

Day 21

DIY

Q1. Problem Statement: Sampling Distribution-I

A wine can be labeled as quality "7" and above only if the alcohol level is greater than 11.3%. If the level of alcohol is high, then the will taste sour enough; if the alcohol level is less than 11.3%, the wine might taste bitter. The quality control person is only allowed to open ten bottles from a batch of wine with quality "7" & "8" to ensure the levels are sufficient.

Load the “*winequality-red.csv*” data into a DataFrame and [erform following tasks:

1. Pick up a sample of ten bottles from "7" and "8" quality using `numpy.random()` function
2. Find out the mean and standard deviation of the sample
3. Plot the alcohol content available in the sample using a distplot
4. Calculate the standard error using formula - $\text{standard_error} = s / \text{np.sqrt}(n)$, Where, s = standard deviation of the sample, and n = sample size

Dataset:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

Sample Output:

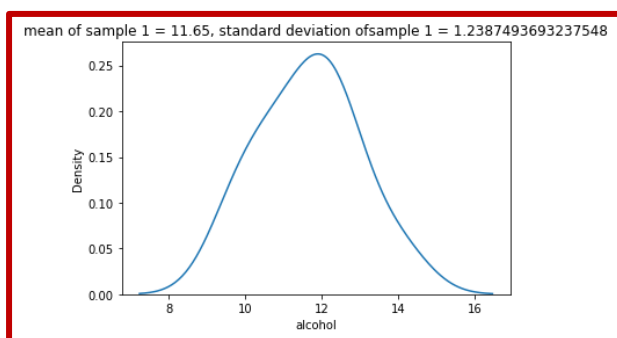
1. Pick up a sample of ten bottles from "7" and "8" quality using `numpy.random()` function

alcohol	
0	12.0
1	12.5
2	11.7
3	14.0
4	12.7
5	12.1
6	11.0
7	10.0
8	10.8
9	9.7

2. Find out the mean and standard deviation of the sample

The mean of the sample is: 11.65
The standard deviation of the sample is: 1.2387493693237548

3. Plot the alcohol content available in the sample using a distplot



4. Calculate the standard error using formula - `standard_error = s/np.sqrt(n)`, Where, `s` = standard deviation of the sample, and `n` = sample size

```
Standard Error is: 0.39172694571601785
```

Q2. Problem Statement: Sampling Distribution-II

Note: This problem is in continuation with the previous one.

Our goal was to get a mean of 11.3, but we got 11.65 for our samples. Since this is a sample and we do not expect it to be precisely 11.3, it is 11.65. Does this batch of the sample accurately reflect the alcohol content of the entire population? To answer these questions, perform the following tasks:

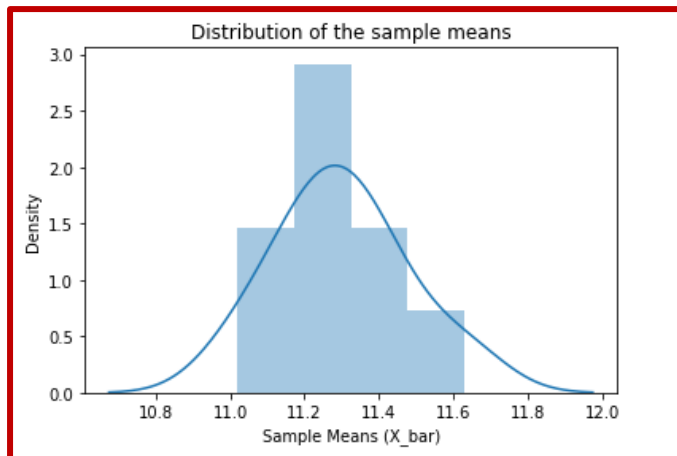
1. Take nine samples of size ten each from the population and find the means and standard deviations of all the nine samples
2. Plot the means of the sampling distribution
3. Find the standard error of the sampling distribution

Sample Output:

1. Take nine samples of size ten each from the population and find the means and standard deviations of all the nine samples

	Sample Means (\bar{X})	Sample Standard Deviation (s)
0	11.02	0.669029
1	11.14	0.971802
2	11.21	1.013361
3	11.63	0.888876
4	11.45	0.908020
5	11.24	0.935094
6	11.32	1.231909
7	11.39	0.773886
8	11.30	0.507937

2. Plot the means of the sampling distribution



3. Find the standard error of the sampling distribution

The Standard Error is: 0.168522995463527

edureka!
a Veranda Enterprise