Quiz 4 - 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

Suppose we've imported the `scipy` module. To the nearest 0.5, what does the following expression evaluate to?

```
scipy.stats.norm.cdf(-2) * 100
```

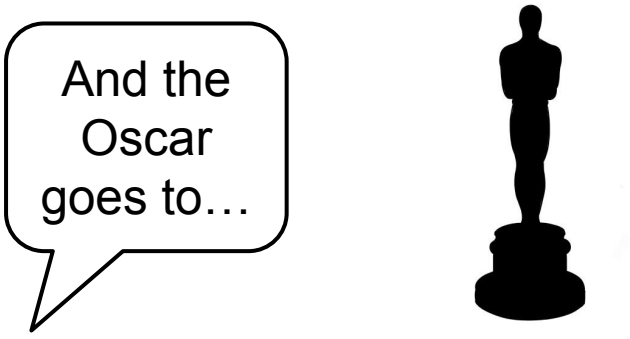
Question 2

Select all the true statements below.

- ☐ The average of the deviations from the mean is a meaningful measure of the spread of the data.
- ☐ It is possible for the standard deviation of a dataset to equal zero.
- ☐ It is possible for the standard deviation of a dataset to be negative.
- ☐ Given the standard deviation of a dataset, we can determine its mean.
- ☐ Given the standard deviation of a dataset, we can determine its variance.

Question 3

The Oscars, or Academy Awards, are the highest awards in the film industry, awarded each year to the best movies of that year. The `oscars` DataFrame contains a row for each movie that has ever been nominated for an Oscar. The `"name"` column contains the name of the movie and the `"rating"` column contains a rating of the movie on a 0 to 100 scale. This number incorporates many factors, but we won't worry about how it is computed.



PID: _____

a) Fill in the blanks below to collect a **simple random sample** of 400 movies from the `oscars` DataFrame, then calculate 10,000 bootstrapped sample mean ratings.

```
my_sample = __(x)__
n_resamples = 10000
boot_means = np.array([])
for i in np.arange(n_resamples):
    resample = __(y)__
    mean = __(z)__
    boot_means = np.append(boot_means, mean)
```

`(x)`: _____

`(y)`: _____

`(z)`: _____

b) In each blank, circle the word that correctly fills in the sentence.

A histogram of ***`boot_means`*** *shows a(n)* [*probability / empirical*] *distribution of a* [*statistic / parameter*].

c) Suppose we use the array `boot_means` to calculate a 90% confidence interval for the mean rating of Oscar-nominated movies. **Select all** correct conclusions we can draw about this interval.

- ☐ There is a 90% chance that the true mean rating of all Oscar-nominated movies falls within this interval.
- ☐ The sample mean rating is within 90% of the true mean rating of all Oscar-nominated movies.
- ☐ If we looked at the ratings of many Oscar-nominated movies, about 90% of them would fall within this range.
- ☐ None of the above.

d) Suppose both of the following expressions evaluate to `True`.

- `my_sample.get("rating").mean() == 61.25`
- `np.std(my_sample.get("rating")) == 15`

What are the endpoints of a 95% CLT-based confidence interval for the mean rating of Oscar-nominated movies?

left endpoint: _____, right endpoint: _____

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