Quiz 4 - DSC 10, Fall 2025

Full Name:

PID:

Quiz Time: ◯ 2PM ◯ 3PM ◯ 4PM ◯ 5PM

Instructions:

- This quiz consists of 5 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.
- Assume we have already run `import babypandas as bpd`, `import numpy as np`, and `import scipy`.

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.

Ryan works at Geisel Library. He counts the number of students in Geisel's second floor at 10 AM for some days in summer and fall, storing them in DataFrames `summer` (on the left) and `fall` (on the right). The first few rows are shown, along with the mean and standard deviation of the `"students"` column.

`summer`

	date	students
0	2025-06-01	28
1	2025-06-07	46
2	2025-06-11	34

mean = 40
SD = 10

`fall`

	date	students
0	2025-10-02	371
1	2025-10-05	473
2	2025-10-09	419

mean = 400
SD = 10

Question 1

At least what proportion of `summer.get("students")` is in the range $[30, 40]$?

◯ ≥ 0.75 ◯ ≥ 0.68 ◯ ≥ 0.25 ◯ ≥ 0.00

Question 2

Ryan runs the code below to make a single array with the contents of the `"students"` columns from both `summer` and `fall`.

```
combined = np.append(summer.get("students"), fall.get("students"))
```

a) Select all of the true statements below:

☐ The mean of `combined` is 220.
☐ The mean of `combined` is in the range $(40, 400)$.
☐ The standard deviation of `combined` is less than 10.
☐ The standard deviation of `combined` is greater than 10.
☐ None of the above.

b) Since he was in a rush, Ryan asked ChatGPT to write code that standardizes `combined`. To his surprise, 90% of the *standardized* data is less than 0. Which of the following best explains this situation?

◯ Ryan should be surprised; this must have incorrectly standardized the data.
◯ Ryan should be surprised; a number of students cannot be negative.
◯ Ryan should not be surprised; it's possible for 100% of the data to be below 0.
◯ Ryan should not be surprised; his data is skewed.

Question 3

For this question only, suppose we know that `summer.get("students")` is normally distributed. You randomly select one number from `summer.get("students")`. Using the function `scipy.stats.norm.cdf`, write **one line of code** that evaluates to the probability that the number you selected is in the range $[30, 45]$.

PID: _____

Question 4

Ryan writes the following code to obtain a 95% bootstrapped confidence interval for the mean number of students in Geisel's second floor at 10AM in the **fall only**:

```
stats = np.array([])
for i in range(3333):
    stat = fall.sample(fall.shape[0],
                       replace=True).get("students").mean()
    stats = np.append(stats, stat)
lower = np.percentile(stats, 2.5)
upper = np.percentile(stats, 97.5)
```

a) Write **one line of code** that evaluates to the **upper** endpoint of the 95% CLT-based confidence interval for this same parameter. Recall that some potentially relevant values are provided with the DataFrames on the previous page.

`upper_clt` = _____

b) **True or False**: The 95% confidence interval generated by the bootstrapping code above **cannot be narrower** than a 95% confidence interval generated by the Central Limit Theorem.

◯ True
◯ False

c) **True or False**: The Central Limit Theorem tells us that the data in `fall.get("students")` is roughly normally distributed.

◯ True
◯ False

d) **True or False**: To obtain a 95% bootstrapped confidence interval for the median, we can simply change `.mean()` to `.median()` in the original code above.

◯ True
◯ False

e) **True or False**: We can use the CLT to create a valid 95% confidence interval for the median.

◯ True
◯ False

Question 5

Suppose Ryan is interested in collecting data and using it to form a 95% confidence interval for the **proportion of UCSD students that have ever been to Geisel library**.

Write an expression that evaluates to the minimum sample size required for the confidence interval to have a width of at most 0.2.

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.