



HOMEWORK II

O objetivo do segundo *homework* é revisar e familiarizar-se com os conceitos de probabilidade e variáveis aleatórias discretas e contínuas. Podem consultar os slides das aulas para os conceitos teóricos e os códigos em R para a parte computacional [The objective of the second homework is to revise and familiarise with the concepts of probability and discrete and continuous random variables. You might refer to the lecture slides for the theoretical background and for code in R for the computational part].

INSTRUÇÕES: Você deve entregar os seguintes itens [You must submit the following]:

- RELATÓRIO: Resolva cada exercício utilizando cálculos teóricos e descreva, de forma clara, as principais etapas do seu raciocínio. Em seguida, compare os resultados obtidos com aqueles gerados utilizando a linguagem R. Os pontos estão distribuídos igualmente entre as questões [Solve each exercise using theoretical calculations, and clearly describe the main steps of your reasoning. Then, compare your results with those obtained using the R programming environment. Points are equally distributed among the questions].
 - O relatório deve seguir uma estrutura lógica e bem organizada, garantindo clareza e facilitando a avaliação. Para cada exercício, inclua um parágrafo introdutório curto, apresentando brevemente o objetivo geral da tarefa e uma descrição do conjunto de dados utilizado. Descreva as ferramentas, pacotes e métodos estatísticos aplicados. Explique a resolução da questão, com os cálculos relevantes, gráficos (quando aplicável), tabelas e, principalmente, as interpretações [The report should follow a logical and well-organized structure to ensure clarity and ease of evaluation. For each exercise, include a short introductory paragraph briefly presenting the general objective of the task and a description of the dataset used. Describe the tools, packages, and statistical methods applied. Discuss the solution to the question, including relevant calculations, graphs (when applicable), tables, and especially interpretations].
 - Foque em clareza e concisão. Não inclua saídas brutas do R sem explicação, sempre interprete os resultados apresentados. Utilize gráficos e tabelas de forma eficiente, garantindo que cada elemento visual tenha título, legenda e seja pertinente à análise. Evite informações redundantes e mantenha uma formatação limpa e consistente ao longo do documento [Focus on clarity and conciseness. Do not include raw R output without explanation, always interpret the results. Use graphs and tables efficiently, ensuring each visual element has a title, legend, and is relevant to the analysis. Avoid redundancy and keep formatting clean and consistent throughout the document].
 - Utilize uma linguagem técnica, objetiva e imprecisa. Evite expressões como “eu acho” ou “nós acreditamos”. Prefira construções como “os resultados indicam que...”

ou “observou-se que...”. Estruture bem os parágrafos, organize o conteúdo por seções e identifique claramente as respostas para cada item da atividade [Use a technical, objective, and impersonal tone. Avoid expressions like “I think” or “we believe”. Prefer formulations such as “the results indicate that...” or “it was observed that...”. Structure your paragraphs well, organize the content into sections, and clearly identify the answers to each item of the assignment].

- LISTAGEM DE CÓDIGO: Envie o código utilizado para realizar a análise. O código deve ser funcional e executável, com todos os pacotes e dependências devidamente especificados. Ele pode ser incluído ao final do relatório como apêndice ou enviado separadamente em um arquivo compactado (.zip) [Submit the code used to perform the analysis. The code must be functional and executable, with all packages and dependencies clearly specified. It may be included at the end of the report as an appendix or submitted separately in a compressed file (.zip)].
- Crie e compartilhe com a professora um repositório Git (por exemplo, no GitHub)¹, de forma que seja possível acompanhar o desenvolvimento do projeto ao longo do tempo. O repositório deve conter: (i) o relatório final em PDF, (ii) todo o código-fonte utilizado na análise, (iii) quaisquer arquivos auxiliares (como scripts de acesso a dados ou arquivos de configuração), e (iv) um arquivo README com descrição do projeto, instruções claras de execução (incluindo dependências) e a identificação das contribuições individuais de cada membro do grupo. Submissões sem repositório acessível não serão avaliadas [Create and share with the instructor a Git-based repository (e.g., on GitHub)², so that the development can be followed throughout the process. The repository must include: (i) the final report in PDF format, (ii) all source code used in the analysis, (iii) any supporting files (such as data access scripts or configuration files), and (iv) a README file containing a project description, clear execution instructions (including dependencies), and a statement of each group member’s contributions. Submissions without an accessible repository will not be evaluated].

É permitido consultar recursos externos; no entanto, todas as fontes devem ser devidamente citadas. Caso opte por utilizar ferramentas de IA na realização do trabalho, faça isso de forma responsável, para esclarecer conceitos ou verificar seu raciocínio, e não para gerar soluções completas. Utilizar IA para produzir respostas completas compromete seu aprendizado e infringe as normas de integridade acadêmica. Inclua tanto o prompt quanto a resposta gerada pela IA em um apêndice do relatório ou em um arquivo separado no repositório Git [You are allowed to consult external resources; however, all sources must be properly cited. If you choose to use AI tools while working on your homework, do so responsibly, to clarify concepts or check your reasoning, not to generate complete solutions. Using AI to produce full answers undermines your learning and violates academic integrity policies. Include both the prompt and the AI-generated output in an appendix of your homework report or in a separate file in your Git repository].

¹ Você pode utilizar outra plataforma baseada em Git, desde que o repositório esteja acessível sem restrições à professora.

O trabalho tem que ser realizado individualmente. O relatório pode ser enviado pelo SIGAA com o prazo comunicado no sistema. Os atrasos serão penalizados: <24h=20% de penalidade, <48h=40% de penalidade, etc. [The work must be done individually. The report must be sent through SIGAA, with deadline given into the system. Late submissions will be penalised (<24h: -20%, <48: -40%, etc)].

QUESTÃO 1

Em um restaurante muito frequentado, aproximadamente 70% dos clientes pedem uma sobremesa após o prato principal. Seja X a variável aleatória que representa o número de clientes que pedem sobremesa em uma amostra aleatória de $n = 50$ clientes [In a popular restaurant, 7 out of 10 customers order a dessert after their main course. Let X be the random variable describing the "number of customers who order a dessert" in a random sample of size $n = 50$].

1. Determine a função de distribuição de [Define the distribution function of] X .
2. Construa os gráficos da função massa de probabilidade (PMF) e da função distribuição acumulada (CDF) de [Plot the probability mass function (PMF) and the cumulative distribution function (CDF) of] X .
3. Calcule o valor esperado, a variância e o desvio padrão de [Compute the expected value, the variance and standard deviation of] X .
4. Calcule a probabilidade de [Compute the probability]:
 - (a) $P(X \geq 20)$.
 - (b) $P(30 < X < 43)$.
 - (c) $P(X = 31)$.
5. Suponha que o restaurante estoque sobremesas com base na demanda esperada. Como o uso da distribuição de X poderia ajudar a reduzir desperdício e evitar falta de produtos [Suppose the restaurant stocks desserts based on expected demand. How could using the distribution of X help reduce waste while avoiding shortages]?
6. Como mudanças em p (por exemplo, sobremesa se torna mais popular, $p = 0.8$) ou em n (número de clientes) afetariam a forma e as probabilidades de X [How would changes in p (e.g., dessert becomes more popular, $p = 0.8$) or in n (number of customers) affect the shape and probabilities of X]?

¹ Você pode utilizar o modelo do HW [You might use the template provided for HW].

QUESTÃO 2

Um site realiza uma pesquisa online e oferece uma recompensa a um usuário selecionado aleatoriamente que responde a uma série de perguntas. Cada um dos 10 milhões de visitantes diários tem, independentemente, probabilidade $p = 10^{-7}$ de ganhar a recompensa [A website runs an online survey and offers a reward to a randomly selected user who answers a series of questions. Each of the 10 million daily visitors independently has probability $p = 10^{-7}$ of winning the reward].

1. Encontre uma aproximação simples e adequada para a função de massa de probabilidade (PMF) do número de vencedores em um dia, X . Justifique claramente se essa aproximação é apropriada para os valores dados de n and p [Find a simple, good approximation for the probability mass function (PMF) of the number of winners in a day, X . Clearly justify if this approximation is appropriate for the given n and p].
2. Calcule o valor esperado, $E[X]$, e a variância $\text{Var}(X)$, usando tanto a distribuição exata quanto a aproximada. Comente sobre a semelhança entre os resultados [Compute the expected value, $E[X]$, and variance, $\text{Var}(X)$, using the exact and the approximate distribution. Comment on the similarity of the results].
3. Suponha que você ganhe a recompensa, mas que possa haver outros vencedores. Seja $W \sim \text{Pois}(1)$ o número de vencedores além de você. Se houver vários vencedores, o prêmio é sorteado aleatoriamente entre todos eles. Encontre a probabilidade de que você realmente receba o prêmio [Suppose you win the reward, but there may be other winners. Let $W \sim \text{Pois}(1)$ represent the number of winners other than you. If there are multiple winners, the prize is awarded randomly among all winners. Find the probability that you actually win the prize].
4. Gere um grande número de simulações diárias para o número de vencedores. Crie uma comparação visual entre os resultados empíricos e a aproximação considerada no item ???. Descreva brevemente o que a visualização indica sobre a qualidade da aproximação [Generate a large number of days' worth of simulated outcomes for the number of winners. Create a visual comparison of the empirical results against the approximation you chose in item ???. Briefly describe what the visualisation indicates about the quality of the approximation].

QUESTÃO 3

Você é responsável por monitorar a temperatura de uma CPU multicore em uma unidade de processamento embarcada. Sob carga normal, a temperatura da CPU apresenta flutuações devido a mudanças na carga de trabalho, nas condições ambientais e na eficiência do sistema de resfriamento. Testes mostram que a temperatura em regime estacionário da CPU segue uma distribuição normal com temperatura média $\mu = 62^\circ\text{C}$ e desvio-padrão $\sigma = 3,5^\circ\text{C}$. Sua tarefa é simular medições de temperatura da CPU e analisar suas propriedades estatísticas [You are responsible for monitoring the temperature of a

multi-core CPU in an embedded processing unit. Under normal load, the CPU's temperature fluctuates due to changes in workload, ambient conditions, and cooling efficiency. Testing shows that the CPU's steady-state temperature follows a normal distribution with mean temperature $\mu = 62^\circ\text{C}$ and standard deviation $\sigma = 3.5^\circ\text{C}$. Your task is to simulate CPU temperature measurements and analyse their statistical properties].

1. Crie uma função que gere valores com distribuição normal usando a transformação de Box-Muller³, a partir de entradas aleatórias uniformes. Especificamente [Create a function that generates normally distributed values using the Box-Muller transform⁴, starting from uniform random inputs. Specifically](#):

- (a) Gere duas variáveis aleatórias uniformes independentes [[Generate two independent uniform random variables](#)]: $U_1, U_2 \sim \text{Unif}(0, 1)$.
- (b) Calcule dois valores normais padrão usando as fórmulas de Box-Muller [[Compute two standard normal values using the Box-Muller formulas](#)]:

$$Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2), \quad Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

Z_1 e Z_2 são variáveis aleatórias independentes com distribuição normal padrão. Concatene-as para formar um vetor Z de valores normais padrão [[\$Z_1\$ and \$Z_2\$ are independent random variables with a standard normal distribution. Concatenate them to form a vector \$Z\$ of standard normal values](#)].

- (c) Converta cada valor normal padrão para a distribuição de temperatura da CPU [[Convert each standard normal value to the CPU temperature distribution](#)]:

$$T = 62 + 3.5 Z$$

2. Use seu gerador de números aleatórios para gerar 1.000 medições de temperatura da CPU [[Use your RNG to generate 1,000 CPU temperature measurements](#)].

Gere mais 1.000 valores de temperatura utilizando o gerador de números aleatórios normal embutido do R, com a mesma média e desvio-padrão [[Generate an additional 1,000 temperature values using R's built-in normal RNG with the same mean and standard deviation](#)].

3. Para ambos os conjuntos de dados simulados, calcule [[For both simulated datasets, compute](#)]:
 - (a) Média amostral [[Sample mean](#)].
 - (b) Desvio-padrão amostral [[Sample standard deviation](#)].
 - (c) Temperatura mínima e máxima observada [[Minimum and maximum observed temperature](#)].

³ https://pt.wikipedia.org/wiki/Transformação_de_Box-Muller

⁴ https://en.wikipedia.org/wiki/Box-Muller_transform

- (d) Probabilidade empírica e teórica [Empirical and theoretical probability] $P(T > 68)$.
- (e) Probabilidade empírica e teórica [Empirical and theoretical probability] $P(60 < T < 65)$.
- (f) Probabilidade teórica [Theoretical probability] $P(T > 75)$.

Algum dos conjuntos de dados simulados (1.000 amostras) contém valores acima de 75°C? Caso não, explique por que eventos raros requerem tamanhos de amostra grandes para serem observados [Do either of the simulated datasets (1,000 samples) contain any values above 75°C? If not, explain why rare events require large sample sizes to be observed].

4. Visualize os resultados criando [Visualize the results by creating]:
 - (a) Um histograma das temperaturas simuladas da CPU (pode plotar os dois conjuntos de dados separadamente ou sobrepostos) [A histogram of your simulated CPU temperatures (you may plot the two datasets separately or overlay them)].
 - (b) A função densidade de probabilidade (PDF) normal teórica (média 62 °C, desvio padrão 3,5 °C) sobreposta ao histograma [Theoretical normal PDF (mean 62 °C, standard deviation 3.5 °C) overlaid on the histogram].
5. Discuta seus resultados respondendo às seguintes perguntas: As distribuições empíricas da temperatura da CPU se assemelham à curva normal teórica? Quão próximas estão a média amostral e o desvio-padrão amostral dos valores esperados 62 °C e 3,5 °C? Há diferenças perceptíveis entre o conjunto de dados gerado com seu RNG manual e o produzido pelo RNG embutido do R? Como essa simulação pode ajudar na avaliação de estratégias de resfriamento ou de escalonamento dinâmico de clock? Por que geradores de números aleatórios uniformes são a base dos sistemas de RNG [Discuss the results by answering the following questions: Do the empirical CPU temperature distributions resemble the theoretical normal curve? How close are your sample mean and standard deviation to the expected values of 62 °C and 3.5 °C? Are there any noticeable differences between the dataset generated with your manual RNG and the one produced using R's built-in RNG? How can this simulation help in evaluating cooling strategies or dynamic clock scaling? Why are uniform random generators the backbone of RNG systems]?