



# UFC

## ESTATÍSTICA PARA ENGENHARIA

### HOMEWORK 2

#### Alunos:

JOAO VICTOR MARQUES FALCÃO, 567357

LUCAS MARTINS MENEZES, 565788

# 1 Sobremesa de Restaurante

O comportamento de consumo em estabelecimentos comerciais é um fator de grande interesse tanto do ponto de vista econômico quanto estatístico. Em restaurantes de grande movimento, decisões aparentemente simples, como optar ou não por uma sobremesa após o prato principal, refletem padrões de preferência dos clientes e podem ser analisadas de forma quantitativa a partir de modelos probabilísticos.

O Problema 1 aborda exatamente esse tipo de situação. Considera-se um restaurante muito frequentado, no qual aproximadamente 70% dos clientes optam por pedir sobremesa após a refeição principal. A partir desse cenário, observa-se uma amostra aleatória de  $n = 50$  clientes, com o objetivo de analisar o número de indivíduos que realizam esse pedido.

Do ponto de vista estatístico, cada cliente representa um ensaio independente, no qual existem apenas dois resultados possíveis: pedir sobremesa ou não pedir sobremesa. Assim, o foco da análise está na contagem do número de sucessos — definidos como pedidos de sobremesa — ao longo de um número fixo de observações, sob uma probabilidade constante de ocorrência. Esse tipo de estrutura fornece a base para a modelagem da variável aleatória associada ao problema e para a aplicação de conceitos fundamentais de variáveis aleatórias discretas e distribuições de probabilidade.

## 1.1 Função de Distribuição

Seja  $X$  a variável aleatória que representa o número de clientes que pedem sobremesa após o prato principal em uma amostra aleatória de  $n = 50$  clientes. Cada cliente observado pode ser entendido como um ensaio independente, no qual existem apenas dois resultados possíveis: pedir sobremesa ou não pedir sobremesa.

De acordo com o enunciado do problema, a probabilidade de um cliente pedir sobremesa é aproximadamente constante e igual a  $p = 0,7$ . Dessa forma, o experimento consiste em um número fixo de ensaios independentes, cada um com dois possíveis desfechos e com probabilidade constante de sucesso, sendo o sucesso definido como o pedido de sobremesa.

Essas características satisfazem integralmente as condições necessárias para a aplicação do modelo de distribuição binomial. Assim, a variável aleatória  $X$  segue uma distribuição binomial com parâmetros  $n = 50$  e  $p = 0,7$ , sendo formalmente representada por:

$$X \sim \text{Bin}(50, 0,7)$$

## 1.2 Função Massa de Probabilidade (PMF) e Função Distribuição Acumulada (CDF)

Uma vez estabelecido que  $X \sim \text{Bin}(50, 0,7)$ , podemos construir tanto a **função massa de probabilidade (PMF)** quanto a **função distribuição acumulada (CDF)** da variável aleatória.

### PMF (Probability Mass Function)

A PMF descreve a probabilidade de a variável aleatória assumir exatamente um valor  $x$ , isto é,  $P(X = x)$ . Para a distribuição binomial, a PMF é dada por:

$$P(X = x) = \binom{50}{x} (0,7)^x (0,3)^{50-x}, \quad x = 0, 1, 2, \dots, 50$$

## CDF (Cumulative Distribution Function)

A CDF representa a probabilidade acumulada até um certo valor, isto é,  $P(X \leq x)$ . Como  $X$  é discreta, a CDF é definida como a soma das probabilidades da PMF até o valor  $x$ :

$$F(x) = P(X \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{50}{k} (0,7)^k (0,3)^{50-k}$$

Como se trata de uma variável aleatória discreta, o gráfico da CDF possui o formato de *escada*, apresentando saltos apenas nos valores inteiros de  $x$ .

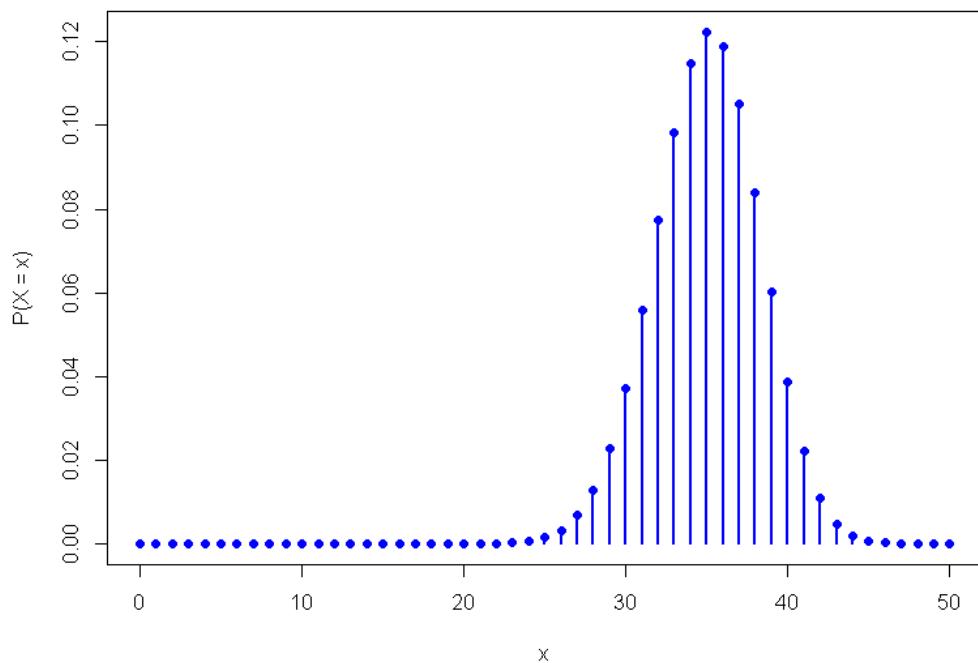
## Construção dos gráficos no R

Para construir os gráficos da PMF e da CDF, utilizamos as funções nativas do R `dbinom()` (para a PMF) e `pbinom()` (para a CDF), considerando o suporte  $x = 0, 1, \dots, 50$ .

```
1 # Parametros da distribuicao binomial
2 n <- 50
3 p <- 0.7
4
5 # Suporte da variavel aleatoria
6 x <- 0:n
7
8 # PMF e CDF
9 pmf <- dbinom(x, size = n, prob = p)
10 cdf <- pbinom(x, size = n, prob = p)
11
12 # Grafico da PMF
13 plot(x, pmf,
14       type = "h",
15       lwd = 2,
16       col = "blue",
17       main = "PMF - Binomial(n=50, p=0.7)",
18       xlab = "x",
19       ylab = "P(X=x)")
20 points(x, pmf, pch = 16, col = "blue")
21
22 # Grafico da CDF
23 plot(x, cdf,
24       type = "s",
25       lwd = 2,
26       col = "red",
27       main = "CDF - Binomial(n=50, p=0.7)",
28       xlab = "x",
29       ylab = "P(X<=x)")
```

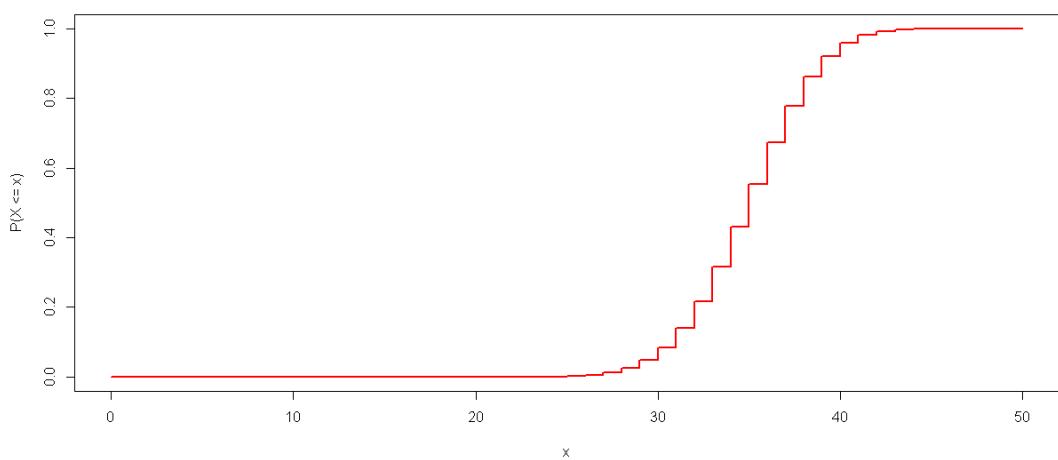
Como mostrado na Figura 1, a função massa de probabilidade (PMF) da Binomial(50, 0,7) é unimodal e concentra-se em torno do valor esperado  $E[X] = np = 35$ . A Figura 2 apresenta o crescimento em degraus típico da CDF de uma variável discreta.

**PMF - Binomial( $n = 50, p = 0.7$ )**



**Figura 1:** PMF da distribuição Binomial( $n = 50, p = 0,7$ ).

**CDF - Binomial( $n = 50, p = 0.7$ )**



**Figura 2:** CDF da distribuição Binomial( $n = 50, p = 0,7$ ).

A partir dos gráficos obtidos, observa-se que a PMF apresenta um comportamento unimodal, com maior concentração de probabilidade em torno do valor esperado da distribuição, dado por  $E[X] = np = 50 \cdot 0,7 = 35$ . O que, institivamente, faz todo o sentido. Se temos 50 clientes e a estatística geral implica que 70% de todos os clientes pedem sobremesa após o prato principal, o mais comum é que 70% dos 50 clientes observados peçam sobremesa. nesse caso, existe aproximadamente 12% de chance de que, aproximadamente, 35 clientes peçam sobremesa. Já a CDF cresce de forma monotônica e em degraus, acumulando a probabilidade total até atingir 1 em  $x = 50$ .

### 1.3 Valor Esperado, Variância e Desvio Padrão de $X$

Uma vez estabelecido que a variável aleatória  $X$  segue uma distribuição binomial com parâmetros  $n = 50$  e  $p = 0,7$ , podemos calcular suas principais medidas teóricas: o valor esperado, a variância e o desvio padrão.

Para uma variável aleatória com distribuição binomial  $X \sim \text{Bin}(n, p)$ , o valor esperado é dado por:

$$E[X] = np$$

Substituindo os valores do problema, obtemos:

$$E[X] = 50 \cdot 0,7 = 35$$

Assim, espera-se que, em média, 35 dos 50 clientes observados peçam sobremesa após o prato principal.

#### Dedução da Variância da Distribuição Binomial

Sabemos, com o que foi apresentado explicitamente em sala de aula, que a variância  $\text{Var}(X)$  de uma distribuição binomial é:

$$\text{Var}(X) = np(1 - p).$$

Ainda assim, podemos obter essa equação a partir de outras ferramentas mais elementais a fim de possamos obter a equação simplificada da variância de qualquer distribuição sem decorar. De forma geral, a variância de uma variável aleatória discreta pode ser expressa pela identidade fundamental:

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Para deduzir a variância da distribuição binomial, considera-se que a variável aleatória  $X$  pode ser escrita como a soma de  $n$  variáveis aleatórias de Bernoulli independentes:

$$X = \sum_{i=1}^n Y_i,$$

onde cada  $Y_i$  assume valor 1 se o  $i$ -ésimo cliente pede sobremesa e 0 caso contrário, com  $P(Y_i = 1) = p$ .

Para uma variável de Bernoulli, temos:

$$E[Y_i] = p \quad \text{e} \quad E[Y_i^2] = E[Y_i] = p,$$

uma vez que  $Y_i^2 = Y_i$ .

Utilizando a linearidade da esperança matemática, obtemos:

$$E[X] = E\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n E[Y_i] = np,$$

confirmando o valor esperado já calculado.

Agora, calculamos  $E[X^2]$ . Observando que:

$$X^2 = \left(\sum_{i=1}^n Y_i\right)^2 = \sum_{i=1}^n Y_i^2 + 2 \sum_{i < j} Y_i Y_j,$$

temos:

$$E[X^2] = \sum_{i=1}^n E[Y_i^2] + 2 \sum_{i < j} E[Y_i Y_j].$$

Como os ensaios são independentes:

$$E[Y_i Y_j] = E[Y_i] E[Y_j] = p^2,$$

e, portanto:

$$E[X^2] = np + 2 \binom{n}{2} p^2 = np + n(n-1)p^2.$$

Aplicando a fórmula fundamental da variância:

$$\text{Var}(X) = E[X^2] - (E[X])^2,$$

obtemos:

$$\text{Var}(X) = [np + n(n-1)p^2] - (np)^2 = np + n(n-1)p^2 - n^2p^2.$$

Simplificando a expressão:

$$\text{Var}(X) = np(1-p).$$

## Aplicação ao Problema

Substituindo os valores  $n = 50$  e  $p = 0,7$ , temos:

$$\text{Var}(X) = 50 \cdot 0,7 \cdot (1 - 0,7) = 50 \cdot 0,7 \cdot 0,3 = 10,5.$$

O desvio padrão é dado pela raiz quadrada da variância:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{10,5} \approx 3,24.$$

Esses resultados indicam que, embora o número médio esperado de clientes que pedem sobremesa seja 35, ocorrem variações naturais em torno desse valor, com uma dispersão típica de aproximadamente 3,24 clientes.

Para confirmar esses valores obtidos manualmente podemos utilizar o seguinte script simples em R, também aplicando as fórmulas que já havíamos obtido.

```

1 valor_esperado <- sum(x * pmf)
2
3 # Variância via definição
4 variância <- sum(x^2 * pmf) - valor_esperado^2
5
6 # Desvio padrão
7 desvio_padrao <- sqrt(variancia)
8
9 valor_esperado
10 variância
11 desvio_padrao

```

```

> valor_esperado
[1] 35
> variância
[1] 10.5
> desvio_padrao
[1] 3.24037

```

**Figura 3:** Saída do R confirmando  $E[X]$ ,  $\text{Var}(X)$  e  $\sigma_X$  para  $X \sim \text{Bin}(50, 0,7)$ .

A validação computacional (ver Figura 3) confirma os valores teóricos obtidos.

## 1.4 Cálculo de Probabilidades

Considerando a variável aleatória já estabelecida e analisada  $X \sim \text{Bin}(50, 0,7)$ . Nesta subseção, calculamos as probabilidades solicitadas utilizando a função de distribuição acumulada da distribuição binomial.

(a)  $P(X \geq 20)$

A probabilidade de  $X$  assumir valores maiores ou iguais a 20 pode ser calculada utilizando o complemento da função distribuição acumulada:

$$P(X \geq 20) = 1 - P(X \leq 19).$$

Assim, temos:

$$P(X \geq 20) = 1 - \sum_{k=0}^{19} \binom{50}{k} (0,7)^k (0,3)^{50-k}.$$

Numericamente, obtemos:

$$P(X \geq 20) = 1 - P(X \leq 19) \approx 0,9999971533.$$

(b)  $P(30 < X < 43)$

Como  $X$  é uma variável aleatória discreta, o evento  $30 < X < 43$  equivale a  $31 \leq X \leq 42$ . Logo:

$$P(30 < X < 43) = P(31 \leq X \leq 42) = P(X \leq 42) - P(X \leq 30).$$

Portanto:

$$P(30 < X < 43) = \sum_{k=0}^{42} \binom{50}{k} (0,7)^k (0,3)^{50-k} - \sum_{k=0}^{30} \binom{50}{k} (0,7)^k (0,3)^{50-k}.$$

Como  $X$  é discreta,  $30 < X < 43 \iff X \in \{31, 32, \dots, 42\}$ . Assim:

$$P(30 < X < 43) = \sum_{k=31}^{42} P(X = k) \approx 0,9079331984.$$

**(c)**  $P(X = 31)$

A probabilidade pontual de  $X$  assumir exatamente o valor 31 é dada diretamente pela função massa de probabilidade da binomial:

$$P(X = 31) = \binom{50}{31} (0,7)^{31} (0,3)^{19}.$$

A probabilidade pontual é:

$$P(X = 31) = \binom{50}{31} (0,7)^{31} (0,3)^{19} \approx 0,0557572778.$$

### Validação Computacional no R

Os valores obtidos teoricamente foram calculados e validados no software R utilizando a função `pbinom()` para probabilidades acumuladas e `dbinom()` para probabilidades pontuais:

```

1 # Parametros da distribuicao
2 n <- 50
3 p <- 0.7
4
5 # (a) P(X >= 20)
6 prob_a <- 1 - pbinom(19, size = n, prob = p)
7
8 # (b) P(30 < X < 43) = P(31 <= X <= 42)
9 prob_b <- pbinom(42, size = n, prob = p) - pbinom(30, size = n,
prob = p)
10
11 # (c) P(X = 31)
12 prob_c <- dbinom(31, size = n, prob = p)
13
14 prob_a
15 prob_b
16 prob_c

```

Quantidade	Teórico	R	Dif. abs.
$E[X]$	35	35	0
$\text{Var}(X)$	10,5	10,5	0
$\sigma_X$	3,24037	3,24037	0
$P(X \ge 20)$	0,9999971533	0,9999971533	$\approx 0$
$P(30 < X < 43)$	0,9079331984	0,9079331984	$\approx 0$
$P(X = 31)$	0,0557572778	0,0557572778	$\approx 0$

```

> prob_a
[1] 0.9999972
> prob_b
[1] 0.9079332
> prob_c
[1] 0.05575728

```

**Figura 4:** Saída do R confirmando  $P(X \geq 20)$ ,  $P(30 < X < 43)$  e  $P(X = 31)$  para  $X \sim \text{Bin}(50, 0,7)$ .

Esses resultados confirmam os cálculos analíticos apreapresentados (ver Figura 4), permitindo quantificar a probabilidade de ocorrência de diferentes intervalos e valores específicos do número de clientes que pedem sobremesa em uma amostra de 50 indivíduos.

## 1.5 Como a distribuição de $X$ ajuda a reduzir desperdício e evitar falta de produtos

Uma política simples baseada apenas no valor esperado  $E[X] = np$  fornece um ponto de partida (demanda média). No entanto, operar exatamente em  $E[X]$  ignora a variabilidade natural da demanda, representada por  $\text{Var}(X) = np(1 - p)$ . Na prática, o restaurante precisa escolher um estoque  $s$  (número de sobremesas disponíveis) de modo a controlar a probabilidade de faltar sobremesa.

Considere novamente  $X \sim \text{Bin}(n, p)$  como o número de clientes, dentre  $n$  observados em um período de interesse, que pedem sobremesa, onde  $p$  é a probabilidade de um cliente pedir sobremesa. O uso explícito da distribuição de  $X$  permite transformar uma decisão operacional (quantas sobremesas produzir/estocar) em um problema quantitativo de **gestão de risco**, equilibrando dois custos opostos:

- **Desperdício (overstock):** produzir sobremesas demais implica perda por vencimento, armazenamento e custo de insumos.
- **Falta (stockout):** produzir sobremesas de menos implica perda de vendas, insatisfação de clientes e possível impacto reputacional.

Se o restaurante decide estocar  $s$  sobremesas, então ocorrerá **falta** quando  $X > s$ . Logo, a probabilidade de falta é:

$$P(\text{falta}) = P(X > s) = 1 - P(X \leq s).$$

De forma equivalente, a probabilidade de **atender toda a demanda** é:

$$P(\text{atender}) = P(X \leq s).$$

Assim, o restaurante pode escolher  $s$  com base em um **nível de serviço** desejado. Por exemplo, se a gerência quer atender a demanda em pelo menos 95% dos dias, escolhe-se  $s$  tal que:

$$P(X \leq s) \geq 0,95.$$

O menor valor inteiro  $s$  que satisfaz essa condição é, essencialmente, o **percentil 95** da distribuição de  $X$ . Essa abordagem é preferível a estocar apenas  $E[X]$ , pois incorpora a variabilidade e controla o risco de falta.

Além disso, pode-se quantificar **desperdício esperado**. Se  $s$  é o estoque e  $X$  a demanda, o número de sobremesas não vendidas (sobras) é:

$$W = \max(s - X, 0).$$

Logo, o desperdício esperado é:

$$E[W] = E[\max(s - X, 0)] = \sum_{x=0}^s (s - x) P(X = x).$$

Analogamente, a **falta esperada** (demanda não atendida) é:

$$L = \max(X - s, 0), \quad E[L] = \sum_{x=s+1}^n (x - s) P(X = x).$$

Essas expressões permitem comparar quantitativamente políticas de estoque diferentes: aumentar  $s$  reduz  $P(\text{falta})$  e  $E[L]$ , mas aumenta  $E[W]$ ; diminuir  $s$  reduz desperdício, mas aumenta falta. Portanto, a distribuição de  $X$  fornece uma base matemática para um equilíbrio racional entre custos, em vez de decisões puramente intuitivas.

**Síntese operacional.** Ao estimar  $p$  (por dados históricos) e definir o horizonte com  $n$  clientes, o restaurante pode:

- escolher  $s$  como um percentil (ex.: 90%, 95%, 99%) de  $X$  para controlar risco de falta;
- calcular  $P(X > s)$  para estimar a probabilidade de ruptura;
- estimar  $E[W]$  e  $E[L]$  para comparar desperdício esperado e falta esperada;
- ajustar  $s$  conforme o custo relativo de falta versus desperdício.

## 1.6 Efeitos de mudanças em $p$ e em $n$ na forma e nas probabilidades de $X$

A distribuição binomial é determinada por dois parâmetros:  $n$  e  $p$ . Mudanças nesses parâmetros afetam diretamente:

- o centro (média),
- a dispersão (variância/desvio padrão),
- o formato (assimetria e concentração de probabilidades).

**Mudança em  $p$  (ex.: sobremesa mais popular,  $p = 0,8$ ).** Para  $X \sim \text{Bin}(n, p)$ :

$$E[X] = np \quad \text{e} \quad \text{Var}(X) = np(1 - p).$$

Assim, mantendo  $n = 50$ :

$$E[X]_{p=0.7} = 50 \cdot 0.7 = 35 \quad \longrightarrow \quad E[X]_{p=0.8} = 50 \cdot 0.8 = 40.$$

Ou seja, o centro da distribuição desloca-se para a direita: espera-se mais pedidos de sobremesa.

Quanto à variância:

$$\text{Var}(X)_{p=0.7} = 50 \cdot 0.7 \cdot 0.3 = 10.5,$$

$$\text{Var}(X)_{p=0.8} = 50 \cdot 0.8 \cdot 0.2 = 8.0.$$

Note que, ao aumentar  $p$  de 0,7 para 0,8, a variância **diminui**, pois o termo  $p(1 - p)$  é máximo em  $p = 0.5$  e diminui quando  $p$  se aproxima de 0 ou 1. Isso significa que a demanda fica mais **concentrada** ao redor do novo valor esperado. Em termos práticos, o restaurante passa a prever uma demanda maior, mas com menor variabilidade relativa.

**Forma (assimetria) ao variar  $p$ .** A distribuição binomial é aproximadamente simétrica quando  $p \approx 0.5$ . Para  $p > 0.5$ , a distribuição tende a ser levemente assimétrica à esquerda (maior massa de probabilidade em valores altos, com cauda para valores menores). Para  $p < 0.5$ , tende a ser assimétrica à direita. Assim, ao aumentar  $p$  de 0,7 para 0,8, a distribuição torna-se ainda mais concentrada em valores altos e com cauda mais pronunciada à esquerda.

**Mudança em  $n$  (número de clientes observados).** Mantendo  $p$  fixo, aumentar  $n$  aumenta a média e altera a dispersão:

$$E[X] = np \quad \text{cresce linearmente com } n,$$

$$\text{Var}(X) = np(1 - p) \quad \text{também cresce linearmente com } n.$$

Logo, com mais clientes, o número esperado de sobremesas aumenta, e a variabilidade absoluta também aumenta. Entretanto, um ponto importante é que a **variabilidade relativa** tende a diminuir, pois o coeficiente de variação (aproximado) é:

$$\text{CV} \approx \frac{\sqrt{np(1 - p)}}{np} = \sqrt{\frac{1 - p}{np}}.$$

Portanto, conforme  $n$  aumenta, CV diminui, indicando que a demanda torna-se relativamente mais previsível em comparação ao seu tamanho médio.

**Implicações nas probabilidades.** Alterações em  $p$  e  $n$  mudam probabilidades de eventos como:

$$P(X \geq k), \quad P(a \leq X \leq b), \quad P(X = k),$$

pois deslocam e/ou “apertam” a distribuição. Em particular:

- aumentar  $p$  desloca a massa de probabilidade para valores maiores de  $X$ , aumentando  $P(X \geq k)$  para  $k$  moderados;
- aumentar  $n$  aumenta a escala do problema (mais pedidos esperados), e concentra a distribuição relativamente ao redor de  $np$ , tornando mais útil a análise via percentis e níveis de serviço para dimensionar estoque.

Por fim, podemos concluir que o uso de  $X \sim \text{Bin}(n, p)$  permite que decisões de estoque sejam tomadas com base em níveis de risco controláveis (probabilidade de falta) e em medidas quantitativas de desperdício e falta esperados. Mudanças em  $p$  e  $n$  alteram diretamente a média e a variância de  $X$ , deslocando a distribuição e modificando a concentração de probabilidades, o que exige ajuste das políticas de estoque para manter o equilíbrio entre desperdício e ruptura de produtos.

## 2 Premiação Digital

Plataformas digitais frequentemente utilizam pesquisas online e campanhas de incentivo como estratégias para aumentar engajamento, coletar dados de usuários e estimular a participação em formulários ou questionários. Uma prática comum nesse tipo de ação é a oferta de recompensas a usuários selecionados aleatoriamente dentre um grande volume de visitantes, com o intuito de tornar a participação mais atrativa.

Do ponto de vista estatístico, esse cenário se torna particularmente interessante porque envolve um número muito grande de tentativas independentes (visitas diárias) e uma probabilidade extremamente pequena de sucesso (ganhar a recompensa) para cada indivíduo. Assim, ainda que a chance individual seja quase nula, o volume de visitantes pode gerar um número não desprezível de vencedores ao longo de um dia, tornando necessário modelar adequadamente a variável aleatória associada ao processo.

Nesta questão, considera-se que um site possui  $n = 10^7$  visitantes diáriamente, e que cada visitante, independentemente, tem probabilidade  $p = 10^{-7}$  de ganhar a recompensa. Definimos  $X$  como a variável aleatória que representa o número total de vencedores em um dia. A partir dessa definição, o problema propõe: (i) identificar uma aproximação simples e apropriada para a função massa de probabilidade de  $X$ , (ii) comparar estatisticamente os valores esperados e variâncias obtidos pela distribuição exata e pela aproximação, (iii) analisar a probabilidade de um vencedor realmente receber o prêmio quando há concorrência com outros vencedores, e (iv) validar empiricamente, por simulação, a qualidade da aproximação adotada.

A análise deste problema evidencia um caso clássico de modelagem probabilística em larga escala, onde métodos de aproximação tornam-se essenciais para simplificar cálculos e interpretar fenômenos raros em sistemas com alto volume de observações.

Utilizaremos como base de todos os códigos de validação a seguinte definição em R:

```
1 # Par metros
2 n <- 1e7
3 p <- 1e-7
4 lambda <- n * p    # = 1
5
6 cat("n=", n, "\n")
7 cat("p=", p, "\n")
8 cat("lambda=", n*p, "lambda", "\n\n")
```

### 2.1 Aproximação para a PMF de $X$ e Justificativa

Seja  $X$  o número de vencedores em um dia. Como cada um dos  $n = 10^7$  visitantes diáriamente possui, de forma independente, probabilidade  $p = 10^{-7}$  de ganhar a recompensa, o modelo **exato** para  $X$  é binomial:

$$X \sim \text{Bin}(n = 10^7, p = 10^{-7}).$$

Entretanto, trabalhar diretamente com a PMF binomial torna-se pouco prático para valores tão grandes de  $n$ . Além disso, este problema se enquadra em um caso clássico de **eventos raros**, pois  $p$  é extremamente pequeno, enquanto  $n$  é extremamente grande. Nessas condições, a distribuição binomial admite uma aproximação simples e bastante precisa pela distribuição de Poisson.

A aproximação de Poisson é adequada quando:

- $n$  é grande,

- $p$  é pequeno,
- $\lambda = np$  é de ordem moderada (fixo).

Calculando o parâmetro:

$$\lambda = np = 10^7 \cdot 10^{-7} = 1.$$

Logo, uma aproximação simples e adequada é:

$$X \approx \text{Pois}(\lambda = 1).$$

Assim, a **função massa de probabilidade aproximada** é:

$$P(X = k) \approx \frac{e^{-\lambda} \lambda^k}{k!} = \frac{e^{-1} 1^k}{k!} = \frac{e^{-1}}{k!}, \quad k = 0, 1, 2, \dots$$

### Justificativa Matemática da Aproximação

A PMF exata da binomial é:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Quando  $n \rightarrow \infty$  e  $p \rightarrow 0$  de modo que  $np = \lambda$  permaneça constante, vale o limite clássico:

$$\binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{e^{-\lambda} \lambda^k}{k!},$$

que corresponde exatamente à PMF de uma variável Poisson( $\lambda$ ).

Como neste problema temos  $n$  extremamente grande,  $p$  extremamente pequeno, e  $\lambda = np = 1$  fixo e moderado, conclui-se que a aproximação por Poisson é válida para simplificar o cálculo de probabilidades e a análise do número diário de vencedores, ainda que, evidentemente, tenhamos  $X$  como uma distribuição binomial.

## 2.2 Valor Esperado e Variância: Distribuição Exata vs. Aproximada

Pelo enunciado, o número de vencedores em um dia é modelado exatamente por uma distribuição binomial:

$$X \sim \text{Bin}(n = 10^7, p = 10^{-7}).$$

No Item 1, argumentou-se que, devido ao regime de *eventos raros* (isto é,  $n$  muito grande e  $p$  muito pequeno), uma aproximação adequada é:

$$X \approx \text{Pois}(\lambda), \quad \lambda = np.$$

Como  $np = 10^7 \cdot 10^{-7} = 1$ , a aproximação é:

$$X \approx \text{Pois}(1).$$

A seguir, calculamos  $E[X]$  e  $\text{Var}(X)$  sob as duas distribuições e comparamos os resultados.

## Distribuição Exata: Binomial

Para  $X \sim \text{Bin}(n, p)$ , valem as fórmulas:

$$E[X] = np, \quad \text{Var}(X) = np(1 - p).$$

Substituindo  $n = 10^7$  e  $p = 10^{-7}$ :

$$E[X]_{\text{exato}} = np = 10^7 \cdot 10^{-7} = 1.$$

$$\text{Var}(X)_{\text{exato}} = np(1 - p) = 1 \cdot (1 - 10^{-7}) = 0,9999999.$$

## Distribuição Aproximada: Poisson

Para uma variável aleatória  $Y \sim \text{Pois}(\lambda)$ , temos:

$$E[Y] = \lambda, \quad \text{Var}(Y) = \lambda.$$

Como  $\lambda = np = 1$ :

$$E[X]_{\text{aprox}} = 1, \quad \text{Var}(X)_{\text{aprox}} = 1.$$

## Comparação e Comentário

Observa-se que o valor esperado é **idêntico** nas duas abordagens:

$$E[X]_{\text{exato}} = E[X]_{\text{aprox}} = 1.$$

Já a variância difere apenas por um termo extremamente pequeno:

$$\text{Var}(X)_{\text{exato}} = 0,9999999 \quad \text{e} \quad \text{Var}(X)_{\text{aprox}} = 1.$$

A diferença absoluta é:

$$|\text{Var}(X)_{\text{aprox}} - \text{Var}(X)_{\text{exato}}| = 1 - 0,9999999 = 10^{-7}.$$

Em termos relativos, isso representa uma discrepância da ordem de  $10^{-7}$ , isto é, de certa forma, desprezível, para fins de modelagem e tomada de decisão. Essa aproximação convergente ocorre pois, quando  $p$  é muito pequeno, o fator  $(1 - p)$  é extremamente próximo de 1, fazendo com que:

$$np(1 - p) \approx np = \lambda.$$

Portanto, neste âmbito de análise, a aproximação de Poisson não apenas simplifica cálculos, como também preserva com alta fidelidade as principais medidas teóricas de posição e dispersão.

Para a validação dos cálculos manuais, foi utilizado o seguinte script em R:

```
1 EX_bin <- n * p
2 Var_bin <- n * p * (1 - p)
3
4 EX_pois <- lambda
5 Var_pois <- lambda
6
7 cat("==BINOMIAL(EXATO)==\n")
8 cat("E[X]=", EX_bin, "\n")
```

```

9 cat("Var(X)=", Var_bin, "\n\n")
10
11 cat("== POISSON(APROX) ==\n")
12 cat("E[X]=", EX_pois, "\n")
13 cat("Var(X)=", Var_pois, "\n\n")
14
15 cat("== DIFERENAS ==\n")
16 cat("|Var_pois - Var_bin| =", abs(Var_pois - Var_bin), "\n")
17 cat("Diferença relativa na variância =", abs(Var_pois - Var_bin)
     / Var_pois, "\n\n")

```

Obtendo o seguinte resultado:

```

> # -----
> # Item 2.2: E[X] e var(X)
> # -----
> EX_bin <- n * p
> Var_bin <- n * p * (1 - p)
> EX_pois <- lambda
> Var_pois <- lambda
> cat("== BINOMIAL (EXATO) ==\n")
== BINOMIAL (EXATO) ==
> cat("E[X] =", EX_bin, "\n")
E[X] = 1
> cat("var(X) =", Var_bin, "\n\n")
Var(X) = 0.9999999

> cat("== POISSON (APROX) ==\n")
== POISSON (APROX) ==
> cat("E[X] =", EX_pois, "\n")
E[X] = 1
> cat("var(X) =", Var_pois, "\n\n")
Var(X) = 1

> cat("== DIFERENÇAS ==\n")
== DIFERENÇAS ==
> cat("|var_pois - Var_bin| =", abs(Var_pois - Var_bin), "\n")
|var_pois - Var_bin| = 1e-07
> cat("Diferença relativa na variância =", abs(Var_pois - Var_bin) / Var_pois, "\n\n")
Diferença relativa na variância = 1e-07

```

**Figura 5:** Saída do R para  $E[X]$  e  $\text{Var}(X)$  na Binomial exata e na aproximação de Poisson ( $\lambda = 1$ ).

Para efeitos de comparação, podemos por os resultados analíticos e teóricos em uma tabela:

Medida	Binomial (exata)	Poisson (aprox.)	Dif. abs.
$E[X]$	1	1	0
$\text{Var}(X)$	0,9999999	1	$10^{-7}$

## 2.3 Probabilidade de Receber o Prêmio na Presença de Outros Vencedores

Mesmo que o usuário em questão tenha sido selecionado como vencedor, pode haver outros vencedores no mesmo dia. O enunciado define  $W \sim \text{Pois}(1)$  como o número de vencedores *além* do usuário. Assim, o número total de vencedores é  $W + 1$ . Quando há múltiplos vencedores, o prêmio é sorteado uniformemente entre todos os vencedores; logo, a probabilidade condicional de o usuário receber o prêmio, dado  $W = w$ , é:

$$P(\text{receber o prêmio} \mid W = w) = \frac{P(\text{receber o prêmio} \cap W = w)}{P(W = w)}.$$

Chamemos  $A = \text{receber o prêmio}$ , para que possamos reduzir à equação:

$$P(A \mid W = w) = \frac{P(A \cap W = w)}{P(W = w)}.$$

Assim, podemos interpretar que  $P(A \cap W = w) \equiv P(A \mid W = w)P(W = w)$  usando a Regra de Bayes. Com isso, podemos afirmar, que:

$$P(\text{receber o prêmio} \mid W = w) = \frac{1}{w + 1}.$$

Portanto, a probabilidade incondicional de realmente receber o prêmio é a esperança (expectation):

$$P(\text{receber o prêmio}) = \mathbb{E}\left[\frac{1}{W + 1}\right] = \sum_{w=0}^{\infty} \frac{1}{w + 1} P(W = w).$$

Como  $W \sim \text{Pois}(1)$ , tem-se  $P(W = w) = e^{-1} \frac{1^w}{w!} = e^{-1} \frac{1}{w!}$ . Substituindo:

$$P(\text{receber o prêmio}) = \sum_{w=0}^{\infty} \frac{1}{w + 1} e^{-1} \frac{1}{w!} = e^{-1} \sum_{w=0}^{\infty} \frac{1}{(w + 1)w!}.$$

Observa-se que  $(w + 1)w! = (w + 1)!$ , logo:

$$P(\text{receber o prêmio}) = e^{-1} \sum_{w=0}^{\infty} \frac{1}{(w + 1)!}.$$

Fazendo a mudança de índice  $j = w + 1$  (assim  $j = 1, 2, 3, \dots$ ):

$$P(\text{receber o prêmio}) = e^{-1} \sum_{j=1}^{\infty} \frac{1}{j!}.$$

Por fim, utiliza-se a expansão conhecida  $e = \sum_{j=0}^{\infty} \frac{1}{j!}$ , de modo que

$$\sum_{j=1}^{\infty} \frac{1}{j!} = e - 1.$$

Assim,

$$P(\text{receber o prêmio}) = e^{-1}(e - 1) = 1 - e^{-1}.$$

Numericamente,

$$P(\text{receber o prêmio}) = 1 - e^{-1} \approx 1 - 0,3679 = 0,6321.$$

Conclui-se que, mesmo sendo um dos vencedores, a probabilidade de de fato receber o prêmio (quando o sorteio é uniforme entre todos os vencedores) é aproximadamente 0,6321, ou seja, cerca de 63,2%.

Para a validação dos cálculos manuais, usamos o seguinte script em R:

```

1 p_teorico <- 1 - exp(-lambda)
2 cat("\n== ITEM 2.3 ==\n")
3 cat("Valor teórico: ", 1 - exp(-1), "\n", p_teorico, "\n")
4
5 # Soma truncada
6 Wmax <- 60
7 w <- 0:Wmax
8 p_soma <- sum((1 / (w + 1)) * dpois(w, lambda = lambda))
9 tail_prob <- 1 - ppois(Wmax, lambda = lambda)
10
11 cat("Aproximação por soma (0..", Wmax, ") =", p_soma, "\n", sep="")
12 cat("Massa na cauda ignorada P(W>", Wmax, ") =", tail_prob, "\n", sep="")
13
14 # Simula o
15 B <- 5e5
16 W <- rpois(B, lambda = lambda)
17 p_sim <- mean(1 / (W + 1))
18
19 cat("Estimativa por simulação =", p_sim, "\n")
20 cat("Erro absoluto vs teórico =", abs(p_sim - p_teorico), "\n\n")

```

Obtendo, assim, os seguintes resultados:

```

==== ITEM 2.3 ====
> cat("Valor teórico: 1 - exp(-1) =", p_teorico, "\n")
Valor teórico: 1 - exp(-1) = 0.6321206
> # Soma truncada
> Wmax <- 60
> w <- 0:Wmax
> p_soma <- sum((1 / (w + 1)) * dpois(w, lambda = lambda))
> tail_prob <- 1 - ppois(Wmax, lambda = lambda)
> cat("Aproximação por soma (0..60) =", p_soma, "\n", sep="")
Aproximação por soma (0..60) = 0.6321206
> cat("Massa na cauda ignorada P(W>60) =", tail_prob, "\n", sep="")
Massa na cauda ignorada P(W>60) = 0
> # Simulação
> B <- 5e5
> W <- rpois(B, lambda = lambda)
> p_sim <- mean(1 / (W + 1))
> cat("Estimativa por simulação =", p_sim, "\n")
Estimativa por simulação = 0.6323161
> cat("Erro absoluto vs teórico =", abs(p_sim - p_teorico), "\n\n")
Erro absoluto vs teórico = 0.000195507

```

**Figura 6:** Saída do R validando  $P(\text{receber o prêmio}) = 1 - e^{-1}$  via cálculo direto, soma truncada e simulação.

## 2.4 Simulação e Comparação Visual: Binomial vs. Poisson

Podemos obter os gráficos a partir da seguinte programação:

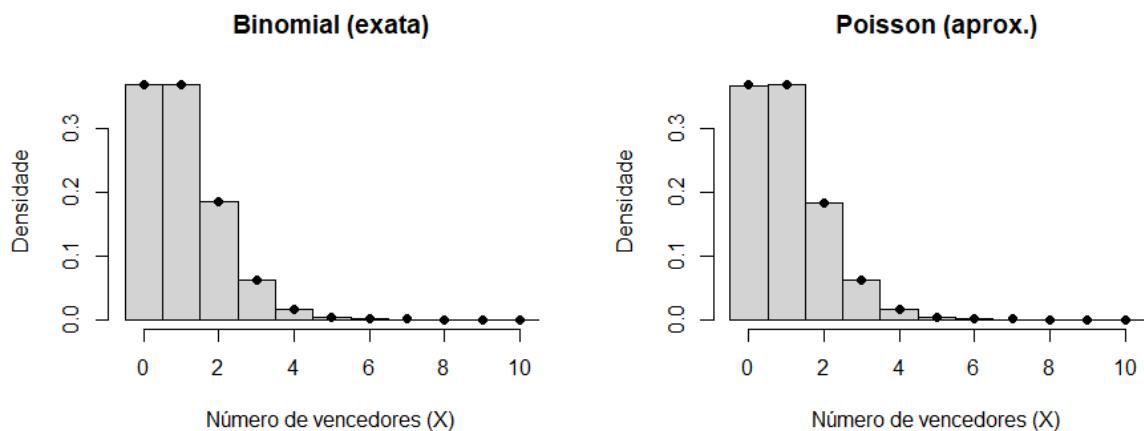
```
1 set.seed(42)
2
3 # Parametros do problema
4 n <- 1e7
5 p <- 1e-7
6 lambda <- n * p # = 1
7
8 # Numero de "dias" simulados
9 B <- 200000 # pode aumentar se quiser (ex.: 1e6)
10
11 # Simulacoes
12 x_bin <- rbinom(B, size = n, prob = p) # modelo exato
13 x_pois <- rpois(B, lambda = lambda) # aproximacao
14
15 # ---- Comparacao visual (histogramas) ----
16 # Para lambda=1, os valores ficam concentrados em poucos inteiros
17 # (0,1,2,3,...)
17 max_k <- max(quantile(c(x_bin, x_pois), 0.999), 10)
18 breaks_int <- seq(-0.5, max_k + 0.5, by = 1)
19
20 par(mfrow = c(1, 2))
21
22 hist(
23   x_bin, breaks = breaks_int, freq = FALSE,
24   main = "Binomial(exata)",
25   xlab = "Número de vencedores(X)",
26   ylab = "Densidade"
27 )
28 # PMF teorica binomial nos inteiros
29 k <- 0:max_k
30 points(k, dbinom(k, size = n, prob = p), pch = 16)
31
32 hist(
33   x_pois, breaks = breaks_int, freq = FALSE,
34   main = "Poisson(aprox.)",
35   xlab = "Número de vencedores(X)",
36   ylab = "Densidade"
37 )
38 # PMF teorica Poisson nos inteiros
39 points(k, dpois(k, lambda = lambda), pch = 16)
40
41 par(mfrow = c(1, 1))
42
43 # ---- Comparacao direta (lado a lado por k) ----
44 # Frequencias empiricas
45 tab_bin <- table(factor(x_bin, levels = k))
46 tab_pois <- table(factor(x_pois, levels = k))
47
48 emp_bin <- as.numeric(tab_bin) / B
49 emp_pois <- as.numeric(tab_pois) / B
```

```

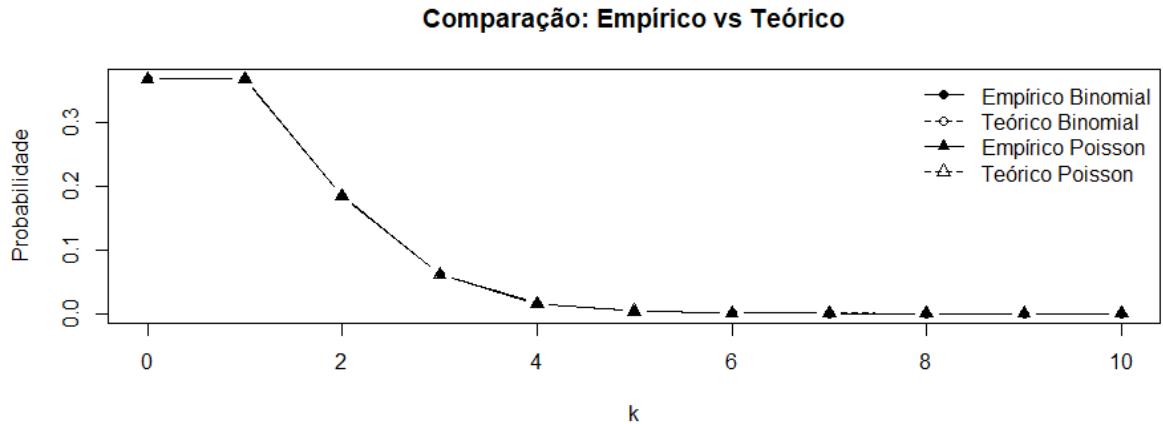
50
51 # Probabilidades teoricas
52 th_bin <- dbinom(k, size = n, prob = p)
53 th_pois <- dpois(k, lambda = lambda)
54
55 # Grafico comparando empirico vs teorico (em um so)
56 plot(
57   k, emp_bin, type = "b", pch = 16,
58   xlab = "k",
59   ylab = "Probabilidade",
60   main = "Comparacao: Empirico vs Teorico"
61 )
62 lines(k, th_bin, type = "b", pch = 1, lty = 2)
63 lines(k, emp_pois, type = "b", pch = 17)
64 lines(k, th_pois, type = "b", pch = 2, lty = 2)
65
66 legend(
67   "topright",
68   legend = c("Empirico Binomial", "Teorico Binomial", "Empirico
69     Poisson", "Teorico Poisson"),
70   pch = c(16, 1, 17, 2),
71   lty = c(1, 2, 1, 2),
72   bty = "n"
73 )

```

E observá-los nos plots:



**Figura 7:** Comparação empírica: histogramas (em densidade) para  $X$  sob Binomial exata e Poisson aproximada ( $\lambda = 1$ ).



**Figura 8:** Comparação por valores inteiros  $k$ : frequências empíricas  $\hat{P}(X = k)$  versus PMFs teóricas para Binomial e Poisson.

Para avaliar empiricamente a qualidade da aproximação de Poisson do Item 1, foram realizadas  $B$  simulações independentes do número diário de vencedores  $X$  sob dois modelos: o modelo exato binomial ( $X \sim \text{Bin}(10^7, 10^{-7})$ ) e o modelo aproximado de Poisson ( $X \sim \text{Pois}(\lambda)$  com  $\lambda = np = 1$ ). A comparação visual foi feita por meio de histogramas em escala de densidade (com classes centradas em inteiros) e pela comparação direta das probabilidades por valor  $k$ , isto é, das frequências relativas empíricas  $\hat{P}(X = k)$  com as probabilidades teóricas  $P(X = k)$ .

As visualizações indicam forte concordância entre os resultados empíricos e as PMFs teóricas, com a massa de probabilidade concentrada em pequenos valores inteiros (principalmente  $k = 0, 1, 2, 3$ ), como esperado quando  $\lambda = 1$ . Além disso, as distribuições geradas pelos dois modelos são praticamente indistinguíveis na região de maior probabilidade, sugerindo que o erro introduzido pela aproximação é desprezível do ponto de vista prático. Pequenas discrepâncias podem ocorrer apenas nas caudas (valores de  $k$  pouco prováveis), o que é compatível com a variabilidade amostral finita. Assim, a simulação confirma que a aproximação por Poisson é adequada para descrever o número diário de vencedores neste regime de eventos raros.

### 3 Monitorando CPU

#### 3.1 Geração de valores normais via Transformação de Box–Muller

Nesta etapa, o objetivo é gerar amostras da temperatura de uma CPU assumindo regime estacionário

$$T \sim \mathcal{N}(\mu = 62, \sigma = 3,5),$$

a partir de variáveis aleatórias uniformes independentes, utilizando a transformação de Box–Muller.

#### Conversão para a distribuição de temperatura da CPU

Para transformar uma normal padrão  $Z \sim \mathcal{N}(0, 1)$  em uma normal com média  $\mu$  e desvio-padrão  $\sigma$ , utiliza-se a transformação linear:

$$T = \mu + \sigma Z.$$

Como  $\mu = 62$  e  $\sigma = 3,5$ , cada valor gerado é convertido por:

$$T = 62 + 3,5 Z.$$

Em particular, para os dois valores do item (b):

$$T_1 = 62 + 3,5 Z_1, \quad T_2 = 62 + 3,5 Z_2,$$

produzindo amostras compatíveis com  $\mathcal{N}(62, 3,5^2)$ .

### Implementação em R

O código abaixo implementa a geração de  $n$  valores normais via Box–Muller e, opcionalmente, já retorna a variável na escala desejada (por padrão,  $\mu = 0$  e  $\sigma = 1$ ).

```

1 box_muller_norm <- function(n, mu = 0, sigma = 1) {
2   m <- ceiling(n / 2)
3
4   u1 <- runif(m)
5   u1[u1 == 0] <- .Machine$double.eps # evita log(0)
6   u2 <- runif(m)
7
8   r <- sqrt(-2 * log(u1))
9   theta <- 2 * pi * u2
10
11  z1 <- r * cos(theta)
12  z2 <- r * sin(theta)
13
14  z <- c(z1, z2)[1:n]
15  mu + sigma * z
16 }
```

## 3.2 Geração de 1.000 medições de temperatura: Box–Muller vs. gerador normal do R

Em seguida, foram gerados 1000 valores utilizando o gerador normal embutido do R (função `rnorm`), e utilizando o Box-Muller criado, com os mesmos parâmetros de média e desvio-padrão, com o objetivo de comparar os dois métodos de geração.

### Implementação em R

O trecho de código a seguir gera as duas amostras, cada uma com 1.000 observações:

```

1 set.seed(42)
2
3 # 1) 1.000 temperaturas via Box-Muller (funcao do item 1)
4 temp_bm <- box_muller_norm(1000, mu = 62, sigma = 3.5)
5
6 # 2) 1.000 temperaturas via gerador normal embutido do R
7 temp_rnorm <- rnorm(1000, mean = 62, sd = 3.5)
```

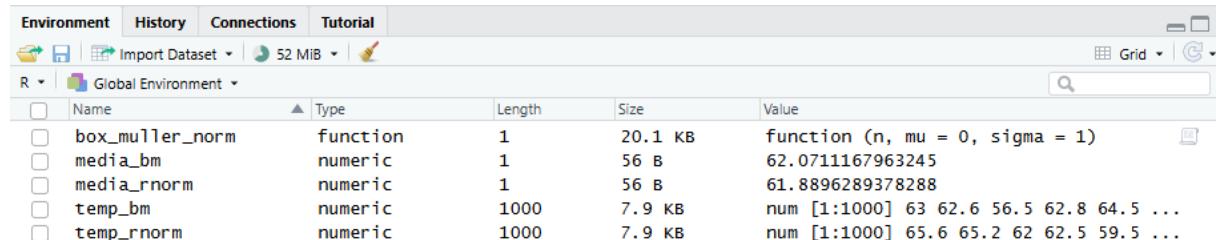
### 3.3 Estatísticas descritivas e probabilidades: duas amostras simuladas

Foram considerados dois conjuntos de dados simulados com  $n = 1000$  observações cada: (i) temperaturas geradas via Box–Muller (`temp_bm`) e (ii) temperaturas geradas pelo gerador normal embutido do R (`temp_rnorm`). Para cada conjunto, foram calculados: média e desvio padrão amostrais.

#### a) Média amostral em R:

```
1 # Media amostral
2 # =====
3 # Box-Muller
4 media_bm <- mean(temp_bm)
5
6 # rnorm
7 media_rnorm <- mean(temp_rnorm)
8
9 media_bm
10 media_rnorm
```

Os valores podem ser observados no "environment" do Rstudio após o cálculo feito no código supracitado:



	Name	Type	Length	Size	Value
□	box_muller_norm	function	1	20.1 KB	function (n, mu = 0, sigma = 1)
□	media_bm	numeric	1	56 B	62.0711167963245
□	media_rnorm	numeric	1	56 B	61.8896289378288
□	temp_bm	numeric	1000	7.9 KB	num [1:1000] 63 62.6 56.5 62.8 64.5 ...
□	temp_rnorm	numeric	1000	7.9 KB	num [1:1000] 65.6 65.2 62 62.5 59.5 ...

**Figura 9:** Saída do R para as médias amostrais: Box–Muller e `rnorm`.

As médias amostrais estimadas foram  $\bar{T}_{BM} \approx 62,07^{\circ}\text{C}$  (Box–Muller) e  $\bar{T}_R \approx 61,89^{\circ}\text{C}$  (`rnorm`). Ambas estão próximas, e a diferença entre elas (cerca de  $0,18^{\circ}\text{C}$ ) é compatível com a variabilidade amostral esperada para  $n = 1000$ , indicando, a priori, que os dois métodos reproduzem adequadamente o comportamento médio da distribuição.

#### b) Desvio-padrão amostral em R:

```
1 # Desvio-padrão amostral
2 # =====
3
4 # Box-Muller
5 sd_bm <- sd(temp_bm)
6
7 # rnorm
8 sd_rnorm <- sd(temp_rnorm)
9
10 sd_bm
11 sd_rnorm
```

Da mesma forma que no item anterior, podemos observar o resultado do cálculos de desvio-padrão no mesmo environment:

The screenshot shows the RStudio interface with the 'Environment' tab selected. The 'Global Environment' pane lists the following variables:

Name	Type	Length	Size	Value
box_muller_norm	function	1	20.1 KB	function (n, mu = 0, sigma = 1) { ... }
media_bm	numeric	1	56 B	62.071167963245
media_rnorm	numeric	1	56 B	61.8896289378288
sd_bm	numeric	1	56 B	3.59863583027187
sd_rnorm	numeric	1	56 B	3.49998930885173
temp_bm	numeric	1000	7.9 KB	num [1:1000] 63 62.6 56.5 62.8 64.5 ...
temp_rnorm	numeric	1000	7.9 KB	num [1:1000] 65.6 65.2 62 62.5 59.5 ...

**Figura 10:** Saída do R para os desvios-padrão amostrais: Box–Muller e `rnorm`.

Os desvios-padrão amostrais obtidos foram  $s_{BM} \approx 3,60^{\circ}\text{C}$  (Box–Muller) e  $s_R \approx 3,50^{\circ}\text{C}$  (`rnorm`). Ambos os valores estão muito próximos, indicando que as duas amostras reproduzem adequadamente a dispersão esperada. A pequena diferença observada é compatível com a variabilidade amostral para  $n = 1000$ .

### c) Mínimos e Máximos amostrais em R:

Utilizando essas linhas de códigos no que já programamos em R, obtemos as temperaturas mínima e máxima das amostras:

```

1 # Temperatura minima e maxima observada
2 # =====
3
4 # Box-Muller
5 min_bm <- min(temp_bm)
6 max_bm <- max(temp_bm)
7
8 # rnorm
9 min_rnorm <- min(temp_rnorm)
10 max_rnorm <- max(temp_rnorm)
11
12 min_bm; max_bm
13 min_rnorm; max_rnorm

```

Então, observamos os resultados:

The screenshot shows the RStudio interface with the 'Environment' tab selected. The 'Global Environment' pane lists the following variables:

Name	Type	Length	Size	Value
box_muller_norm	function	1	20.1 KB	function (n, mu = 0, sigma = 1) { ... }
max_bm	numeric	1	56 B	72.9711252202993
max_rnorm	numeric	1	56 B	74.2335649239496
media_bm	numeric	1	56 B	62.071167963245
media_rnorm	numeric	1	56 B	61.8896289378288
min_bm	numeric	1	56 B	49.6370061445843
min_rnorm	numeric	1	56 B	50.1989132110732
sd_bm	numeric	1	56 B	3.59863583027187
sd_rnorm	numeric	1	56 B	3.49998930885173
temp_bm	numeric	1000	7.9 KB	num [1:1000] 63 62.6 56.5 62.8 64.5 ...
temp_rnorm	numeric	1000	7.9 KB	num [1:1000] 65.6 65.2 62 62.5 59.5 ...

**Figura 11:** Saída do R para mínimos e máximos observados nas amostras: Box–Muller e `rnorm`.

Quanto aos extremos observados, obteve-se  $T_{\min, \text{BM}} \approx 49,64^\circ\text{C}$  e  $T_{\max, \text{BM}} \approx 72,97^\circ\text{C}$  para a amostra gerada por Box–Muller, enquanto para `rnorm` obteve-se  $T_{\min, \text{R}} \approx 50,20^\circ\text{C}$  e  $T_{\max, \text{R}} \approx 74,23^\circ\text{C}$ . As diferenças entre os extremos nas duas amostras decorrem de flutuações aleatórias e não indicam divergência sistemática entre os métodos de geração.

#### d) Probabilidade empírica e teórica $P(T > 68)$ :

Conseguimos calcular a probabilidade empírica a partir do seguinte código:

```

1 # Probabilidade empirica P(T > 68)
2 # =====
3
4 # Box-Muller
5 p_emp_bm_gt68 <- mean(temp_bm > 68)
6
7 # rnorm
8 p_emp_rnorm_gt68 <- mean(temp_rnorm > 68)
9
10 p_emp_bm_gt68
11 p_emp_rnorm_gt68

```

E observá-lo no environment:

	Name	Type	Length	Size	Value
□	box_muller_norm	function	1	20.1 KB	function (n, mu = 0, sigma = 1)
□	max_bm	numeric	1	56 B	72.9711252202993
□	max_rnorm	numeric	1	56 B	74.2335649239496
□	media_bm	numeric	1	56 B	62.0711167963245
□	media_rnorm	numeric	1	56 B	61.8896289378288
□	min_bm	numeric	1	56 B	49.6370061445843
□	min_rnorm	numeric	1	56 B	50.1989132110732
□	p_emp_bm_gt68	numeric	1	56 B	0.052
□	p_emp_rnorm_gt68	numeric	1	56 B	0.034
□	sd_bm	numeric	1	56 B	3.59863583027187
□	sd_rnorm	numeric	1	56 B	3.49998930885173
□	temp_bm	numeric	1000	7.9 KB	num [1:1000] 63 62.6 56.5 62.8 64.5 ...
□	temp_rnorm	numeric	1000	7.9 KB	num [1:1000] 65.6 65.2 62 62.5 59.5 ...

**Figura 12:** Saída do R para as estimativas empíricas de  $P(T > 68)$  nas duas amostras simuladas.

Agora, para obter a  $P(T > 68)$ , assumindo  $T \sim \mathcal{N}(62, 3,5^2)$ , padroniza-se:

$$P(T > 68) = P\left(\frac{T - 62}{3,5} > \frac{68 - 62}{3,5}\right) = P\left(Z > \frac{6}{3,5}\right), \quad Z \sim \mathcal{N}(0, 1).$$

Logo,

$$P(T > 68) = 1 - \Phi\left(\frac{6}{3,5}\right) = 1 - \Phi(1,7143) \approx 0,04324.$$

Nas simulações com  $n = 1000$ , obteve-se  $\hat{P}_{\text{BM}}(T > 68) = 0,052$  (Box–Muller) e  $\hat{P}_{\text{R}}(T > 68) = 0,034$  (`rnorm`). Ambos os valores estão próximos da probabilidade teórica ( $\approx 0,043$ ) e as diferenças observadas são compatíveis com a variabilidade amostral esperada para uma amostra finita.

### e) Probabilidade empírica e teórica $P(60 < T < 65)$ :

Novamente, calculamos a probabilidade empírica a partir do seguinte código:

```

1 # Probabilidade empirica P(60 < T < 65)
2 # =====
3
4 # Box-Muller
5 p_emp_bm_60_65 <- mean(temp_bm > 60 & temp_bm < 65)
6
7 # rnorm
8 p_emp_rnorm_60_65 <- mean(temp_rnorm > 60 & temp_rnorm < 65)
9
10 p_emp_bm_60_65
11 p_emp_rnorm_60_65

```

E observá-lo no environment:

	Name	Type	Length	Size	Value
□	box_muller_norm	function	1	20.1 KB	function (n, mu = 0, sigma = 1)
□	max_bm	numeric	1	56 B	72.9711252202993
□	max_rnorm	numeric	1	56 B	74.2335649239496
□	media_bm	numeric	1	56 B	62.0711167963245
□	media_rnorm	numeric	1	56 B	61.8896289378288
□	min_bm	numeric	1	56 B	49.6370061445843
□	min_rnorm	numeric	1	56 B	50.1989132110732
□	p_emp_bm_60_65	numeric	1	56 B	0.513
□	p_emp_bm_gt68	numeric	1	56 B	0.052
□	p_emp_rnorm_60_65	numeric	1	56 B	0.52
□	p_emp_rnorm_gt68	numeric	1	56 B	0.034
□	sd_bm	numeric	1	56 B	3.59863583027187
□	sd_rnorm	numeric	1	56 B	3.49998930885173
□	temp_bm	numeric	1000	7.9 KB	num [1:1000] 63 62.6 56.5 62.8 64.5 ...
□	temp_rnorm	numeric	1000	7.9 KB	num [1:1000] 65.6 65.2 62 62.5 59.5 ...

**Figura 13:** Saída do R para as estimativas empíricas de  $P(60 < T < 65)$  nas duas amostras simuladas.

Assumindo  $T \sim \mathcal{N}(62, 3,5^2)$ , temos:

$$P(60 < T < 65) = P\left(\frac{60 - 62}{3,5} < \frac{T - 62}{3,5} < \frac{65 - 62}{3,5}\right) = P\left(\frac{-2}{3,5} < Z < \frac{3}{3,5}\right), \quad Z \sim \mathcal{N}(0, 1).$$

Logo,

$$P(60 < T < 65) = \Phi\left(\frac{3}{3,5}\right) - \Phi\left(\frac{-2}{3,5}\right) = \Phi(0,8571) - \Phi(-0,5714).$$

Usando valores tabelados (ou computacionais) da CDF normal padrão,

$$\Phi(0,8571) \approx 0,80427, \quad \Phi(-0,5714) \approx 0,28385,$$

de modo que

$$P(60 < T < 65) \approx 0,80427 - 0,28385 = 0,52042 \approx 0,520.$$

Nas simulações com  $n = 1000$ , obteve-se  $\hat{P}_{\text{BM}}(60 < T < 65) = 0,513$  (Box-Muller) e  $\hat{P}_{\text{R}}(60 < T < 65) = 0,520$  (rnorm). Ambos os valores são consistentes com a probabilidade teórica ( $\approx 0,520$ ), com diferenças pequenas atribuíveis à variabilidade amostral.

## f) Probabilidade teórica $P(T > 75)$ e explicações:

Assumindo  $T \sim \mathcal{N}(62, 3,5^2)$ , calcula-se a probabilidade teórica de exceder  $75^\circ\text{C}$  por padronização. Define-se

$$Z = \frac{T - \mu}{\sigma} = \frac{T - 62}{3,5} \sim \mathcal{N}(0, 1).$$

Logo,

$$P(T > 75) = P\left(\frac{T - 62}{3,5} > \frac{75 - 62}{3,5}\right) = P\left(Z > \frac{13}{3,5}\right).$$

Calculando o valor do *z-score*:

$$\frac{13}{3,5} = 3,7142857 \approx 3,7143.$$

Assim,

$$P(T > 75) = 1 - \Phi(3,7143).$$

Usando tabela/CDF computacional da normal padrão:

$$\Phi(3,7143) \approx 0,9998981 \implies P(T > 75) \approx 1 - 0,9998981 = 0,0001019 \approx 1,02 \times 10^{-4}.$$

## Por que é difícil observar esse evento em $n = 1000$ amostras?

Considere o indicador  $I_i = \mathbf{1}\{T_i > 75\}$  para a  $i$ -ésima medição. Sob independência, tem-se

$$I_i \sim \text{Bernoulli}(p), \quad \text{com } p = P(T > 75) \approx 1,02 \times 10^{-4}.$$

O número total de ocorrências acima de  $75^\circ\text{C}$  em  $n$  medições é

$$N = \sum_{i=1}^n I_i \sim \text{Bin}(n, p).$$

Portanto, o número *esperado* de ocorrências é

$$\mathbb{E}[N] = np \approx 1000 \times 1,02 \times 10^{-4} \approx 0,102.$$

Como  $\mathbb{E}[N] \ll 1$ , a expectativa é observar *bem menos que uma ocorrência* em 1000 medições, isto é, na maioria das amostras não aparecerá nenhum valor acima de  $75^\circ\text{C}$ .

**Probabilidade de aparecer pelo menos uma ocorrência.** A probabilidade de não ocorrer nenhuma vez (isto é,  $N = 0$ ) é:

$$P(N = 0) = (1 - p)^n.$$

Com  $n = 1000$  e  $p \approx 0,0001019$ :

$$P(N = 0) = (1 - 0,0001019)^{1000}.$$

Uma aproximação padrão para  $p$  pequeno usa  $\ln(1 - p) \approx -p$ , resultando em

$$(1 - p)^{1000} = \exp(-1000p) \approx \exp(-1000p) = e^{-0,102} \approx 0,903.$$

Assim, a probabilidade de observar *pelo menos uma* medição acima de  $75^\circ\text{C}$  é

$$P(N \geq 1) = 1 - P(N = 0) \approx 1 - 0,903 = 0,097,$$

mostrando que há apenas cerca de 9,7% de chance de ver esse evento em 1000 amostras.

**Tamanho amostral necessário para observar o evento com maior frequência.**  
 Se o objetivo for ter, por exemplo, 50% de chance de observar ao menos uma ocorrência, impõe-se

$$P(N \geq 1) = 1 - (1 - p)^n \geq 0,5 \iff (1 - p)^n \leq 0,5.$$

Usando novamente a aproximação  $(1 - p)^n \approx e^{-np}$ :

$$e^{-np} \leq 0,5 \iff np \geq \ln 2 \iff n \geq \frac{\ln 2}{p}.$$

Com  $p \approx 1,02 \times 10^{-4}$ :

$$n \frac{0,6931}{1,02 \times 10^{-4}} \approx 6,8 \times 10^3.$$

Logo, seriam necessárias em torno de 6,800 medições para que a chance de observar ao menos um valor acima de 75°C seja aproximadamente 50%, o que reforça que se trata de um evento raro e que amostras pequenas (como  $n = 1000$ ) tendem a não capturá-lo.

### 3.4 Visualizando resultados

#### a) Um histograma das temperaturas simuladas da CPU:

Podemos obter os histogramas a partir da seguinte programação:

```

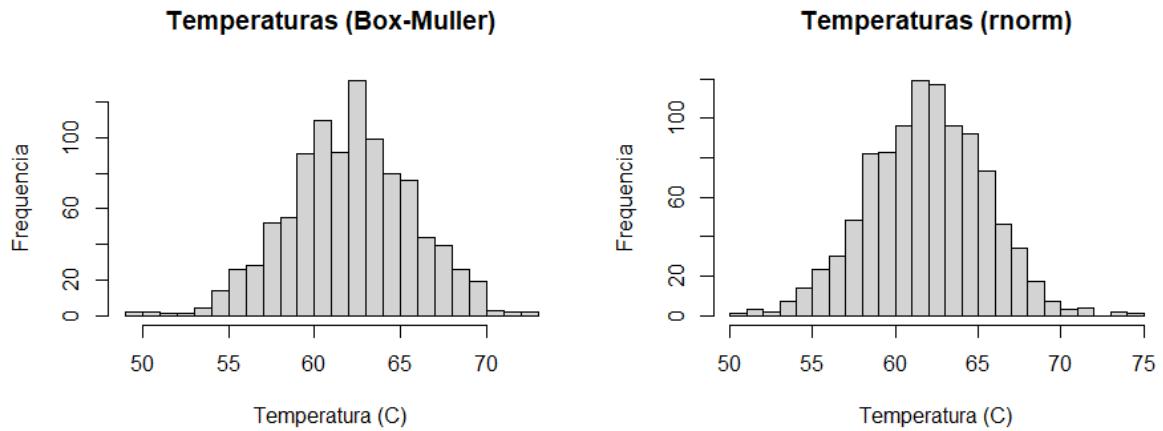
1 # Histograma das temperaturas simuladas (separados)
2 # =====
3
4 par(mfrow = c(1, 2))
5
6 # Box-Muller
7 hist(
8   temp_bm,
9   breaks = 30,
10  main = "Temperaturas_(Box-Muller)",
11  xlab = "Temperatura_(C)",
12  ylab = "Frequencia"
13 )
14
15 # rnorm
16 hist(
17   temp_rnorm,
18   breaks = 30,
19   main = "Temperaturas_(rnorm)",
20   xlab = "Temperatura_(C)",
21   ylab = "Frequencia"
22 )
23
24 par(mfrow = c(1, 1))

```

E observá-los nos plots:

#### b) A função densidade de probabilidade (PDF) normal teórica (média 62 °C, desvio padrão 3,5 °C) sobreposta ao histograma

Denovo, podemos obter as PDF's a partir da seguinte programação:



**Figura 14:** Histogramas das temperaturas simuladas: Box–Muller (à esquerda) e `rnorm` (à direita).

```

1 # Questao 3 - Item 4 (b): PDF normal teorica sobreposta ao
2 # =====
3
4 # Parametros teoricos
5 mu <- 62
6 sigma <- 3.5
7
8 # Box-Muller: hist + PDF
9 hist(
10   temp_bm,
11   breaks = 30,
12   freq = FALSE, # densidade
13   main = "Box-Muller: histograma+PDF teorica",
14   xlab = "Temperatura(C)",
15   ylab = "Densidade"
16 )
17 curve(
18   dnorm(x, mean = mu, sd = sigma),
19   add = TRUE,
20   lwd = 2
21 )
22
23 # rnorm: hist + PDF
24 hist(
25   temp_rnorm,
26   breaks = 30,
27   freq = FALSE, # densidade
28   main = "rnorm: histograma+PDF teorica",
29   xlab = "Temperatura(C)",
30   ylab = "Densidade"
31 )
32 curve(
33   dnorm(x, mean = mu, sd = sigma),
34   add = TRUE,

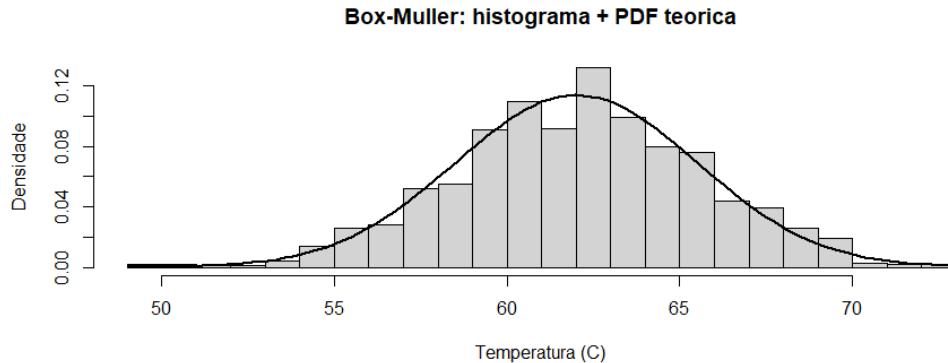
```

```

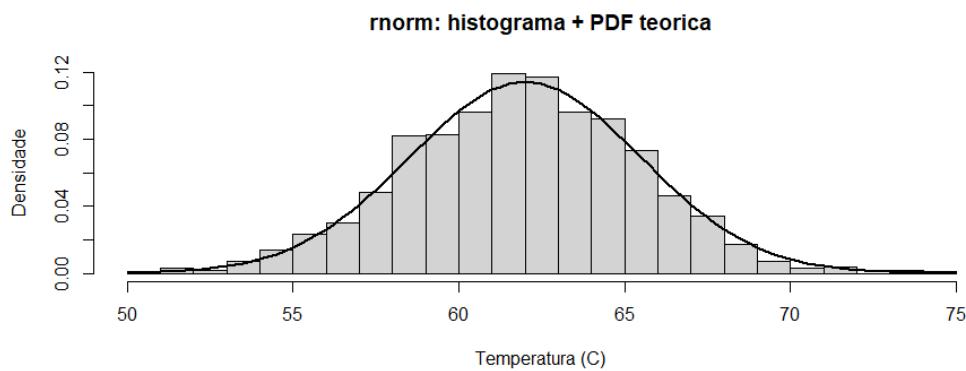
35 lwd = 2
36 )

```

E observá-las nos plots:



**Figura 15:** Box–Muller: histograma em densidade com PDF teórica de  $\mathcal{N}(62, 3,5^2)$  sobreposta.



**Figura 16:** `rnorm`: histograma em densidade com PDF teórica de  $\mathcal{N}(62, 3,5^2)$  sobreposta.

### 3.5 Discussão de resultados:

As distribuições empíricas das temperaturas simuladas apresentam formato aproximadamente simétrico e unimodal, e os histogramas obtidos (com a PDF teórica sobreposta) indicam boa aderência ao comportamento esperado de uma distribuição normal  $\mathcal{N}(62, 3,5^2)$ . A curva teórica se ajusta de forma consistente ao centro e às caudas dos histogramas, com pequenas discrepâncias locais atribuíveis ao tamanho amostral finito ( $n = 1000$ ) e à discretização inerente ao particionamento em classes.

Em relação às medidas de tendência central e dispersão, as médias amostrais estimadas foram próximas do valor esperado  $\mu = 62^\circ\text{C}$ :  $\bar{T}_{\text{BM}} \approx 62,07^\circ\text{C}$  (Box–Muller) e  $\bar{T}_{\text{R}} \approx 61,89^\circ\text{C}$  (`rnorm`). De forma análoga, os desvios-padrão amostrais ficaram próximos de  $\sigma = 3,5^\circ\text{C}$ :  $s_{\text{BM}} \approx 3,60^\circ\text{C}$  e  $s_{\text{R}} \approx 3,50^\circ\text{C}$ . As diferenças observadas entre valores empíricos e teóricos, bem como entre os dois métodos de geração, são compatíveis com flutuações aleatórias esperadas em amostras de tamanho 1000 e não sugerem viés sistemático.

Não foram identificadas diferenças perceptíveis relevantes entre o conjunto gerado via Box–Muller e o conjunto gerado via `rnorm`. Ambos produzem amostras que reproduzem adequadamente o comportamento médio e a variabilidade da distribuição-alvo. Quando

ocorrem pequenas variações (por exemplo, nos extremos ou em probabilidades empíricas), elas são explicadas pela variabilidade amostral: com  $n$  fixo, duas amostras independentes de uma mesma distribuição não são idênticas e podem apresentar diferenças pontuais.

Do ponto de vista aplicado, simulações desse tipo são úteis para avaliar estratégias de resfriamento e escalonamento dinâmico de clock, pois permitem modelar cenários de carga e ruído térmico e, a partir disso, estimar a frequência de ultrapassagem de limites operacionais (por exemplo,  $P(T > 68)$  ou eventos raros como  $P(T > 75)$ ). Mesmo quando os dados reais ainda não estão disponíveis, a simulação fornece uma referência probabilística para definir limiares, dimensionar margens de segurança e comparar políticas de controle térmico sob hipóteses plausíveis.

Por fim, geradores de números aleatórios uniformes são a base de muitos sistemas de RNG em computação porque são relativamente simples de implementar e analisar, permitem gerar outras distribuições por meio de transformações matemáticas, como o método de Box–Muller (normal) ou métodos de inversão e aceitação-rejeição, e viabilizam a reproduzibilidade dos experimentos via *seeds*. Assim, a uniformidade atua como um bloco fundamental a partir do qual se constroem simulações e algoritmos probabilísticos mais complexos de forma controlada.