star Baker awards between contestants who won their season and contestants who did not win their season is the same.

Alternative hypothesis: Contestants who win their season of the show will win more Star Baker awards on average.

Our alternative hypothesis is related to our suspicion that contestants who win more Star Baker awards are more skilled, so they are more likely to win the season.

Question 2.2. Should we use an A/B test to test these hypotheses? If yes, what is our "A" group and what is our "B" group?

Type your answer here, replacing this text.

Check your answers with your neighbors or a staff member before you move on to the next section.

The `bakers` table below describes the number of star baker awards each contest won and whether or not they won their season (`1` if they won, `0` if they did not win). The data was manually aggregated from Wikipedia for seasons 2-11 of the show. We randomized the order of rows as to not spoil the outcome of the show.

```
[44]: bakers = Table.read_table("star_bakers.csv")
      bakers.show(3)
```

```
<IPython.core.display.HTML object>
```

Question 2.3. Create a new table called `means` that contains the mean number of star baker awards for bakers who did not win (`won==0`) and bakers that did win (`won==1`). The table should have the column names `won` and `star baker awards mean`.

```
[45]: means = bakers.group('won', np.mean)
      means
```

```
[45]: won  | star baker awards mean
      0    | 0.651786
      1    | 1.5
```

```
[46]: grader.check("q2_3")
```

```
[46]: q2_3 results: All test cases passed!
```

Question 2.4. Visualize the distribution of Star Baker awards for winners and non-winners. You should use the bins we provided.

Hint: You will want to use the group argument of `tbl.hist`. In order to produce several overlayed histograms based on unique values in a given column, we can do something like `tbl.hist(...,` `group=<col_name>, bins=...)`!

```
[47]:  useful_bins = np.arange(0, 7)
       bakers.hist('star baker awards', group='won', bins=useful_bins)
```
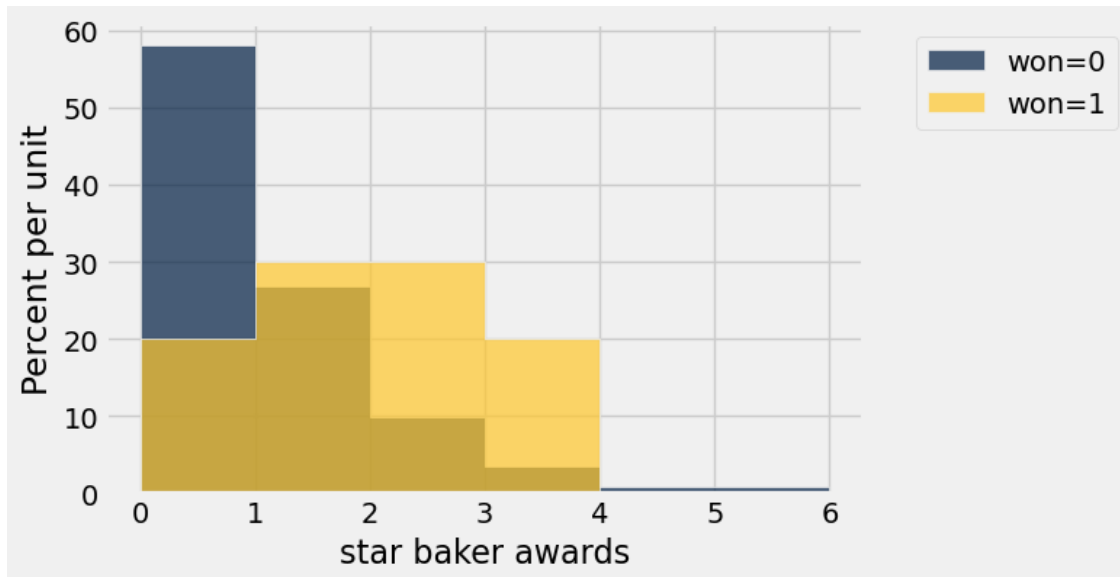

Question 2.5. We want to figure out if there is a difference between the distribution of Star Baker awards between winners and non winners.

What should the test statistic be? Which values of this test statistic support the null, and which values support the alternative?

If you are in lab, confirm your answer with a neighbor or staff member before moving on.

Hint: You should think about what measures we use to describe a distribution.

Type your answer here, replacing this text.

Question 2.6. Set `observed_difference` to the observed test statistic using the `means` table.

```
[48]:  observed_difference = means.column('star baker awards mean').item(1) - means.
       ↪column('star baker awards mean').item(0)
       observed_difference
```

```
[48]:  0.8482142857142857
```

```
[49]:  grader.check("q2_6")
```

```
[49]:  q2_6 results: All test cases passed!
```

Question 2.7. Given a table like `bakers`, a label column `label_col`, and a values column `val_col`, write a function that calculates the appropriate test statistic.

Hint: Make sure that you are taking the directionality of our alternative hypothesis into account.

```
[50]: def find_test_stat(tbl, label_col, val_col):
          new_table = tbl.group(label_col, np.average)
          return new_table.column(val_col+" average").item(1)-new_table.
      ↪column(val_col+" average").item(0)

      find_test_stat(bakers, "won", "star baker awards")
```

[50]: 0.8482142857142857

```
[51]: grader.check("q2_7")
```

[51]: q2_7 results: All test cases passed!

When we run a simulation for A/B testing, we resample by **shuffling the labels** of the original sample. If the null hypothesis is true and the star baker award distributions are the same, we expect that the difference in mean star baker awards will be not change when `"won"` labels are changed.

Question 2.8. Write a function `simulate_and_test_statistic` to compute one trial of our A/B test. Your function should run a simulation and return a test statistic.

```
[52]: def simulate_and_test_statistic(tbl, labels_col, values_col):
          shuffled_column = tbl.sample(with_replacement=False).column(values_col)
          new_table = tbl.drop(values_col).with_column('shuffled label',␣
      ↪shuffled_column)
          return find_test_stat(new_table, labels_col, 'shuffled label')

      simulate_and_test_statistic(bakers, "won", "star baker awards")
```

[52]: 0.3035714285714286

```
[53]: grader.check("q2_8")
```

[53]: q2_8 results: All test cases passed!

Question 2.9. Simulate 5000 trials of our A/B test and store the test statistics in an array called `differences`.

```
[54]: # This cell might take a couple seconds to run
      differences = make_array()
      for i in np.arange(5000):
          test = simulate_and_test_statistic(bakers, 'star baker awards', 'won')
          differences = np.append(differences, test)


      differences
      len(differences)
```
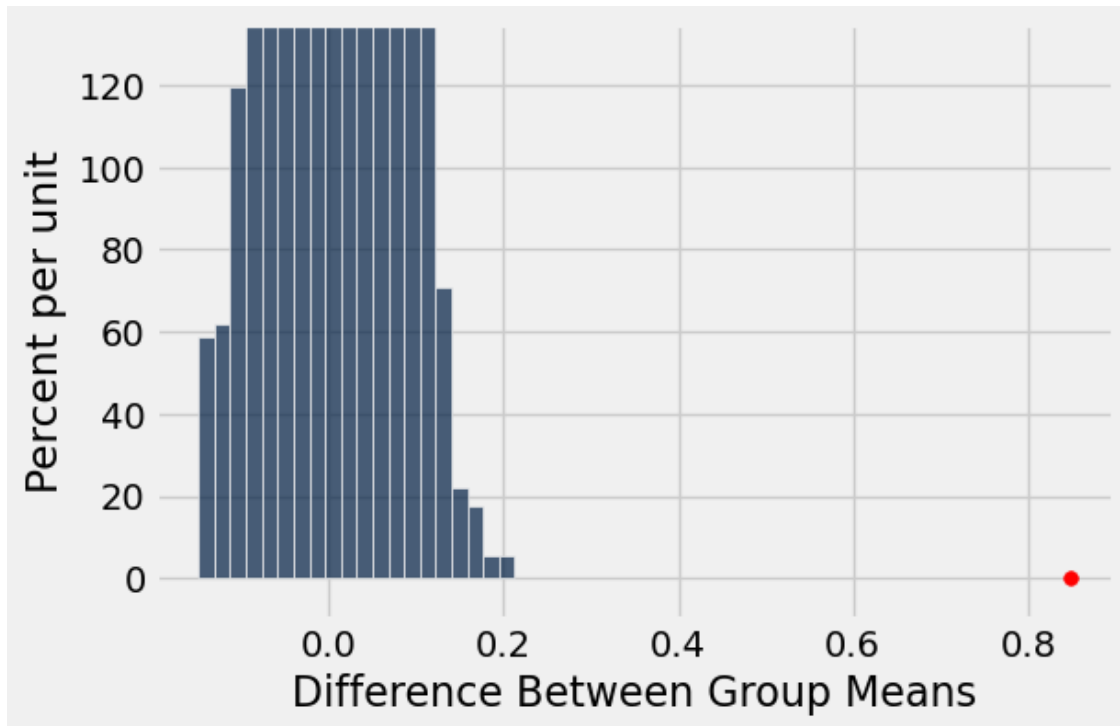
[54]: 5000

5

```
[62]: grader.check("q2_9")
```

[62]: q2_9 results: All test cases passed!

Run the cell below to view a histogram of your simulated test statistics plotted with your observed test statistic.

```
[63]: Table().with_column('Difference Between Group Means', differences).hist(bins=20)
      plots.scatter(observed_difference, 0, color='red', s=30, zorder=2)
      plots.ylim(-0.1, 1.35);
```


Question 2.10. Find the p-value for your test and assign it to `empirical_p`.

```
[64]: empirical_p = sum(differences >= observed_difference) / len(differences)
      empirical_p
```

[64]: 0.0

```
[65]: grader.check("q2_10")
```

[65]: q2_10 results: All test cases passed!

Question 2.11. Using a 5% P-value cutoff, draw a conclusion about the null and alternative hypotheses. Describe your findings using simple, non-technical language. What does your analysis

tell you about the association between star baker awards and winning? What can you claim about causation from your statistical analysis? Call your GSI or AI over and confirm your answer.

Type your answer here, replacing this text.

```
[66]:  grader.check("q2_10")
```

```
[66]:  q2_10 results: All test cases passed!
```

1.3   3. Submission

Minnie wanted to say congrats on finishing Lab 7!! You can now be like Minnie and enjoy your spring break!

Important submission steps: 1. Run the tests and verify that they all pass. 2. Choose **Save Notebook** from the **File** menu, then **run the final cell**. 3. Click the link to download the zip file. 4. Then submit the zip file to the corresponding assignment according to your instructor's directions.

It is your responsibility to make sure your work is saved before running the last cell.

1.4   Submission

Make sure you have run all cells in your notebook in order before running the cell below, so that all images/graphs appear in the output. The cell below will generate a zip file for you to submit. **Please save before exporting!**

```
[67]:  # Save your notebook first, then run this cell to export your submission.
       grader.export(pdf=False, run_tests=True)
```

Running your submission against local test cases…

Your submission received the following results when run against available test cases:

    q1_1 results: All test cases passed!

    q2_3 results: All test cases passed!

    q2_6 results: All test cases passed!

    q2_7 results: All test cases passed!

    q2_8 results: All test cases passed!

    q2_9 results: All test cases passed!

    q2_10 results: All test cases passed!

```
<IPython.core.display.HTML object>
```