Quiz 4 - DSC 10, Winter 2025

Full Name:

PID:

Quiz Time: ◯ 3PM ◯ 3:30PM ◯ 4PM ◯ 4:30PM

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.
- You may use one handwritten sheet of 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

You want to estimate the proportion of all flower fields in the world that use fertilizer. You will create a **68% confidence interval** for this proportion, based on a sample of flower fields. You want your confidence interval to have a width of **at most 0.01**. Using the fact that the standard deviation of any dataset of 0s and 1s is no more than 0.5, calculate the **minimum sample size required**. Give your answer as an **integer**.

sample size $\geq$ ☐

Question 2

A researcher collects data on many flowers, some of which have been treated with fertilizer and some of which have not (untreated). The researcher wants to determine whether there is a relationship between fertilizer use and the height of the flowers.

a) Choose the correct formulation of the null and alternative hypotheses.

◯ **Null:** There is a difference in the average height of treated and untreated flowers.
Alt: There is no difference in the average height of treated and untreated flowers.

◯ **Null:** There is no difference in the average height of treated and untreated flowers.
Alt: There is a difference in the average height of treated and untreated flowers.

◯ **Null:** There is no difference in the average height of treated and untreated flowers.
Alt: Treated flowers grow taller, on average, than untreated flowers.

◯ **Null:** Fertilizer use has no impact on flower height.
Alt: Fertilizer use causes taller flowers.

b) True or False: We could use a permutation test to test these hypotheses.

◯ True ◯ False

Question 3

The DataFrame `flowers` contains information on a sample of flower fields. Each row corresponds to a different flower field. The `"Fertilizer"` column contains Boolean values corresponding to whether each flower field uses fertilizer.

You wonder if it could be the case that flower fields are fertilized at random, with each flower field having an 80% chance of being fertilized, independently of all others. You decide to do a hypothesis test to determine whether this could or could not be the case, testing at the $p = 0.05$ significance level. You will use as your test statistic **the absolute difference between 0.8 and the proportion of fertilized flower fields**.

a) Complete the implementation of the function `one_stat`, which calculates one simulated value of this test statistic, under the assumptions of the null hypothesis. Note that the optional argument `p` in `np.random.choice` specifies the probabilities with which each element is chosen (here, there is a 0.8 probability of selecting `True`).

PID: _____

```
def one_stat():
    sample_size = __(a)__
    random_choice = (np.random.choice([True, False],
                     sample_size, p=[0.8, 0.2]))
    return __(b)__
```

(a):

(b):

b) Complete the implementation of the function `one_stat_differently`, which also calculates one simulated value of this test statistic under the null.

```
def one_stat_differently():
    multi = np.random.multinomial(__(c)__)
    return __(d)__
```

(c):

(d):

c) Fill in the blanks to calculate 10,000 simulated values of the test statistic and collect them in an array called `many_stats`. You can use the functions you've already written to help you.

```
many_stats = __(e)__
for i in np.arange(10000):
    many_stats = __(f)__
```

(a):

(b):

d) Suppose that the observed value of the test statistic is 0.04. What do we conclude?

◯ Reject the null hypothesis. ◯ Accept the null hypothesis.

◯ Fail to reject the null hypothesis. ◯ Not enough information.

e) True or False: We could construct a confidence interval to test these hypotheses.

◯ True ◯ False

Reminders:

- **Write your PID** on the front page and on the top right corner of page 3.
- 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**.
- Stay in your seat until the quiz is over. No leaving early.