CS23334-FUNDAMENTALS OF DATA SCIENCE

DEVA DHARSHINI P (240701107)

11.) RANDAOM SAMPLING AND SAMPLING DISTRIBUTION

Aim:

To understand and implement Random Sampling and analyze the Sampling Distribution to study how sample statistics vary from the population parameters.

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] = []
```

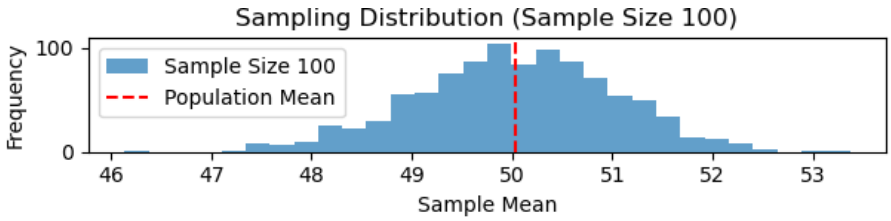
```
for _ in range(num_samples):
 sample = np.random.choice(population, size=size, replace=False)
 sample_means[size].append(np.mean(sample))
```

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

```
<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'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()
```



Result:
Random samples were taken, and the sampling distribution was successfully created.