
Quiz 5 - DSC 10, Winter 2024

Full Name:

PID:

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

Instructions:

- This quiz consists of 3 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.

Question 1

- a) Suppose you want to estimate the proportion of UCSD students that prefer sunny days over rainy days. You plan to survey 900 students, then construct a 95% confidence interval for this proportion. What is the **widest possible width** for the resulting confidence interval? Give your answer as a **fully simplified fraction**.

- b) If you decide to survey 450 students instead of 900 students for your sample, the maximum possible width of your 95% confidence interval would:
- ☐ double ☐ increase by more than double ☐ increase by less than double

Question 2

Bill claims that San Diego is sunny 60% of the year, but you think the percentage is higher. You decide to test the validity of Bill's claim by running a hypothesis test.

- a) Both your null and alternative hypotheses will take this form:

The percentage of sunny days in San Diego is _____ 60%.

- (i) For the null hypothesis, how should we fill in the blank?

☐ equal to ☐ not equal to ☐ greater than ☐ less than

- (ii) For the alternative hypothesis, how should we fill in the blank?

☐ equal to ☐ not equal to ☐ greater than ☐ less than

- b) Say that in 2023, San Diego had 235 sunny days. Using the number of sunny days per year as the test statistic, fill in the following code to run the hypothesis test and store the p-value of your test in `p_val`. We'll assume that all years have 365 days.

```
observed_stat = ____(w)___
```

```
results = np.array([])
```

```
for i in np.arange(10000):
```

```
    result = np.____(x)____(365, ____ (y)____)[0]
```

```
    results = np.append(results, result)
```

```
p_val = np.count_nonzero(results ____(z)____ observed_stat) / 10000
```

w:

, x:

y:

, z:

- c) Suppose the p-value of this hypothesis test is 0.1. Which of the following statements is a correct interpretation of this p-value?
- ☐ If the null hypothesis is true, there is a 10% chance of obtaining a test statistic equal to the observed statistic.
 - ☐ If the null hypothesis is true, there is a 90% chance of obtaining a test statistic equal to the observed statistic.
 - ☐ If the null hypothesis is true, there is a 10% chance of obtaining a test statistic equal to, or at least as extreme as, the observed statistic.
 - ☐ If the alternative hypothesis is true, there is a 90% chance of obtaining a test statistic equal to, or at least as extreme as, the observed statistic.

Question 3

Charlie spends 100 days in each of three California cities (San Diego, San Francisco, and Los Angeles). He records the primary weather for each day in each location. The table below shows these values as proportions. For example, 43 of the 100 days that Charlie spent in San Diego had primarily sunny weather.

Weather	San Diego	San Francisco	Los Angeles
sunny	0.43	0.51	0.34
partly cloudy	0.24	0.18	0.22
cloudy	0.22	0.15	0.26
rainy	0.11	0.16	0.18

- a) What is the total variation distance between the San Diego and San Francisco weather distributions? Give your answer as an **exact decimal**.

- b) Charlie reconsiders his weather categories and decides to count any “partly cloudy” days as “cloudy” days, thereby combining these two categories into one. If he recalculates the TVD between any pair of cities, what will he find?
- ☐ The new TVD is less than or equal to the old TVD.
 - ☐ The new TVD is greater than or equal to the old TVD.
 - ☐ The new TVD is equal to the old TVD.
 - ☐ There is not enough information to determine the relationship between the new TVD and the old TVD.

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