CS-23334 FUNDAMENTALS OF DATA SCIENCE

ABENANTHAN P 240701005

Experiment 11 **Date: 09.10.2025**

11. Random Sampling and Sampling Distribution

Aim:

To explore random sampling from a population and understand the concept of sampling distribution using Python.

Description:

Random sampling and sampling distributions, showing how sample statistics vary across different random samples from a population.

Algorithm:

1. **Generate a Population:** Create a population of data with a specified distribution

2. **Random Sampling:** Perform random sampling from the population to create multiple samples of different size and Compute sample statistics (mean, standard deviation, etc.) for each sample.

3. Sampling Distribution: Plot histograms or density plots of sample statistics and compare the sampling distribution of the sample statistic (mean) with the population distribution.

4. Central Limit Theorem: Demonstrate the Central Limit Theorem by showing that as sample size increases, the sampling distribution of the sample mean approaches a normal distribution regardless of the population distribution.

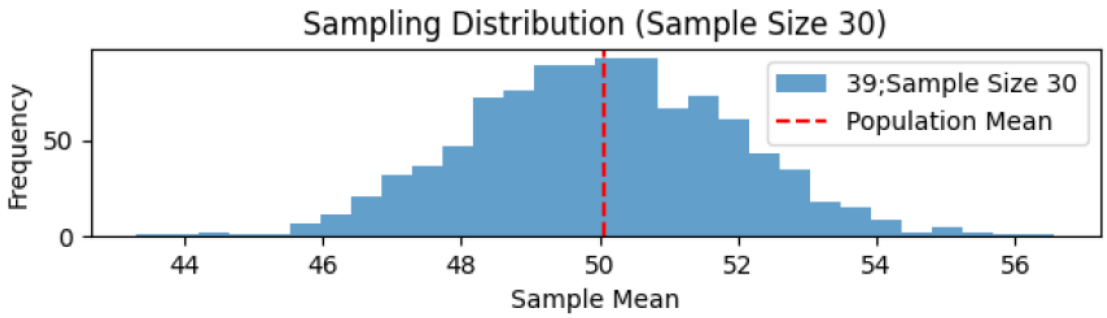
Code With Output:

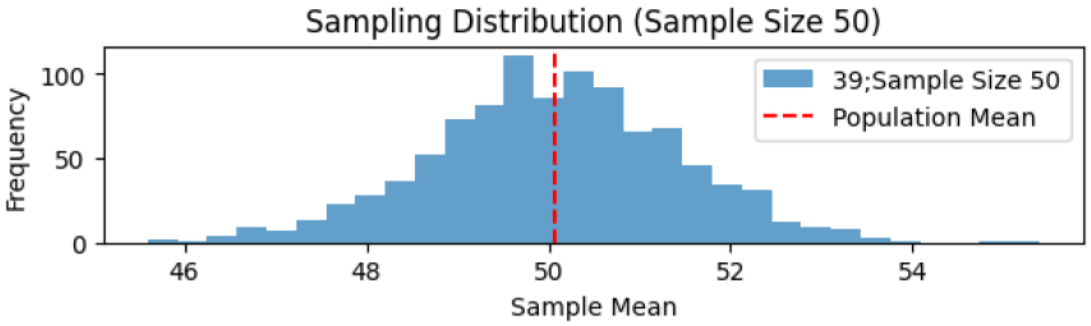
```
import numpy as np
import matplotlib.pyplot as plt

pop_mean = 50
pop_std = 10
pop_size = 100000
popu = np.random.normal(pop_mean, pop_std, pop_size)

sample_sizes = [30, 50, 100]
num_samples = 1000

sample_means = {} #empty

for size in sample_sizes:
    sample_means[size] = []
    for _ in range(num_samples):
        sample = np.random.choice(popu, size=size, replace=False)
        sample_means[size].append(np.mean(sample))

plt.figure(figsize=(12, 8))
```

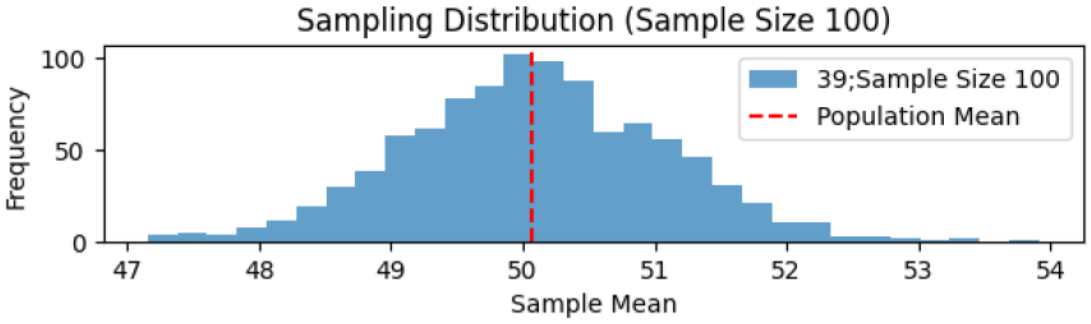
```
<Figure size 1200x800 with 0 Axes>
```

```
<Figure size 1200x800 with 0 Axes>
```

```
for i, size in enumerate(sample_sizes):
    plt.subplot(len(sample_sizes), 1, i+1)
    plt.hist(sample_means[size], bins=30, alpha=0.7, label=f'39;Sample
Size {size}')
    plt.axvline(np.mean(popu), color='red', linestyle='dashed',
linewidth=1.5,label='Population Mean')
    plt.title(f'Sampling Distribution (Sample Size {size})')
    plt.xlabel('Sample Mean')
    plt.ylabel('Frequency')
    plt.legend()
    plt.tight_layout()
    plt.show()
```







Result:

Thus the python program to explore random sampling from a population was completed.