

# DPP- Sampling\_&\_Sampling\_Distributions

---

## Basics of Sampling

### Question 1

**Suppose you want to study customer satisfaction in a mall. Explain how you would use systematic sampling to select participants.**

Solution:

Systematic sampling selects units at regular intervals from an ordered list.

Let:

$N$  = total number of customers

$n$  = desired sample size

$k$  = sampling interval

Formula used:

$$k = N / n$$

A random starting point between 1 and  $k$  is selected. Then every  $k$ -th customer is chosen. This ensures uniform coverage and avoids clustering bias.

### Question 2

**A school has 5 grades, each with different numbers of students. Describe how you would use stratified sampling to estimate the average exam score of all students.**

Solution:

The population is divided into strata based on grades. Random samples are drawn from each grade proportionally.

Formula for overall mean:

$$\text{Overall Mean} = \sum (\text{Stratum Mean} \times \text{Stratum Proportion})$$

This ensures fair representation from all grades.

### Question 3

**A school has 800 students: 400 in junior grades, 250 in senior grades, and 150 in advanced grades. You want to survey 100 students ensuring proportional representation.**

Solution:

Formula used:

$$\text{Sample from group} = (\text{Group size} / \text{Total population}) \times \text{Total sample size}$$

Junior:  $(400/800) \times 100 = 50$

Senior:  $(250/800) \times 100 \approx 31$

Advanced:  $(150/800) \times 100 \approx 19$

#### Question 4

**A store wants to survey its customers. It has a list of 1,000 customer transactions and wants to sample every 20th transaction. If the starting point is transaction #7, list the first 10 transactions.**

Solution:

Let:

a = starting transaction

k = sampling interval

i = index (0,1,2,...)

Formula:

Transaction =  $a + i \times k$

Selected transactions:

7, 27, 47, 67, 87, 107, 127, 147, 167, 187

#### Question 5

**A city has 200 neighborhoods. You select 10 neighborhoods at random and survey all households within them. What is the main advantage of cluster sampling?**

Solution:

Main advantage:

Cluster sampling reduces cost and time by surveying entire groups instead of scattered individuals. It is efficient for geographically dispersed populations.

#### Question 6

**In a population of students, the average study time is 10 hours per week with a standard deviation of 3 hours. A sample of size 36 is taken.**

Solution:

Formula:

Standard Error (SE) =  $\sigma / \sqrt{n}$

SE =  $3 / \sqrt{36} = 0.5$  hours

## Standard Error Calculation

#### Question 7

**If factory light bulb lifespans have a mean of 1,200 hours and a standard deviation of 200 hours, calculate the standard error for samples of size 50.**

Solution:

$$SE = \sigma / \sqrt{n} = 200 / \sqrt{50} \approx 28.28 \text{ hours}$$

## Effect of Sample Size on Variability

### Question 8

A population has a mean of 50 and a standard deviation of 10. Compare the standard error for sample sizes of 16, 64, and 256.

Solution:

SE formula:  $SE = \sigma / \sqrt{n}$

$$n = 16 \rightarrow SE = 2.5$$

$$n = 64 \rightarrow SE = 1.25$$

$$n = 256 \rightarrow SE = 0.625$$

As sample size increases, variability decreases.

## Real-World Applications

### Question 9

A company sends out a survey to 1,000 customers and receives responses from 150.

Solution:

**Sampling method used:** Voluntary response sampling

**Potential bias:**

Non-response bias, as respondents may not represent the entire population.

### Question 10

A researcher collects 10 random samples of size 30. The sample means are: 75, 78, 74, 76, 77, 75, 79, 76, 74, and 77.

Solution:

Mean of sampling distribution:

$$\bar{x} = \Sigma x / n = 76.1$$

Standard deviation of sampling distribution  $\approx 1.7$

## Probability Using CLT

### Question 11

The average weight of apples is 200 g with a standard deviation of 50 g. A random sample of 36 apples is taken. What is the probability that the sample mean is greater than 210 g?

Solution:

$$SE = 50 / \sqrt{36} = 8.33$$

$$Z = (210 - 200) / 8.33 \approx 1.20$$

From Z-table:

$$P(Z < 1.20) = 0.8849$$

$$P(Z > 1.20) = 1 - 0.8849 = 0.115$$

Thus, there is an 11.5% probability that the sample mean exceeds 210 g.

SHIVANSH YADAV