
Quiz 6 - DSC 10, Winter 2024

Full Name:

PID:

Discussion: ☐ A (5PM) ☐ B (6PM) ☐ C (7PM)

Instructions:

- This quiz consists of 2 questions. You have a total of 20 minutes to complete it.
 - Please write **clearly** in the provided answer boxes; we will not grade work that appears elsewhere. Completely fill in bubbles and square boxes; if we cannot tell which option(s) you selected, you may lose points.
 - ☐ A bubble means that you should only **select one choice**.
 - ☐ A square box means you should **select all that apply**.
 - If your answer is a string, make sure to put it in quotes. If your answer is a float, make sure to include a decimal point.
 - No aids are allowed (no notes, no calculators, and no computers).
-

By signing below, you are agreeing that you will behave honestly and fairly during and after this quiz.

Signature:

Version A

Please do not open your quiz until instructed to do so.

It can be hard to find a parking spot on UCSD's campus! The `parking` DataFrame contains UCSD parking occupancy data for two on-campus parking structures. The `"Structure"` column contains either `"Gilman"` or `"Hopkins"`. Each row of `parking` represents one day. The `"Occupancy"` column contains a float representing the proportion of occupied spaces at noon on that day. We'll use this data to test the following hypotheses:

- **Null Hypothesis:** At noon, Gilman and Hopkins are equally occupied. The observed differences in our samples are simply due to random chance.
- **Alternative Hypothesis:** At noon, Hopkins is less occupied than Gilman. The observed differences in our samples cannot be explained by random chance alone.

As our test statistic, we will use the mean noontime `"Occupancy"` of Hopkins minus the mean noontime `"Occupancy"` of Gilman.

Question 1

- a) Suppose the Series `s` is defined as below. Write an expression involving `s` that evaluates to the observed value of the test statistic, and store the result in `observed`.

```
s = parking.groupby("Structure").mean().get("Occupancy")
observed = _____
```

What goes in the blank?

- b) In running the permutation test, we need to do a simulation that runs many times, using a `for`-loop. What should be the first thing we do inside the `for`-loop?
- ☐ Initialize an empty array to store our results.
 - ☐ Define a variable for the number of repetitions.
 - ☐ Permute one of the columns of `parking`.
 - ☐ Calculate the difference in group means.
- c) Suppose we store 5000 simulated test statistics in the array `differences`. Choose the appropriate symbol to fill in the calculation of the p-value below.

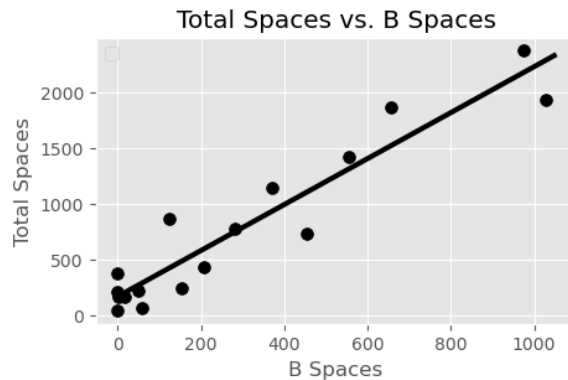
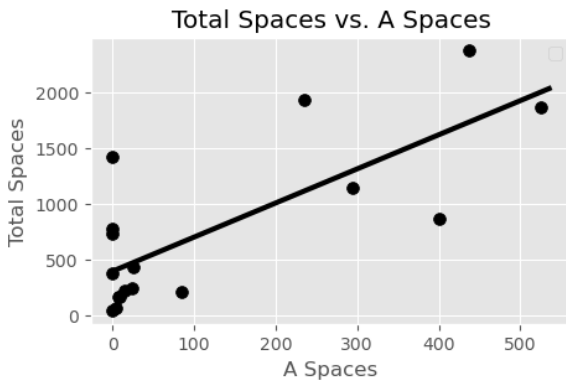
```
p_value = np.count_nonzero(differences _____ observed) / 5000
```

☐ < ☐ <= ☐ == ☐ != ☐ > ☐ >=

- d) Suppose the p-value of our test is 0.01 and we are testing the hypotheses at the 0.05 significance level. Which hypothesis is better supported by the data?
- ☐ Null ☐ Alternative

Question 2

UCSD's parking lots include A spaces for faculty, B spaces for staff, and S spaces for students. The scatter plot on the left shows the relationship between the number of A spaces and the total number of parking spaces in each region of campus. Similarly, the scatter plot on the right shows the relationship between the number of B spaces and the total number of parking spaces in each region of campus. Note that these numbers represent a **count of parking spaces existing** on campus, and have nothing to do with occupancy.



- a) Based on these scatter plots, which pair of variables has a larger correlation coefficient?
- ☐ A Spaces and Total Spaces ☐ B Spaces and Total Spaces
- b) Which of the following variables would most likely be negatively associated with the total number of spaces in a campus region?
- ☐ Total area of the campus region
- ☐ Number of athletic fields in the campus region
- ☐ Number of residents in the campus region
- ☐ Number of electric vehicle charging stations in the campus region
- c) The number of S spaces and the total number of parking spaces in each campus region are linearly related with a correlation coefficient of 0.6. The number of S spaces in the Warren College region of campus is 5 standard deviations below average. What does the regression line predict for the number of total spaces in Warren College, measured in standard units?

- d) True or False: Since number of parking spaces is always a positive integer, the regression line that predicts the total number of spaces from the number of S spaces must have a positive y -intercept.
- ☐ True ☐ False

Before submitting your quiz, make sure your PID is on the front page and on the top right corner of page 3.