CS23334 -Fundamental Of Data Science

240701463

Random Sampling and Sampling Distribution

Exp : 11

Date: 12-10-2025

Aim:

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

Algorithm:

1. **Generate Population:** Create a large data population with specified mean and standard deviation (e.g., normal distribution).

2. **Random Sampling:** Draw multiple samples for different sample sizes (e.g., 30, 50, 100) from the population.

3. **Compute Sample Statistics:** Calculate the mean of each sample.

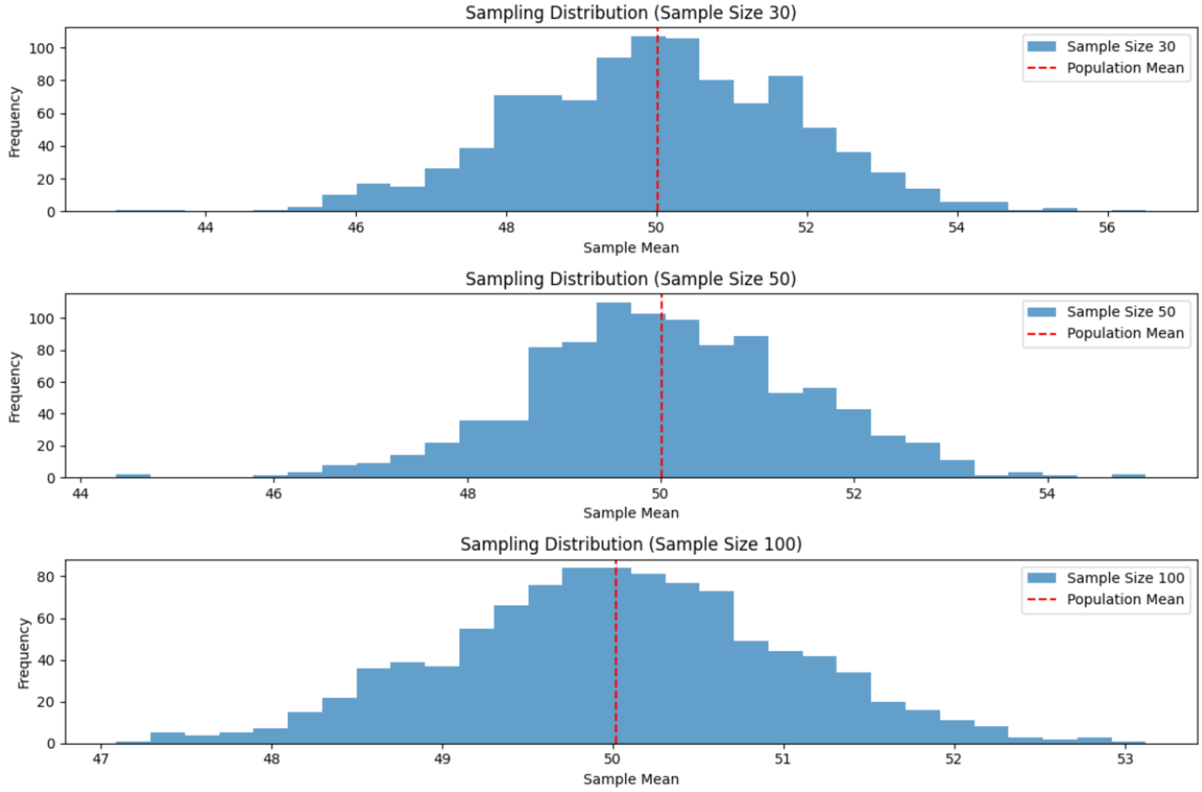
4. **Plot Distributions:** Plot histograms of the sample means (the sampling distribution) for each sample size.

5. **Compare:** Overlay the population mean on the histograms to compare the sampling distribution's center and spread.

Code:

```
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]

num_samples = 1000

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:



Result:

The experiment successfully generated a population and performed random sampling to create the sampling distribution of the sample mean. The output histograms show that for all sample sizes (30, 50, 100), the distribution of sample means is **approximately normal** and is **centered around the true population mean (50)**. Furthermore, as the **sample size increases (from 30 to 100)**, the distribution of the sample means becomes **tighter and less spread out** (lower standard error), visually demonstrating the fundamental concept of the Central Limit Theorem and the relationship between sample size and the precision of sample estimates.