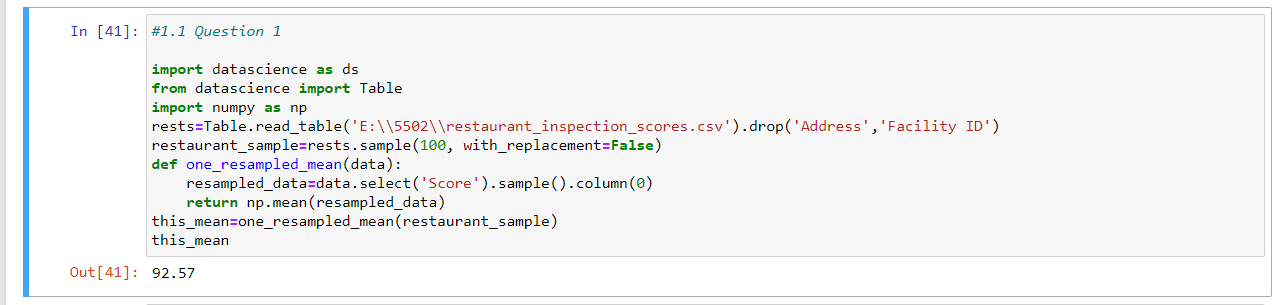
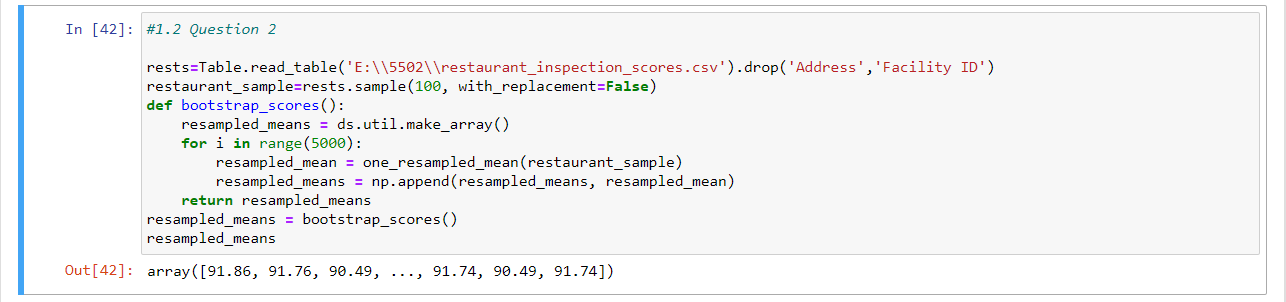
Assignment 6

1. The Bootstrap and the normal curve
   1. Question 1) Complete the function one resampled mean below. It should take in an original table data, with a column Score, and return the mean score of one resampling from data.

The mean score of one resampling from data is 92.57



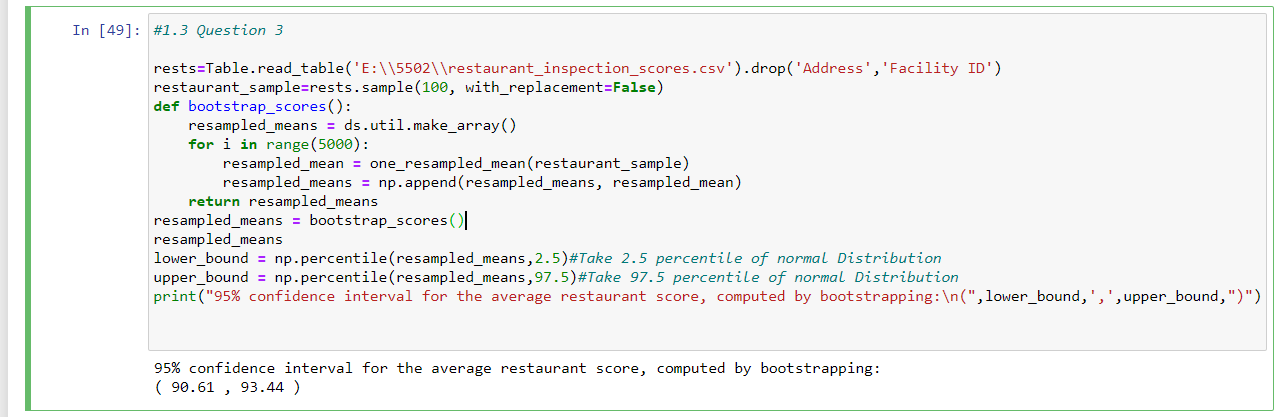
* 1. Question 2) Complete the function bootstrap scores below. It should take no arguments. It should simulate drawing 5000 resamples from restaurant sample and compute the mean restaurant score in each resample. It should return an array of those 5000 resample means.



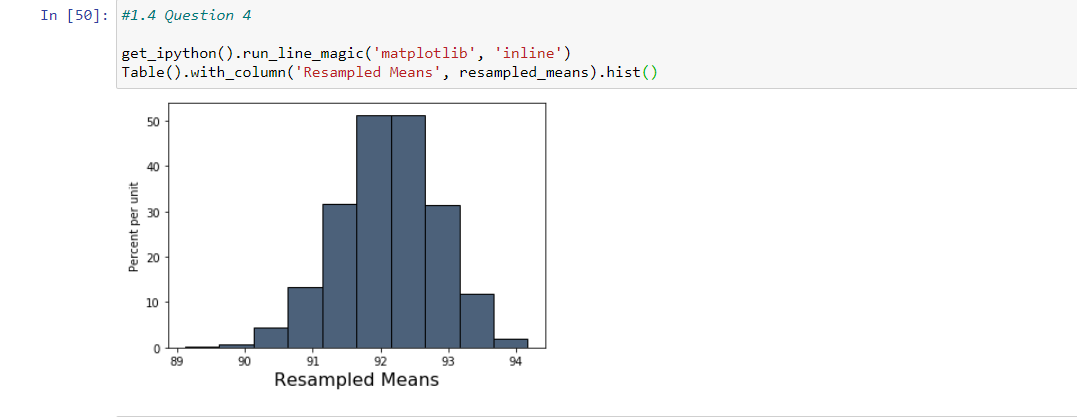
The array of 5000 resample means is ([91.86, 91.76, 90.49, ..., 91.74, 90.49, 91.74])

* 1. Question 3) Compute a 95 percent confidence interval for the average restaurant score using the array resampled means.

95 percent confidence interval for the average restaurant score is 90.61, 93.44



* 1. Question 4) What distribution is the histogram between question 2 and 3 displaying and why does it have that shape?

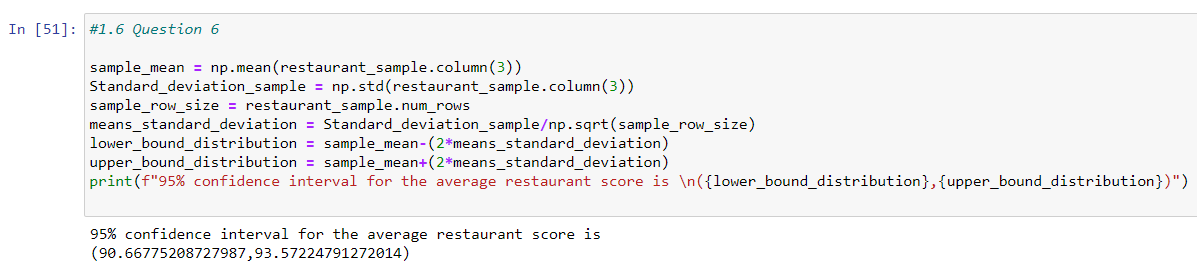


From the histogram shown in the figure, the sample means shape is normal distribution because the central limit theorem. This theorem states that sampling distribution of the mean of any independent random variable will be normal or nearly normal, if the sample size is large enough which is as shown in above figure i.e., normal distribution shape.

* 1. Question 5) Does the distribution of the sampled scores look normally distributed? State yes or no and describe in one sentence why you should expect this result.

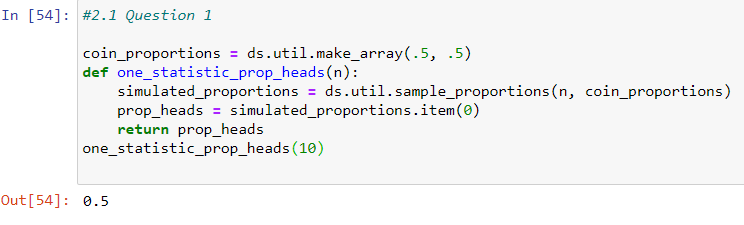
Ans: No, because the central limit theorem does not apply to the distribution of sampled scores. It applies to only sum or average of the sampled scores.

* 1. Question 6) Without referencing the array resampled means or performing any new simulations calculate an interval around the sample mean that covers approx. 95% of the numbers in the resampled means array. This confidence interval should look very similar to the one you computed in question 3.



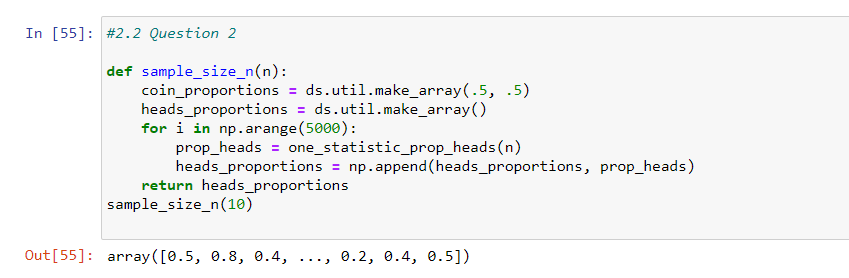
The confidence interval for the average restaurant score is 90.66, 93.57 which looks very similar to Question 3

1. Testing the Central limit theorem:
   1. Questions 1) Define the function one statistic prop heads which should return exactly one simulated statistic of the proportion of heads from n coin flips.



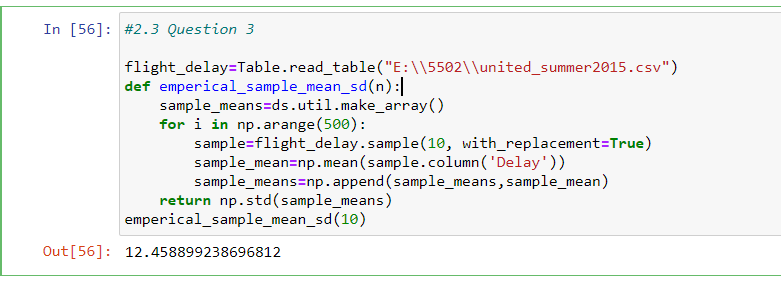
Exactly one simulated statistic of the proportion of heads from n coin flips is 0.5

* 1. Question 2) Write a function called sample\_size\_n that takes in a sample size n. It should return an array that contains 5000 sample proportions of heads, each from n coin flips.



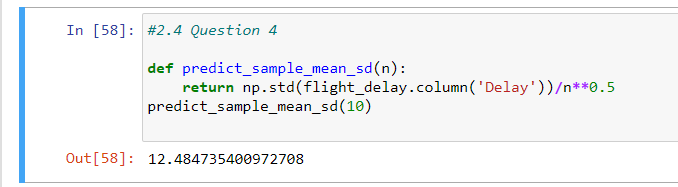
The array is ([0.5, 0.8, 0.4, ..., 0.2, 0.4, 0.5])

* 1. Question 3)Write a function called empirical sample mean SD that takes a sample size n as its argument. The function should simulate 500 samples with replacement of size n from the flight delays dataset, and it should return the standard deviation of the means of those 500 samples.



The standard deviation is 12.45

* 1. Question 4)Now, write a function called predict sample mean SD to find the predicted value of the standard deviation of means according to the relationship between the standard deviation of the sample mean and sample size that is discussed in the textbook. It takes a sample size n (a number) as its argument. It returns the predicted value of the standard deviation of the mean delay time for samples of size n from the flight delays (represented in the table united).



The predicted value of the standard deviation of means according to the relationship between the standard deviation of the sample mean and sample size is 12.48