Sampling Distributions: Guided Lab Conversation and Review

# Overview

This activity is designed to simulate a guided lab conversation between you (the learner) and your professor. The goal is to solidify your understanding of sampling distributions through questions, matching games, recall prompts, and formula reconstruction.  
  
The activity is broken into the following stages:  
1. Quick Review Questions (MCQs)  
2. Matching Terms to Definitions (Grouped by Concept)  
3. Key Formulas with Plain-English Descriptions  
4. Visual Exploration of Sampling Distributions  
5. Formula Recall (Fill-in-the-Blank with Feedback)

# 1. Quick Review Questions

1. Q1: What is a sampling distribution?

* A. A histogram of values in the population
* B. A distribution of repeated sample means from the same population (✔)
* C. A single sample taken from the population
* D. The distribution of standard deviations across different datasets

Explanation: B is correct. A sampling distribution is the distribution of a statistic, like the mean, across many repeated samples.

1. Q2: According to the Central Limit Theorem (CLT), the sampling distribution of the sample mean will be approximately normal if:

* A. The population is normal
* B. The sample size is small
* C. The population is skewed but the sample size is large (✔)
* D. The samples are taken with replacement

Explanation: C is correct. The CLT works even for skewed populations, as long as the sample size is large.

1. Q3: As the sample size n increases, what happens to the standard error (SE) of the sample mean?

* A. It stays the same
* B. It increases
* C. It decreases (✔)
* D. It becomes equal to the population standard deviation

Explanation: C is correct. The standard error shrinks as sample size grows.

1. Q4: If the original population is heavily skewed, what happens to the distribution of the sample mean as you increase the sample size?

* A. It stays skewed
* B. It becomes more normal (✔)
* C. It becomes bimodal
* D. It becomes uniform

Explanation: B is correct. The CLT implies that sample means will be normally distributed as n increases.

1. Q5: Why do we simulate multiple samples when studying sampling distributions?

* A. To generate outliers
* B. To compare population and sample distributions
* C. To see how much sample means vary from sample to sample (✔)
* D. To increase the sample size artificially

Explanation: C is correct. Simulation shows the variability across sample means.

# 2. Matching Terms to Definitions

Match the terms in each category to the correct definition. Answers follow each set for self-checking.

## Population & Sample Concepts

* Population: The entire group you're studying or generalizing to
* Sample: A subset of individuals selected for analysis
* Parameter: A fixed numerical value that describes the full population
* Statistic: A numerical summary calculated from a subset of data

## Distributions

* Sampling Distribution: The distribution of a statistic (like the mean) from many repeated samples
* Empirical Distribution: A distribution based on observed data or simulation
* Theoretical Distribution: A mathematical model describing how a variable behaves ideally
* Distribution: A general term for how values are spread in a dataset

## Center & Spread

* Mean: The average of a set of values
* Standard Deviation: How spread out values are around the mean
* Variance: The square of the standard deviation
* Standard Error: Standard deviation of the sampling distribution of a statistic

## CLT Concepts

* Central Limit Theorem: Describes how the distribution of sample means becomes normal as sample size increases
* Skewed Distribution: A distribution that is not symmetric
* Normal Distribution: A symmetrical bell-shaped distribution

## Simulation

* Sample Size: Number of observations in one sample
* Number of Samples: How many samples are drawn
* Repetition / Replication: Repeating the sampling process
* With Replacement: Sampling where repeats are allowed

## Estimation

* Sampling Variability: Variation in sample statistics across samples
* Law of Large Numbers: Larger samples give more accurate estimates
* Bias: Systematic error in estimation
* Precision: Closeness of repeated estimates

# 3. Key Formulas with Plain-English Descriptions

## Sample Mean

Formula: \bar{x} = \frac{1}{n} \sum\_{i=1}^{n} x\_i

Explanation: Add all the values in a sample and divide by how many values there are.

## Population SD

Formula: \sigma = \sqrt{\frac{1}{N} \sum\_{i=1}^{N} (x\_i - \mu)^2}

Explanation: Take each value, subtract the mean, square it, average the squares, and take the square root.

## Sample SD

Formula: s = \sqrt{\frac{1}{n - 1} \sum\_{i=1}^{n} (x\_i - \bar{x})^2}

Explanation: Same as population SD but divide by n−1 instead of n.

## Standard Error

Formula: SE = \frac{\sigma}{\sqrt{n}} or SE ≈ \frac{s}{\sqrt{n}}

Explanation: Divide the standard deviation by the square root of sample size.

## Central Limit Theorem

Formula: \bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})

Explanation: Sample means follow a normal distribution as n increases.

## Variance of Sample Mean

Formula: Var(\bar{x}) = \frac{\sigma^2}{n}

Explanation: Spread of sample means = population variance divided by sample size.