CSE303 (Section 1)

[Spring 2022]

Home Work 2

Home Work Title: Bootstrap Method

Submitted by:

Rizvee Hassan Prito

2019-3-60-041

Screenshots:

Step 1:

```
df=pd.read_csv('student_age.csv')
```

Reading the ‘student_age’ csv file which is the population.

```
x=random.choice(df,size=(20))
```

Selecting 20 datapoints from the population. Set of this 20 data point is the sample set.

Step 2:

```
x2=random.choice(x,size=(100))
```

From the sample set of 20 data points, resampling the set for 100 data points. This resampling is done in ‘with replacement’ technique.

Step 3:

```
means_list.append(x2.mean())
```

Calculating the mean of the resampling dataset and saving it in a list.

Step 4:

```
for i in range(0,1000):
    x=random.choice(df,size=(20))
    x2=random.choice(x,size=(100))
    means_list.append(x2.mean())
```

Repeating step-1 to step-3 for 1000 times.

Step 5:

```
means=pd.DataFrame(means_list)
mean=means.mean()

std=list(means.std())
print('\nStandard Deviation:',std[0])
```

Converting the list of 1000 means into a DataFrame. Calculating the mean of 1000 means with the bult in mean function of Pandas' DataFrame. Then calculating the standard deviation from the dataframe of 1000 means with the Pandas' built-in function std() and converting it into a list for printing the value accurately.

Standard Error:

```
std_error=list(means.sem())
print('\nStandard Error:',std_error[0])
```

Calculating standard error with the built-in sem() function of Scipy and converting the value into a list for printing it accurately.

Confidence Interval from Mean values:

```
dof=len(means)-1
CL=0.95

t_val=scipy.stats.t.ppf(q=1-CL/2,df=dof)

CI=[float(mean-(std[0]*t_val/np.sqrt(len(means)))) , float(mean+(std[0]*t_val/np.sqrt(len(means))))]
```

First, calculating the degree of freedom from the values of means which is 999. Then finding the T-score by using the given confidence level and degree of freedom in the function of Scipy library of Python. Then using the calculated mean, standard deviation of 1000 means and T-score, calculating the lower value and upper value of Confidence Interval.

Output:

```
Standard Deviation: 0.6414561433577689

Standard Error: 0.02028462432118038

Confidence Interval: [20.582845026960587, 20.585389633039416]
```

Conclusion:

Bootstrap method infers about the nature of the population data points. It evaluates the population data points by the 'with replacement' resampling technique. 'With replacement' method creates a sample in a larger size from a smaller sized sample by randomly choosing the same value one or more than once. This technique makes the statistics evaluation more accurate and unbiased. Bootstrap method also helps us to estimate the confidence interval which is the percentage of certainty that a value would can be found in a certain interval. So, the bootstrapping method is very significant in the field of statistics because this method helps us to infer data about the population, estimate the true standard errors, and can ensures us that data is tested efficiently.