Monte Carlo Simulation Project

Course: Advanced Statistics with Applications in R
Delivery Format: Team project (1–3 students)

🔍 Overview

In this project, your team will design and conduct a Monte Carlo simulation study based on a real-world context. Rather than abstract synthetic scenarios, each team will choose an application domain such as public health, supply chains, sports analytics, finance, or environmental monitoring.

While you will *simulate* data, your design must be rooted in a *realistic* dataset, scenario, or journal article that provides motivation for the distributions and assumptions you use. The focus is on learning how to model randomness and uncertainty using tools from this course.

🎯 Learning Objectives

By the end of this project, you should be able to:

- Translate a real-world decision problem into a statistical simulation framework.
- Choose and justify appropriate probability models and distributions.
- Implement a reproducible Monte Carlo simulation in R.
- Analyze the results using inferential and descriptive tools (e.g., confidence intervals, hypothesis tests - ***Optional***).
- Communicate insights clearly and professionally in written and oral formats.

👥 Team Structure

- Teams of 1–3 students.
- Assign a **Team Leader** to oversee progress and communication.
- Each member must actively contribute and present a portion of the final deliverables.
- Peer evaluations may affect individual grades in cases of unequal participation.

🗂️ Project Tasks

1. Choose a Context & Define Your Simulation Goal

- Start with a real-world question (e.g., from a journal article, online dataset, or report).

- Example: “How would increased demand variability affect order delays in a supply chain?” or “How often would a cancer screening test miss a diagnosis under new sensitivity thresholds?”
- Translate this into a *simulation* problem (define variables, events, outcomes).

2. Build the Statistical Framework

- Identify key random variables and their distributions.
- Distributions must be covered in the course (e.g., Exponential, Poisson, Binomial, Normal, Gamma, Lognormal).
- Justify each choice and define assumptions clearly.
- Explain how you will simulate the process (what loops? what metrics? how many repetitions?).

3. Implement in R

- Use `set.seed()` for reproducibility.
- Your code must be:
  - Well-annotated with comments.
  - Modular, ideally using functions.
  - Saved and submitted as a `.R` or `.Rmd` file.
- Use vectorized operations where possible to ensure efficiency.

4. Analyze & Interpret

- Summarize simulation results using appropriate statistical methods:
  - Descriptive summaries (means, variances).
  - Visualization (histograms, density plots, etc.).
  - Confidence intervals and hypothesis tests (where relevant) - ***Optional***.
- Discuss what these results mean in context — what decision could the client make?

5. Communicate Your Findings

- **Written Report** (5–7 pages):
  - Problem background.
  - Simulation design and methods.
  - Results and interpretation.
  - Limitations and possible extensions.

- **Live Presentation** (15–20 min): Each team member should explain a substantial part.

🧠 AI Usage Policy

- Up to 10% of code or text can originate from AI tools (e.g., ChatGPT, Copilot).
- You must document any AI use in an **Appendix: AI Log**:
  - What tools were used.
  - What parts of the code/text it helped generate.
  - How you reviewed/verified the results.
- Clearly label AI-assisted content in comments or footnotes.

📅 Timeline

Deliverable	Description	Due
Week 4	Submit group info, topic, and a brief idea	June 20
Week 6	Status report with partial code and participation update	July 4
Week 8	Final report, code, and presentation	July 18

✅ Evaluation Rubric Highlights

- Clarity and feasibility of simulation design.
- Justification of statistical assumptions.
- Quality and reproducibility of code.
- Accuracy and depth of analysis.
- Professional communication (written and oral).
- Transparency and ethical use of AI tools.