Lab 3 - The Central Limit Theorem Gauntlet

Overview

In this assignment, you and an **AI agent** (ChatGPT, Gemini, Claude, etc.) will work together to investigate the Central Limit Theorem (CLT) through creativity, curiosity, and a bit of statistical mischief. You remain the chief statistician; the AI is simply your eccentric assistant whose ideas may be helpful, dubious, or unexpectedly brilliant.

You will complete three challenges. Each one asks you to design a scenario, generate or interpret data, and draw conclusions about the behaviour of the sample mean. The challenges are intentionally open-ended. You decide how to collaborate with the AI, what sample sizes to explore, and which visualizations or analyses will support your conclusions. You are encouraged to invent scenarios that are whimsical, bizarre, or wildly imaginative—*provided they remain appropriate and non-offensive.*

Your final work should remain academically clear and concise, even if the contexts themselves are playful or imaginative.

AI Use Policy

You may use your AI partner as a brainstorming companion, a generator of unusual ideas, or a source of datasets. However, you are responsible for all reasoning, interpretation, and conclusions. Your work must reflect your own thinking, not a transcript of a conversation.

Deadline: (...) (Dropbox in Lea)

Guidelines

Part of learning statistics is learning how to ask good analytical questions. This lab is designed to give you freedom to explore, but if you find yourself unsure how to begin, try focusing on how the sampling distribution of the mean changes as your sample size grows, how unusual or extreme populations influence this behaviour, and how the Central Limit Theorem explains (or fails to explain) what you observe.

Here are a few optional questions that may help guide your thinking:

- How do the quirks of your population (skewness, spikes, gaps, heavy tails) appear in the sampling distributions?
- As your sample size increases, which features of the sampling distribution begin to stabilize?
- Does the Central Limit Theorem seem to apply quickly, slowly, or unexpectedly in your scenario?

Challenge #1: Beat the AI

This challenge focuses on comparing AI intuition with empirical evidence.

Task: Can you out-predict a model trained on half the internet? Invent a scenario involving a clearly non-normal variable and ask your AI partner what it expects the sampling distribution of the sample mean to look like for several different sample sizes. Then create your own dataset, run your own simulations, and see how well reality matches the AI's intuition. Focus on where the AI is accurate, where it fails, and why.

What to Submit

- A short description of your scenario.
- Visual evidence from your simulations (e.g., histograms of sample means).
- A concise comparison between the AI's predictions and what you observed.
- What aspect of your sampling distributions was most surprising or counterintuitive, and why do you think it occurred in your particular scenario?

Challenge #2: The Chaotic Distribution Showdown

This challenge explores how the CLT behaves under extreme or unconventional conditions.

Task: Design a population so unruly that even the Central Limit Theorem hesitates. Create a wildly extreme or unconventional distribution—something with sharp spikes, heavy tails, gaps, clusters, strange behaviour, or any other features you can imagine. Use your AI assistant to help bring this population to life and generate a dataset that reflects its quirks. Then explore how the sampling distribution of the mean transforms as your sample size grows.

Some chaotic populations quickly begin to resemble a bell curve. Others cling to their chaos. See where yours lands.

What to Submit

- A brief description of your chaotic scenario.
- A visualization of your population.
- Several sampling-distribution visualizations for different sample sizes.
- A short reflection on how (and how quickly) the distribution of the sample mean changed.
- What feature of your chaotic population seemed to persist the longest as sample size increased, and what does that suggest about how the CLT behaves for extreme distributions?

Challenge #3: CLT Detective Mystery

This challenge asks you to infer a hidden population from limited information.

Task: Your final challenge turns you into a **statistical detective**. Your mission is to reconstruct a hidden population using only the sampling distributions ($n \geq 2$) of its sample means. Treat the sampling distributions as clues. Examine how their shape changes with increasing sample size, and infer the characteristics of the unknown population. Think about skewness, modality, discreteness or continuity, and what real-world process could produce such a distribution.

Approach it like solving a mystery: follow the patterns, connect the hints, and propose a plausible story for the population behind the evidence.

What to Submit

- A short description of your mystery context.
- A brief analysis explaining the likely characteristics of the hidden population.
- At least two plausible real-world explanations consistent with your reconstruction.
- A sentence or two reflecting on whether the AI's suggestions clarified or complicated your investigation.
- Which clue in the sampling distributions most strongly influenced your reconstruction of the hidden population, and why?