CS-23334 FUNDAMENTALS OF DATA SCIENCE

ASHWIN C 240701058

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.

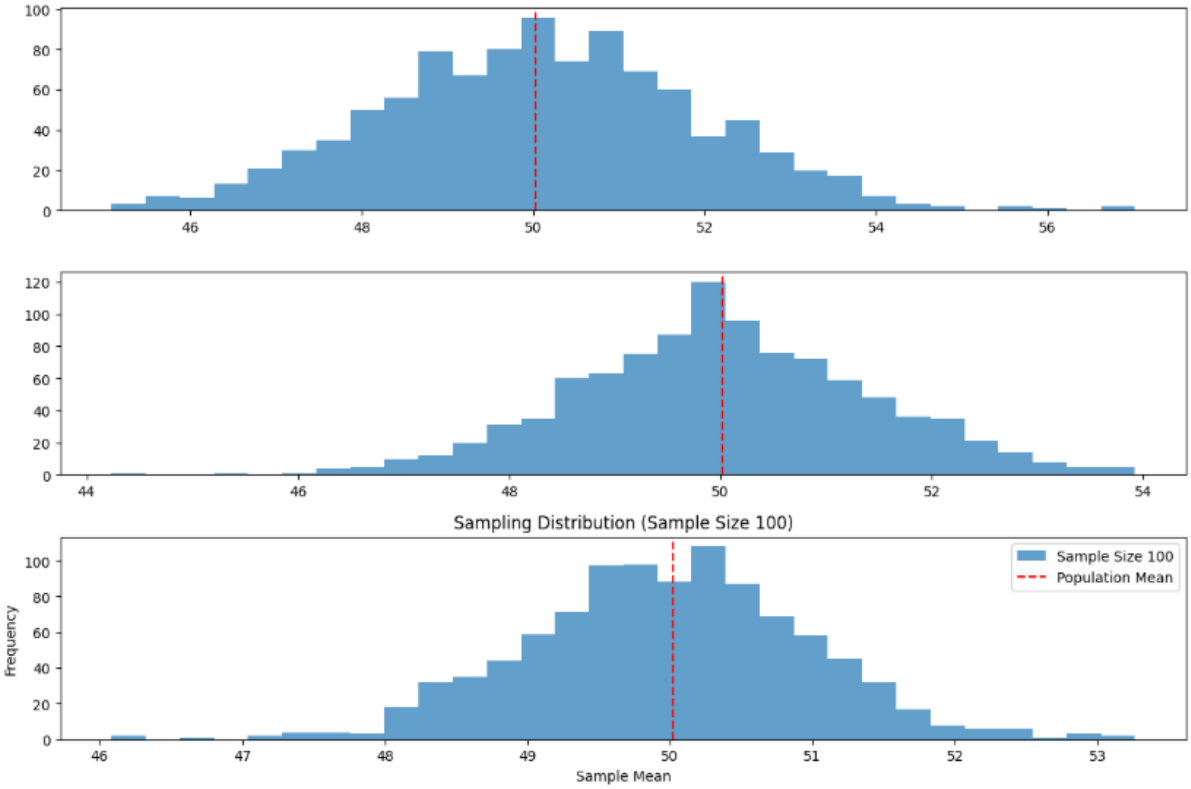
3. Sampling Distribution: Plot histograms or density plots of sample statistics and compare the sampling distribution of the sample statistic (mean) with the population distribution.

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 Example:

```
import numpy as np
import matplotlib.pyplot as plt
population_mean = 50
population_std = 10
population_size = 100000
population = np.random.normal(population_mean, population_std, population_size)
sample_sizes = [30, 50, 100] # different sample sizes to consider
num_samples = 1000 # number of samples for each sample size
sample_means = {}
for size in sample_sizes:
 sample_means[size] = []
 for _ in range(num_samples):
  sample = np.random.choice(population, size=size, replace=False)
  sample_means[size].append(np.mean(sample))
plt.figure(figsize=(12, 8))
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'Sample Size {size}')
 plt.axvline(np.mean(population), 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()
```

Output:


Sampling Distribution (Sample Size 100)
Sample Size 100
Population Mean
Frequency
Sample Mean

Result:

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