Quiz 3 - DSC 10, Spring 2024

Full Name:

PID:

Quiz Time: ◯ 2PM ◯ 2:30PM ◯ 3PM ◯ 3:30PM ◯ 4PM

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

Which of the following statements are true in general? Select all that apply.

☐ Parameters are fixed, but statistics can change depending on the sample.

☐ Parameters and statistics can both fluctuate depending on the sample.

☐ For simple random samples, statistics give better estimates of parameters when the sample size is larger.

☐ The distribution of a statistic is the same regardless of the sample size.

☐ None of the above.

Question 2

The DataFrame `restaurants` contains information about a sample of restaurants in San Diego County. We have each restaurant's `"name" (str)`, `"rating" (int)`, average `"meal_price" (float)`, and type of `"cuisine" (str)`, such as `"Thai"` or `"Italian"`.

a) You are interested in estimating the average `"meal_price"` across all Italian restaurants in San Diego County using only the data in `restaurants`. Fill in the following code so that `italian_means` evaluates to an array of 1000 bootstrapped estimates for this parameter.

```
def bootstrap_means(data, n_samples):
    means = np.array([])
    for i in range(n_samples):
        resample = data.sample(__(a)__, replace = __(b)__)
        means = np.append(means, __(c)__)
    return means

italian_restaurants = __(d)__
italian_means = bootstrap_means(italian_restaurants, __(e)__)
```

(a):

(b):

(c):

(d):

(e):

b) Next, fill in the blanks below so that `italian_CI` evaluates to an 88% bootstrapped confidence interval for the average `"meal_price"` across all Italian restaurants in San Diego County.

```
lower_bound = np.percentile(italian_means, __(a)__)
upper_bound = np.percentile(italian_means, __(b)__)
italian_CI = [lower_bound, upper_bound]
```

(a): _____ (b): _____

c) Suppose `italian_CI` evaluates to [25, 35]. Which of the following statements are correct? Select all that apply.

- ☐ If we randomly selected 1000 Italian restaurants from the population of Italian restaurants in San Diego County, about 880 of them will have an average `"meal_price"` between \$25 and \$35.
- ☐ There is an 88% chance that the average `"meal_price"` of Italian restaurants in San Diego County falls between \$25 and \$35.
- ☐ 88% of all Italian restaurants have an average `"meal_price"` between \$25 and \$35.
- ☐ None of the above.

Question 3

Which of the following can be used to generate a **simple random sample** of `"rating"`s from 10 restaurants in `restaurants`? Select all that apply.

☐ Option 1:

```
sample = restaurants.take(np.arange(10)).get("rating")
```

☐ Option 2:

```
sample = restaurants.sample(10, replace = False).get("rating")
```

☐ Option 3:

```
sample = restaurants.sample(10, replace = True).get("rating")
```

☐ Option 4:

```
positions = np.random.choice(np.arange(0, restaurants.shape[0]),
                             10, replace = False)
sample = restaurants.take(positions).get("rating")
```

☐ Option 5:

```
positions = np.random.choice(np.arange(0, restaurants.shape[0]),
                             10, replace = True)
sample = restaurants.take(positions).get("rating")
```

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