



HOMEWORK II

NOME COMPLETO: LUCAS DE OLIVEIRA SOBRAL

NUMERO DE MATRICULA: 556944

NOME COMPLETO: ÁLVARO JOSÉ PASSOS DE FREITAS NETO

NUMERO DE MATRICULA: 567593

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.

1. Determine a função de distribuição de X .
2. Construa os gráficos da função massa de probabilidade (PMF) e da função distribuição acumulada (CDF) de X .
3. Calcule o valor esperado, a variância e o desvio padrão de X .
4. Calcule a probabilidade de:
 - (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?
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 ?

SOLUÇÃO DA QUESTÃO 1

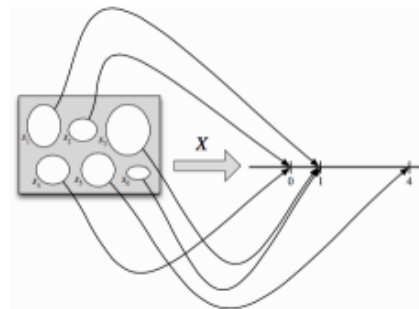
Descrição da atividade

A atividade tem por objetivo abordar o conteúdo de variáveis aleatórias discretas, se destrinchando principalmente na distribuição Binomial, cálculo da PMF e CDF e calcular medidas importantes como valor esperado, variância e desvio padrão. O conteúdo também envolve interpretação de probabilidades em cenários reais e análise de como parâmetros como p (probabilidade de sucesso) e n (tamanho da amostra) influenciam a forma da distribuição, tornando possível tomar decisões mais informadas em contextos de demanda e planejamento.

Variáveis aleatórias:

O que é uma variável aleatória (X)?

- o É uma função que atribui um valor numérico a cada resultado de um experimento aleatório $X(s)$.



Tipos de variáveis aleatórias

- o Variáveis aleatórias discretas: É uma variável aleatória que pode assumir um número contável de valores, seja ele finito ou infinito de valores e explicaremos melhor nessa questão.
- o Variáveis aleatórias contínuas: São variáveis que podem assumir qualquer número dentro de um intervalo contínuo e falaremos melhor depois.

Valor médio esperado $E(X)$

- o Uma das maiores diferenças entre as variáveis aleatórias da probabilidade e variáveis da estatística descritiva se dá pela incapacidade de afirmar com 100% de certeza os valores da amostragem em um experimento, por isso temos uma constante determinística do modelo, chamado de "esperança" média de valores. Entre variáveis discretas e contínuas há uma diferença sutil no cálculo da esperança, que será visto mais à frente.

Variância de uma variável aleatória $Var(\sigma^2$ ou S^2)

- É o valor de o quão afastados os dados estão da média esperada $E(X)$, se a variância tiver um baixo valor então os dados estão concentrados no ponto médio, se estiver com valor alto indica que os dados estão mais espalhados.
- Para variáveis discretas e contínuas a fórmula formal é muito complicada de calcular por isso usamos uma técnica inconsciente, chamada de LOTUS para simplificar os cálculos, apresentada a seguir:

LOTUS Lei do estatístico inconsciente

- É a regra geral para calcular a esperança média de qualquer função de uma variável aleatória, sem precisar descobrir a distribuição dessa nova função, ela diz que aplicar os valores x_i na nova função $g(x)$ e multiplicar pela probabilidade de x_i .
- Fórmula:

$$E[g(x)] = \sum g(x) * P(X = x_i)$$

Distribuição de uma variável aleatória

- É uma forma de descrição matemática que resulta em uma visualização de como os valores estão se distribuindo em um gráfico que a soma de suas probabilidades resulta em 1 e não podem ter probabilidade negativa. Suas principais funções são:
- Função massa de probabilidade FMP ou (PMF) para discretas
- Função densidade de probabilidade FDP ou (PDF) para contínuas
- Função de distribuição acumulada FDA ou (CDF) para discretas e contínuas

Distribuições especiais

- São modelos matemáticos para situações que acontecem na vida real, algumas delas são:

Tipo	Distribuição	Notação	Parâmetros
Discretas	Bernoulli	$X \sim Bern(p)$	p (prob. sucesso)
	Binomial	$X \sim Bin(n, p)$	n (tentativas), p
	Geométrica	$X \sim Geom(p)$	p (prob. sucesso)
	Poisson	$X \sim Pois(\lambda)$	λ (taxa média)
Contínuas	Uniforme	$X \sim U(a, b)$	a (min), b (max)
	Normal	$X \sim N(\mu, \sigma^2)$	μ (média), σ^2 (variância)
	Exponencial	$X \sim Exp(\lambda)$	λ (taxa)

Tabela 1: Principais Distribuições Especiais

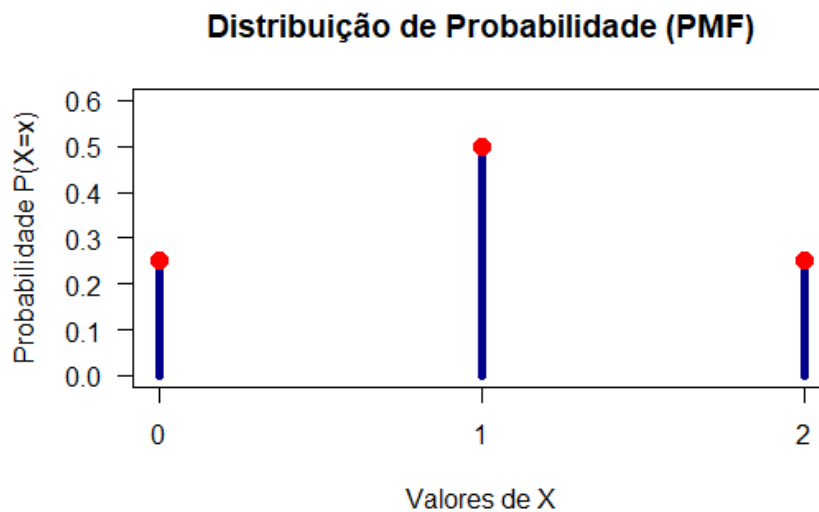
Variáveis aleatórias discretas:

O que são?

- É a variável aleatória que pode assumir um número contável de valores, seja ele finito ou infinito.

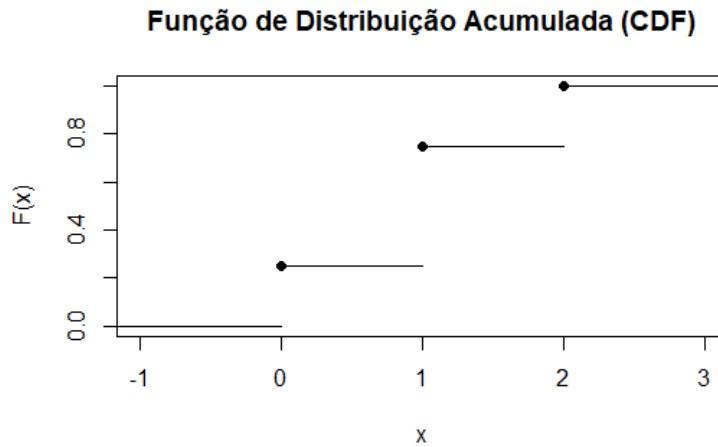
Função massa de probabilidade FMP ou (PMF)

- É a função que nos dá a probabilidade exata de a variável assumir um valor específico, exemplo de PMF com $Rx = (0, 1, 2)$ e $Px = (1/4, 1/2, 1/4)$.



Função de distribuição acumulada discreta FDA ou (CDF)

- Se trata da função que soma todas as probabilidades de 0 até 1, muito parecida com a frequência Acumulada (N_i), ela serve principalmente para termos o valor da probabilidade em um intervalo de valores ou a probabilidade total até o valor X_k , por exemplo: $F_X(x) = P(X \leq x)$.
- Gráfico é em forma de degraus pois ela permanece constante (plana) até atingir um valor possível da variável X_k , onde ela dá um salto.
- CDF do exemplo anterior:



Valor médio esperado $E(X)$ de uma variável discreta

- Para variáveis discretas o cálculo é feito pela média ponderada das amostradas x_i por suas probabilidades $P_X(x_i)$ ou $P(X = x_i)$.
- Fórmula:

$$E(X) = \sum_i x_i P(X = x_i)$$

Variância de uma variável aleatória discreta $Var(X)$

- A variância de uma variável aleatória discreta possui a seguinte fórmula formal:

$$Var(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 \cdot P(X = x)$$

- Mas utilizando a técnica de LOTUS podemos obter:

$$Var(X) = E(X^2) - [E(X)]^2$$

$$Var(X) = \left[\sum_i (x_i)^2 \cdot P(X = x_i) \right] - [E(x)]^2$$

Distribuições especiais das variáveis aleatórias discretas:

Distribuição de Bernoulli [$X \sim Bern(p)$]:

- Modela um único experimento com dois resultados possíveis, sucesso (1) ou fracasso (0).

- PMF:

$$P(X = x) = \begin{cases} p & \text{se } x = 1 \text{ (sucesso)} \\ 1 - p & \text{se } x = 0 \text{ (fracasso)} \\ 0 & \text{caso contrário} \end{cases}$$

- Esperança:

É exatamente a média ponderada do único valor de sucesso $E(X) = 1 \cdot p = p$

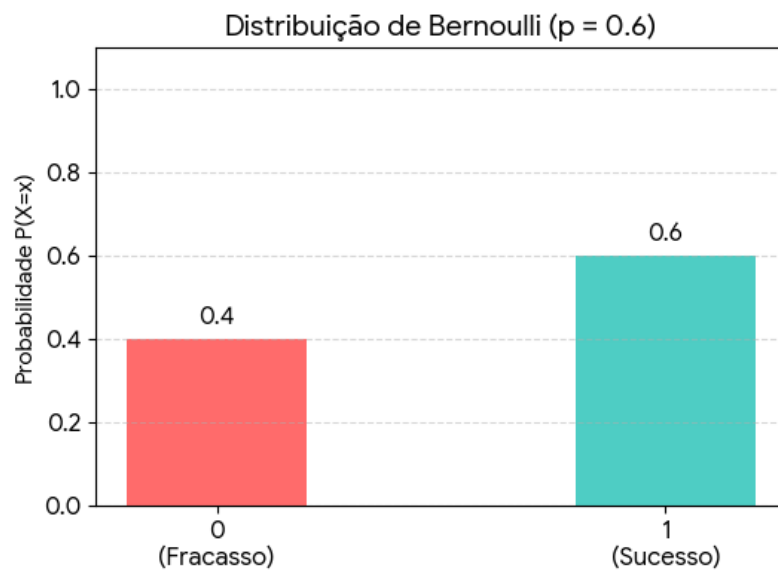
$$E(X) = p$$

- Variância:

Varia no máximo 0.5 se as chances de fracasso e sucesso forem iguais e mínimo se a probabilidade estiver perto de 0 ou 1.

$$Var(X) = p(1 - p)$$

- Gráfico PMF para exemplo Bern(0.6):



Distribuição Binomial [$X \sim \text{Bin}(n, p)$]

- É o modelo que repete o teste de Bernoulli n vezes, para obter a chance de se obter exatamente k sucessos.

- PMF:

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

para $k \in \{0, 1, 2, \dots, n\}$

- o Entendendo a PMF:
 - $\binom{n}{k}$ ou $\frac{n!}{(n-k)! \cdot k!}$: Fator combinatório. Conta de quantas formas podemos organizar os k sucessos nas n posições (ex: para $n = 3, k = 2$, temos SSF, SFS, FSS).
 - p^k : Probabilidade de ocorrerem exatamente k sucessos.
 - $(1 - p)^{n-k}$: Probabilidade dos $n - k$ fracassos restantes ocorrerem.

- o Esperança:

É dado pelo produto do número de amostras pela probabilidade de sucesso destas.

$$E(X) = n \cdot p$$

- o Variância:

É o produto de amostras, sucessos e fracassos, sendo este a dispersão dos dados.

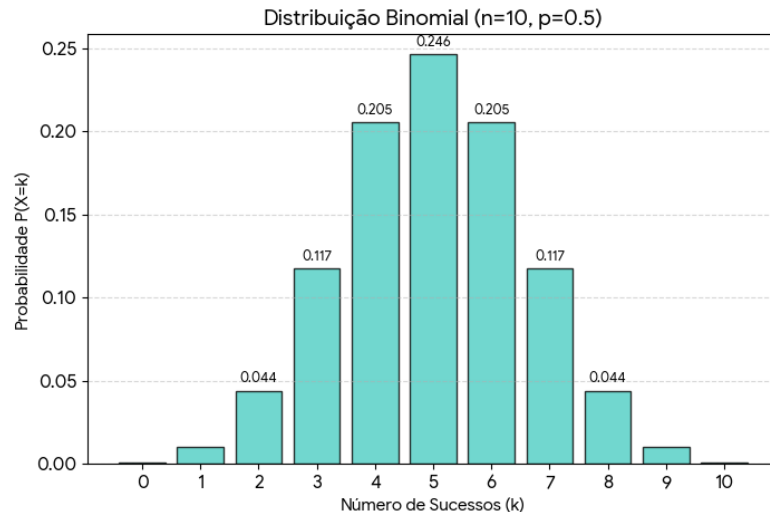
$$Var(X) = n \cdot p \cdot (1 - p)$$

- o Desvio-padrão:

Raiz quadrada da variância.

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

- o Gráfico PMF para 10 amostras com 50% de sucesso:



Respostas dos itens da questão 1:

1.1 Resposta:

A função X tem 2 resultados possíveis, pedir ou não sobremesa, classificando como distribuição binomial, pois além de ter 2 resultados possíveis ele faz $n > 1$ tentativas, denotando:

$$X \sim Bin(50, 0.7)$$

1.2 Resposta:

Para construir esses gráficos, utilizaremos as funções nativas do R. Note que para a PMF usamos em variáveis discretas o spike plot e para a CDF o gráfico de degraus.

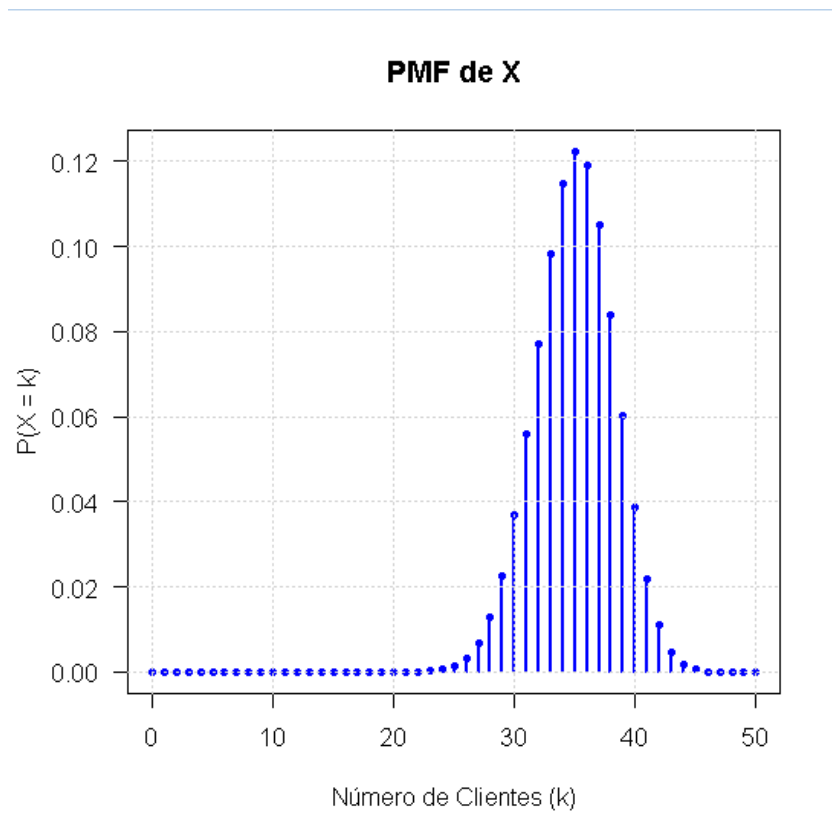
Segue o código básico usado na questão:

```
# Definindo os parametros
n <- 50
p <- 0.7
k <- 0:n

# Calculando as distribuicoes
pmf <- dbinom(k, size = n, prob = p)
cdf <- pbinom(k, size = n, prob = p)
```

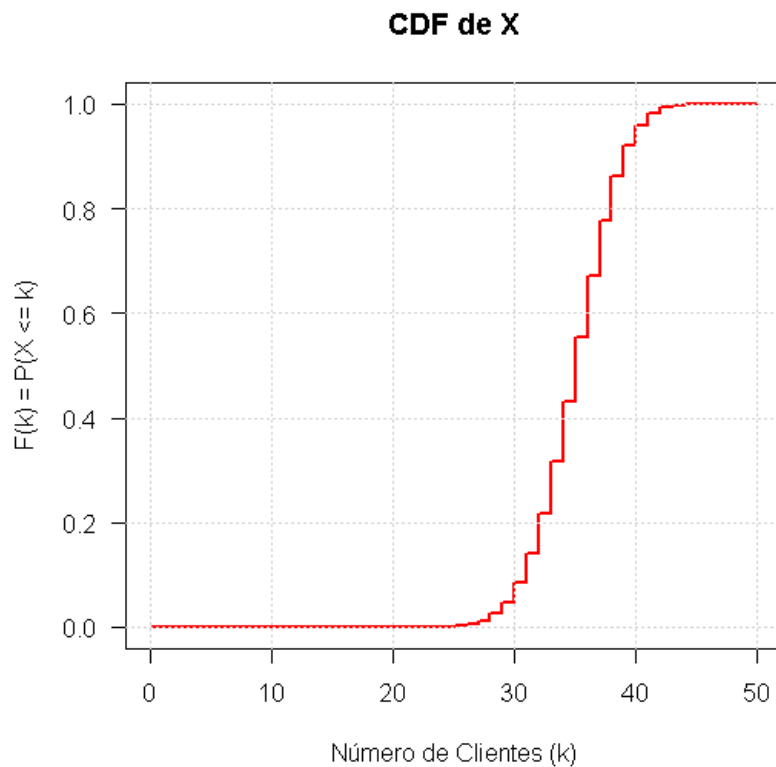
O código para gerar a PMF e o gráfico gerado:

```
# Grafico da Funcao Massa de Probabilidade (PMF)
plot(k, pmf, type = "h", lwd = 2, col = "blue",
     main = "PMF de X",
     xlab = "Numero de Clientes (k)",
     ylab = "P(X=k)",
     las = 1)
points(k, pmf, pch = 16, col = "blue", cex = 0.7)
grid()
```

O código para gerar a CDF e o gráfico gerado:

```
# Grafico da Funcao Distribuicao Acumulada (CDF)
plot(k, cdf, type = "s", lwd = 2, col = "red",
     main = "CDF de X",
     xlab = "Numero de Clientes (k)",
     ylab = "F(k) = P(X ≤ k)",
     las = 1)
grid()
```



No contexto do restaurante, a variável aleatória X representa o número de clientes, entre 50 atendidos, que pedem sobremesa, considerando uma probabilidade de 70% para cada cliente. O gráfico da PMF mostra que os valores mais prováveis de X concentram-se em torno da média $E[X]=35$, refletindo o comportamento esperado de uma distribuição binomial. Os picos discretos evidenciam a natureza discreta do número de pedidos.

O gráfico da CDF apresenta um crescimento em degraus, característico de variáveis discretas, indicando a probabilidade acumulada de até k clientes pedirem sobremesa. A probabilidade acumulada aumenta mais rapidamente próxima à média e se aproxima de 1 à medida que k se aproxima de 50.

1.3 Resposta:

Valor Esperado:

$$E(X) = n \cdot p = 50 \cdot 0.7 = 35$$

Variância:

$$Var(X) = n \cdot p \cdot (1 - p) = 50 \cdot 0.7 \cdot (1 - 0.7) = 10.5$$

Desvio padrão:

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{10.5} \approx 3.24$$

1.4 Resposta:

A função de massa de probabilidade (PMF) que descreve esse modelo é definida pela expressão:

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

De acordo com nossa questão temos os seguintes valores:

- $n = 50$
- $p = 0.7$

Usaremos essa formula para calcular as probabilidades dos itens

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

Note que, para calcular essa probabilidade, precisamos calcular a soma das probabilidades individuais entre 20 e 50. Portanto teremos o k variando de 20 a 50 logo nossa formula fica:

$$P(X \geq 20) = \sum_{k=20}^{50} \binom{50}{k} \cdot (0,7)^k \cdot (0,3)^{50-k} \quad (2)$$

Uma vez que minha calculadora não aguentou fazer o somatório, eu recorri a usar a calculadora online do site desmos.com para resolver a fórmula, descobrindo que:

$$P(X \geq 20) = 0.999997 \approx 1 \quad (3)$$

O valor obtido para $P(X \geq 20)$ é praticamente igual a 1, o que indica que é quase certo que, em um grupo de 50 clientes, pelo menos 20 peçam sobremesa. Esse resultado é coerente com o contexto do problema, já que a probabilidade individual de um cliente pedir sobremesa é alta ($p = 0.7$), fazendo com que valores baixos de X sejam extremamente improváveis. Assim, observar menos de 20 clientes pedindo sobremesa seria um evento raro nesse cenário.

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

Do mesmo modo, para calcular essa probabilidade, precisamos calcular o somatório das probabilidades individuais entre 30 e 43, logo, com isso definido, nosso k varia entre 31 e 42, portanto nossa fórmula fica:

$$P(30 < X < 43) = \sum_{k=31}^{42} \binom{50}{k} \cdot (0,7)^k \cdot (0,3)^{50-k} \quad (4)$$

Usando novamente a calculadora online do site desmos.com para resolver a fórmula, descobri que:

$$P(30 < X < 43) = 0.9079 \quad (5)$$

A probabilidade $P(30 < X < 43) = 0,9079$ indica que há uma alta chance de que o número de clientes que pedem sobremesa esteja entre 31 e 42. Esse intervalo inclui a média da distribuição ($E[X] = 35$), o que explica a grande concentração de probabilidade. O resultado mostra que, na maioria das amostras de 50 clientes, o número de pedidos de sobremesa tende a se manter próximo do valor esperado.

c) $P(X = 31)$:

Por fim, para calcular essa probabilidade, precisamos calcular a probabilidade para quando $k = 31$, portanto nossa fórmula fica:

$$P(X = 31) = \binom{50}{31} \cdot 0,7^{31} \cdot 0,3^{19} \quad (6)$$

Usando novamente a calculadora online do site desmos.com para resolver a fórmula, descobri que:

$$P(X = 31) = 0.0558 \quad (7)$$

A probabilidade $P(X = 31) = 0,0558$ representa a chance de exatamente 31 clientes, entre os 50, pedirem sobremesa. Embora esse seja um valor relativamente próximo da média, trata-se de um único valor específico, o que explica a probabilidade menor quando comparada a intervalos de valores. Ainda assim, o resultado indica que 31 pedidos de sobremesa constituem um cenário plausível dentro do comportamento esperado da distribuição.

1.5 Resposta:

Analisando o valor médio de 35 e desvio padrão 3.24, podemos inferir que o estoque de sobremesas deveria ficar sempre com no máximo a média mais 2 vezes o desvio padrão, sendo 41,48 ou seja 42 sobremesas, garantindo atingir a maioria dos casos pois considerando que o restaurante é muito frequentado dificilmente teria um desperdício alto, usamos o desvio padrão pois a variável sobremesa não está na mesma medida que a variância, ou seja, ela não é quadrática.

1.6 Resposta:

A alteração da probabilidade de sucesso (p) impacta diretamente na posição central e na simetria da distribuição. Por exemplo, se a sobremesa se tornasse mais popular, passando de $p = 0.7$ para $p = 0.8$, o valor esperado de clientes que solicitam o item aumentaria proporcionalmente, deslocando o pico da curva para a direita. A distribuição aumentaria a probabilidade de ocorrência de valores altos, tornando quase nula a chance de observar poucos pedidos. Matematicamente, isso alteraria a média, o que, nesse caso, elevaria a média de 35 para 40 clientes.

Já a variação no número de clientes (n) alteraria a escala e a precisão da distribuição. Ao aumentar o tamanho da amostra, a distribuição binomial tende a se comportar de forma mais suave, aproximando-se de uma Distribuição Normal, de acordo com o Teorema Central do Limite. Embora o desvio padrão absoluto aumente, a variabilidade em relação ao total de clientes diminui, o que significa que os resultados observados tendem a se agrupar de forma mais previsível.

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.

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 e p .
2. Calcule o valor esperado, $E[X]$, e a variância $Var(X)$, usando tanto a distribuição exata quanto a aproximada. Comente sobre a semelhança entre os resultados.
3. Suponha que você ganhe a recompensa, mas que possa haver outros vencedores. Seja $W \sim Pois(1)$ 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.
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 2.1. Descreva brevemente o que a visualização indica sobre a qualidade da aproximação.

SOLUÇÃO DA QUESTÃO 2

Descrição da atividade

Esta atividade tem por objetivo explorar variáveis aleatórias discretas em cenários de probabilidade muito pequena e número de tentativas extremamente grande, permitindo

discutir quando a distribuição Binomial pode ser adequadamente aproximada pela distribuição Poisson. O conteúdo envolve justificar matematicamente essa aproximação, comparar valores esperados e variâncias entre o modelo exato e o aproximado e interpretar a distribuição em situações reais, como sorteios online. Além disso, inclui o uso de simulações para avaliar empiricamente a qualidade da aproximação e compreender como eventos raros se comportam em larga escala, reforçando a importância de modelos probabilísticos na tomada de decisões e na análise de sistemas de grande volume.

Mais distribuições especiais discretas

Distribuição de Poisson [$X \sim Pois(\lambda)$]

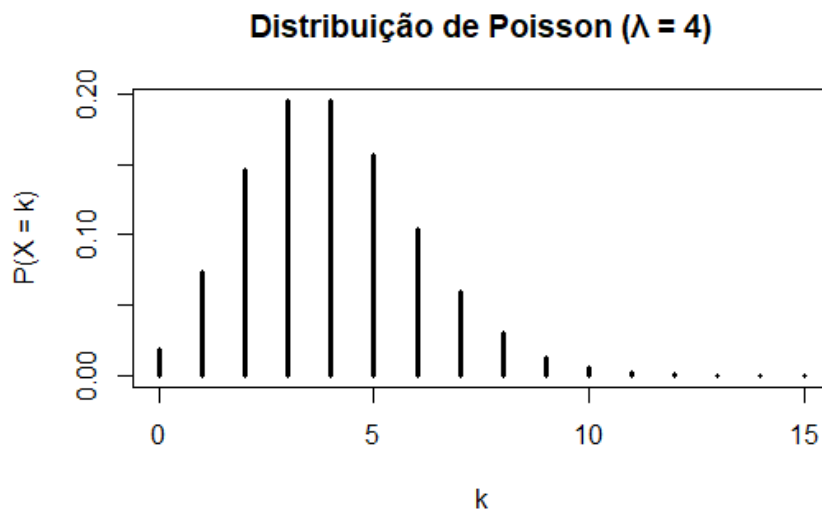
- É o número de eventos que ocorrem em um intervalo fixo de tempo ou espaço, dado que esses eventos ocorrem com uma taxa média constante e independentemente do tempo desde o último evento.
- PMF:

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Para $k \in 0, 1, 2, 3, \dots$

$\lambda(lambda)$ = taxa média de ocorrência no intervalo.

- Esperança = Variância = λ São proporcionais
- Gráfico:



Respostas dos itens da questão 2:

2.1 Resposta:

Podemos usar a distribuição Binomial como uma aproximação aproximada do problema, pois descobriremos a quantidade de vencedores com mesma probabilidade em um único cenário X simulando 1 dia, aplicando:

PMF da $Bin(10^7, 10^{-7})$:

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

para $k \in \{0, 1, 2, \dots, n\}$

$$P(X = k) = \binom{10^7}{k} \cdot (10^{-7})^k \cdot (1 - 10^{-7})^{10^7-k}$$

Expandindo a fórmula:

$$P(X = k) = \frac{10^7!}{(10^7 - k)!} \cdot (10^{-7})^k \cdot (1 - 10^{-7})^{10^7-k}$$

Percebe-se que o grande número de elementos e com probabilidade muito baixa para cada um deles torna a distribuição Binomial impraticável e com a informação que a esperança $E(X) = 10^7 \cdot 10^{-7} = 1$ podemos usá-la como taxa de frequência λ a ser usada mais adequadamente na distribuição de Poisson.

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Para $k \in 0, 1, 2, 3, \dots$

$$P(X = k) = \frac{e^{-1} \cdot 1^k}{k!}$$

Simplificando:

$$P(X = k) = \frac{e^{-1}}{k!}$$

Agora temos uma aproximação mais adequada e simples, que calcula o número de vencedores sem valores impraticáveis.

2.2 Resposta:

Primeiro faremos a distribuição exata, que no caso é a binomial, para isso usaremos algumas formulas.

Para calcular o valor esperado faremos:

$$E[X] = n \cdot p \quad (8)$$

E para calcular a variância faremos:

$$Var(X) = n \cdot p \cdot (1 - p) \quad (9)$$

Sabendo que no contexto dessa questão as variáveis são:

- $n = 10^7$: Número total de visitantes diários;
- $p = 10^{-7}$: Probabilidade individual de um usuário ser premiado;

Portanto teremos:

$$E[X] = 10^7 \cdot 10^{-7} = 1 \quad (10)$$

$$Var(X) = 10^7 \cdot 10^{-7} \cdot (1 - 10^{-7}) = 0.9999999 \quad (11)$$

Agora faremos os cálculos para a distribuição aproximada que é a de Poisson, usando as seguintes formulas.

Para calcular o valor esperado faremos:

$$E[X] = \lambda \quad (12)$$

E para calcular a variância faremos:

$$Var(X) = \lambda \quad (13)$$

Note que a única variável que precisamos é o λ . Para calculá-lo, usaremos a seguinte fórmula e faremos as substituições igual ao item anterior.

$$\lambda = n \cdot p = 10^7 \cdot 10^{-7} = 1 \quad (14)$$

Dessa forma ficamos com os seguintes valores:

$$E[X] = 1 \quad (15)$$

$$Var(X) = 1 \quad (16)$$

Ao comparar os resultados obtidos, observa-se que o valor esperado ($E[X]$) é idêntico em ambos os modelos, resultando em exatamente 1. No que diz respeito à variância

($Var(X)$), o cálculo exato pela distribuição Binomial fornece 0.9999999, enquanto a aproximação de Poisson fornece exatamente 1. A diferença entre os dois valores é muito pequena. Esta semelhança ocorre porque a distribuição de Poisson é o limite da distribuição Binomial quando $n \rightarrow \infty$ e $p \rightarrow 0$, mantendo o produto $n \cdot p$ constante. Como o fator de correção da variância binomial $(1-p)$ é extremamente próximo de 1, a dispersão dos dados nos dois modelos torna-se praticamente equivalente.

2.3 Resposta:

Para calcular a probabilidade de eu ser o ganhador dentre diversos outros ganhadores, sendo estes $W \sim Pois(1)$, realizaremos o cálculo da esperança média desse evento acontecer, que se dá pelo somatório da média ponderada dos valores por suas probabilidades, de acordo com o conceito de variáveis aleatórias discretas.

$$P(X = 1_{(vencedor)}) = E\left[\frac{1}{1_{(ganhadorgarantido)} + w_{(outrosganhadores)}}\right]$$

$$P(X = 1) = \sum_0^{\infty} \frac{1}{1+w} \cdot P(W = w)$$

Dado $W \sim Pois(1)$, então $P(W = w) = \frac{e^{-1}}{w!}$

$$P(X = 1) = \sum_0^{\infty} \frac{1}{(1+w)} \cdot \frac{e^{-1}}{w!}$$

Tirando constante e^{-1} do somatório:

$$P(X = 1) = e^{-1} \sum_0^{\infty} \frac{1}{(1+w) \cdot w!}$$

Podemos reescrever $(1+w) \cdot w!$ como $(1+w)!$ dado a definição de fatorial.

$$P(X = 1) = e^{-1} \sum_0^{\infty} \frac{1}{(1+w)!}$$

Com isso podemos reduzir o denominador $(1+w)$ chamando-o pelo índice k , fazendo aparecer a série de Taylor, após isso trocamos os índices do somatório para reordenar os valores conhecidos.

$$P(X = 1) = e^{-1} \sum_1^{\infty} \frac{1}{k!}$$

O somatório de $\frac{1}{k!}$ indo de $0 \dots \infty$ nos dá o valor de euler, então de $1 \dots \infty$ nos dá $e - 1$.

$$P(X = 1) = e^{-1} \cdot (e - 1)$$

Utilizando calculadora científica, temos:

$$P(X = 1) = 0.632120558$$

Ou seja, a probabilidade de eu ser o ganhador final é de aproximadamente 63.212%

2.4 Resposta:

Para avaliar a precisão da aproximação de Poisson em relação ao modelo binomial exato, foi realizada uma simulação utilizando o R. A simulação reproduziu o cenário do site 10.000 vezes e, em cada repetição, ele processou 10 milhões de tentativas ($n = 10^7$) com uma probabilidade de sucesso de 10^{-7} (p), contabilizando o total de vencedores diários. Este procedimento gera um conjunto de dados que reflete o comportamento real esperado para a variável aleatória X.

Segue o código usado:

```
# Definindo as variaveis
n <- 10^7
p <- 10^-7
lambda <- n * p
n_simulacoes <- 10000

# Gerando a simulacao
dados_simulados <- rbinom(n_simulacoes, size = n, prob = p)

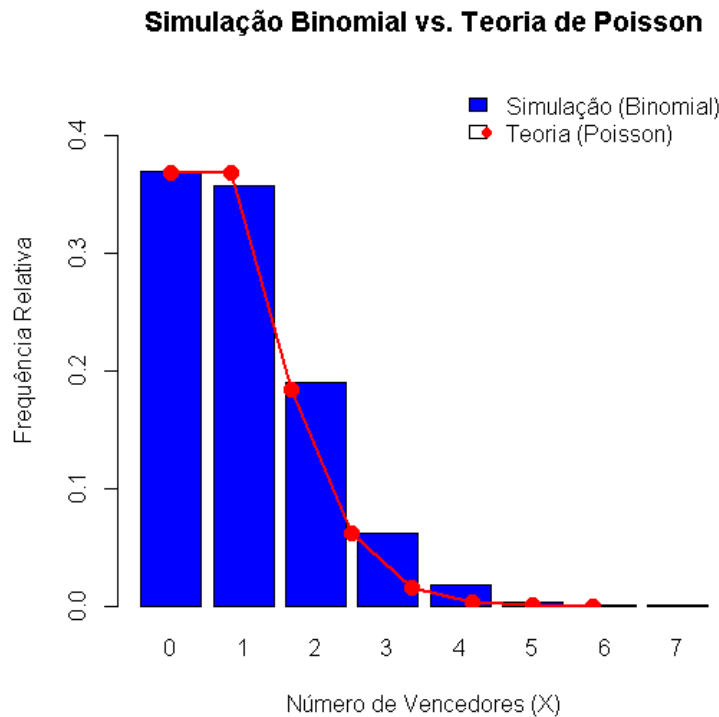
contagem <- table(dados_simulados)
freq_relativa <- as.numeric(contagem) / n_simulacoes
valores_k <- as.numeric(names(contagem))

# Criando o grafico
barplot(freq_relativa,
        names.arg = valores_k,
        col = "blue",
        main = "Simulacao_Binomial_vs_Teoria_de_Poisson",
        xlab = "Numero_de_Vencedores(X)",
        ylab = "Frequencia_Relativa",
        ylim = c(0, 0.45))

# Adicionando os pontos da Distribuicao de Poisson Teorica para comparar
pontos_poisson <- dpois(0:max(valores_k), lambda)
points(0:max(valores_k) + 0.7, pontos_poisson, col = "red", pch = 16, cex
      = 1.5)
lines(0:max(valores_k) + 0.7, pontos_poisson, col = "red", lwd = 2)

legend("topright", legend = c("Simulacao_Binomial", "Teoria_Poisson"),
      , fill = c("blue", NA), col = c(NA, "red"), pch = c(NA, 16), bty = "n"
      )
```

Na minha simulação o gráfico gerado foi:



O gráfico indica que a aproximação de Poisson é de alta qualidade, pois as frequências relativas obtidas pela simulação binomial coincidem quase perfeitamente com os valores teóricos da distribuição de Poisson. As maiores probabilidades concentram-se nos mesmos valores de X , e as diferenças observadas são mínimas e atribuíveis à variabilidade amostral. Isso confirma que, para valores muito grandes de n e probabilidades muito pequenas de p , a distribuição de Poisson é uma aproximação adequada e confiável da distribuição binomial.

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.

1. Crie uma função que gere valores com distribuição normal usando a transformação de

Box-Muller, a partir de entradas aleatórias uniformes. Especificamente:

- (a) Gere duas variáveis aleatórias uniformes independentes: $U_1, U_2 \sim Unif(0, 1)$
- (b) Calcule dois valores normais padrão usando as fórmulas de Box-Muller:

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

Z_1 e Z_2 são variáveis aleatórias independentes com distribuição normal padrão. Concatene-as para formar um vetor Z de valores normais padrão.

- (c) Converta cada valor normal padrão para a distribuição de temperatura da CPU:

$$T = 62 + 3.5Z$$

2. Use seu gerador de números aleatórios para gerar 1.000 medições de temperatura da CPU. Gere mais 1.000 valores de temperatura utilizando o gerador de números aleatórios normal embutido do R, com a mesma média e desvio-padrão.
3. Para ambos os conjuntos de dados simulados, calcule:
 - (a) Média amostral.
 - (b) Desvio-padrão amostral.
 - (c) Temperatura mínima e máxima observada.
 - (d) Probabilidade empírica e teórica $P(T > 68)$
 - (e) Probabilidade empírica e teórica $P(60 < T < 65)$
 - (f) Probabilidade teórica $P(T > 75)$. 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.
4. Visualize os resultados criando
 - (a) Um histograma das temperaturas simuladas da CPU (pode plotar os dois conjuntos de dados separadamente ou sobrepostos).
 - (b) A função densidade de probabilidade (PDF) normal teórica (média 62 °C, desvio padrão 3,5 °C) sobreposta ao histograma.
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?

SOLUÇÃO DA QUESTÃO 3

3.1.1 Descrição da atividade

A atividade aborda o estudo de variáveis aleatórias contínuas por meio da simulação de uma temperatura de CPU que segue distribuição Normal, utilizando tanto um gerador manual baseado na transformação de Box–Muller quanto o gerador normal embutido no R. O conteúdo envolve compreender como transformar números uniformes em valores normalmente distribuídos, analisar estatisticamente médias, desvios-padrão e probabilidades empíricas e teóricas, além de investigar eventos raros e sua relação com tamanhos amostrais. A atividade também introduz a comparação entre dados simulados e a densidade teórica, explorando conceitos de aderência à distribuição Normal, avaliação de sistemas físicos (como resfriamento da CPU) e o papel fundamental dos números uniformes na construção de geradores de variáveis aleatórias.

3.1.2 Variáveis aleatórias contínuas:

O que são?

- o Nas variáveis aleatórias contínuas são medidos valores imensuráveis que podem assumir qualquer número dentro de um intervalo contínuo, por exemplo: tempo, altura, voltagem, temperatura.

Valor médio esperado nas variáveis contínuas $E(X)$

- o Diferente das discretas que faz a média ponderada entre as probabilidades e valores possíveis, nas contínuas como temos infinitos valores teremos que integrar a função $f(x)$ em um intervalo desejado.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Mas geralmente para uma função qualquer temos a noção da LOTUS, afirmando que, para calcular o valor esperado de uma transformação $g(X)$ de uma variável aleatória X , não é necessário conhecer a distribuição de $g(X)$. Basta usar a PDF original de X :

$$E(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Função densidade de probabilidade FDP ou (PDF)

- o A função densidade de probabilidade descreve como a probabilidade está distribuída ao longo do intervalo contínuo de valores da variável aleatória. Diferentemente da função massa de probabilidade (PMF), usada em variáveis discretas, a PDF não fornece diretamente probabilidades pontuais, pois a probabilidade de termos o valor igual às suas infinitas casas decimais é $P(X = x_0) = 0$, por conta desse fato a probabilidade é obtida pela área sob a curva da PDF entre dois valores:

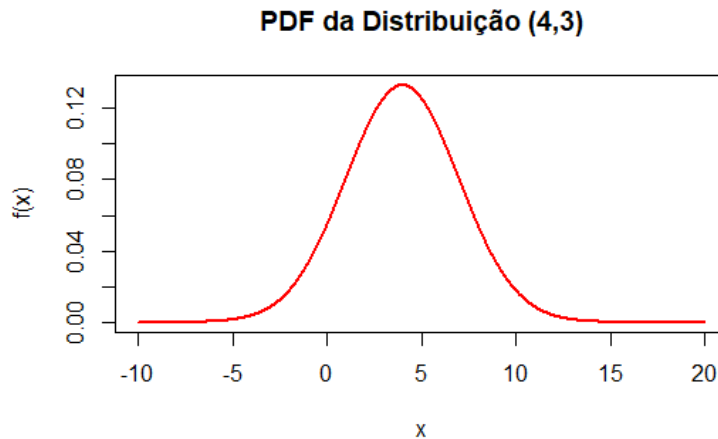
$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Propriedades da Função Densidade de Probabilidade (PDF) A função densidade de probabilidade $f(x)$ deve satisfazer duas propriedades fundamentais:

- A probabilidade não pode ser negativa: $f(x) \geq 0$ para todo x .
- A área total é 1, a integral de $-\infty$ a $+\infty$ deve ser igual a 1 para que a soma das probabilidades seja 100%.

Exemplo de gráfico de PDF

- Com média 4 e desvio padrão 3.

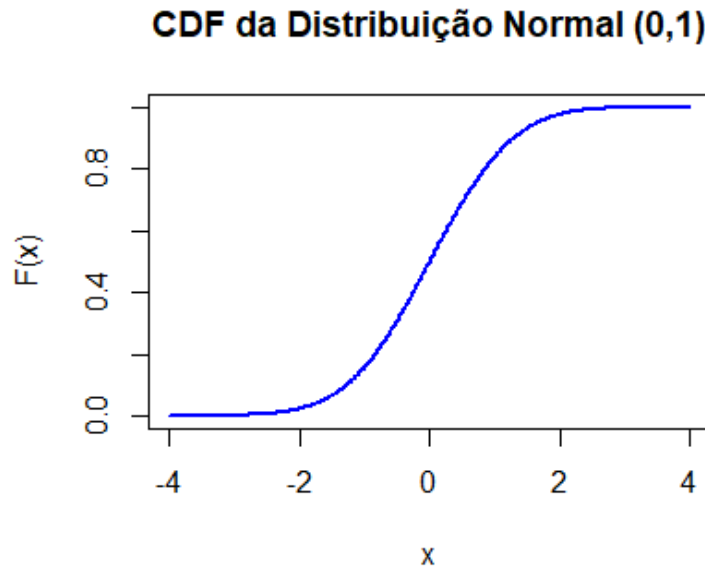


Função de distribuição acumulada contínua FDA ou (CDF)

- Muito parecido com a CDF das variáveis aleatórias discretas as variáveis aleatórias contínuas terão o mesmo conceito, mas que em vez de um somatório das probabilidades de cada valor, aqui teremos uma integral de $-\infty$ até um valor x , utilizado normalmente para descobrir a probabilidade até um certo ponto $\leq X$.

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Outra diferença é no formato do gráfico, que não será mais em formato de escada e sim de curva, segue exemplo de CDF com média 0 e desvio padrão 1.



Propriedades da Função de distribuição acumulada contínua FDA ou (CDF)

- **Limite inferior igual a zero:**

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$

Antes de qualquer região de densidade, a probabilidade acumulada é nula.

- **Limite superior igual a um:**

$$\lim_{x \rightarrow +\infty} F(x) = 1.$$

Após considerar toda a densidade, toda a probabilidade já foi acumulada.

- **Derivada da CDF é a PDF (quando derivável):**

$$f(x) = \frac{d}{dx} F(x).$$

Portanto, a PDF é a “taxa de crescimento” da CDF.

- **Valores entre 0 e 1:**

$$0 \leq F(x) \leq 1.$$

Porque probabilidades sempre pertencem ao intervalo unitário.

3.1.3 Distribuições contínuas importantes:

Distribuição Uniforme [$X \sim U(a, b)$]

- Todos os valores do intervalo possuem igualmente a mesma probabilidade, possuindo um valor inicial a e um valor final b .
- PDF:

$$f(X) = \frac{1}{b - a}$$

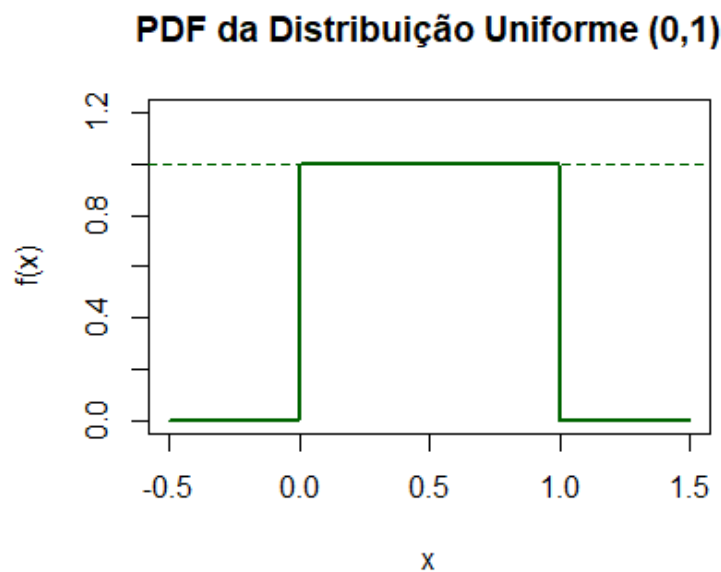
- Esperança:

$$E(X) = \frac{a + b}{2}$$

- Variância:

$$Var(X) = \frac{(b - a)^2}{12}$$

- Gráfico da PDF: Exemplo com $a = 0$ e $b = 1$



Distribuição Normal [$X \sim N(\mu, \sigma^2)$]

- A distribuição Normal ou Gaussiana é a distribuição que modela variáveis que surgem pela soma de muitos efeitos pequenos, normalmente vistas no dia a dia, exemplo: altura humana, notas de um teste...

- PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Outra forma:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Legenda:
 μ : média
 σ desvio padrão
 σ^2 variância

- Esperança:

$$E(X) = \mu$$

- Variância:

$$Var(X) = \sigma^2$$

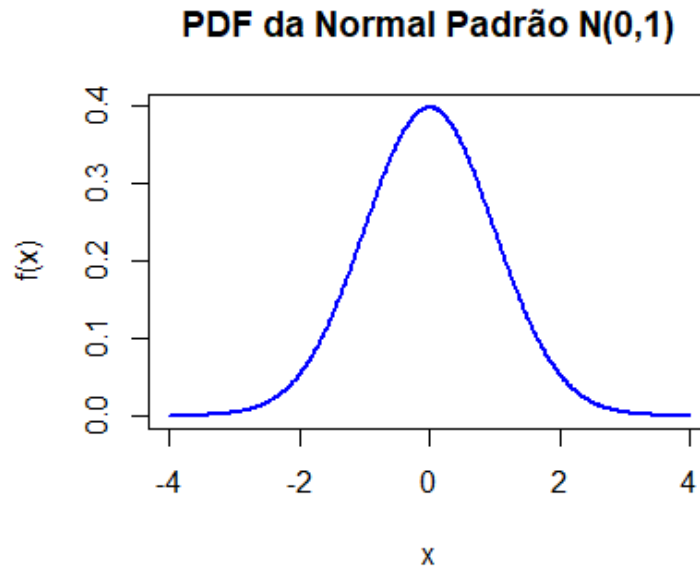
Características da Normal

- É simétrica, ou seja, a média, moda e mediana são iguais e ficam no centro.
- Regra do 68,95,99,7 para $X \sim N(\mu, \sigma)$, que diz:
68% dos valores estão em $\mu \pm \sigma$
95% dos valores estão em $\mu \pm 2\sigma$
99,7% dos valores estão em $\mu \pm 3\sigma$

Distribuição Normal Padrão [$X \sim N(0, 1)$]

É a Normal em que a média é 0 e o desvio padrão é 1.

- Gráfico da PDF de uma Normal Padrão.



3.1.4 Padronização e Cálculo de Probabilidades:

Como calcular a probabilidade de uma função contínua difícil?

- Para calcular a probabilidade de uma função contínua difícil, precisaremos que ela seja próxima da Normal, com isso seguimos o procedimento de normalizar os dados para uma distribuição de média 0 e desvio padrão 1, após isso utilizamos a tabela de integrais da função de distribuição Normal Padrão, os próximos tópicos detalharão melhor.

Padronização Z-score:

- O método de padronização dos dados de uma função qualquer para após isso utilizar a tabela de integrais se dá pela fórmula:

$$Z = \frac{x - \mu}{\sigma}$$

Tabela de integrais para distribuição Normal Padrão:

- Após padronizar os dados podemos aplicar qualquer ponto x no Z-score e obter seu valor de Z através da tabela.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985
3.0	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9990	0.9990

Tabela 2: Tabela acumulada da Normal Padrão $\Phi(z)$ estendida até $z = 3.0$ e coluna 0.07.

3.1.5 Geração de Variáveis Contínuas:

Transformação Box–Muller

- o A Transformação de Box–Muller é um método utilizado para gerar variáveis aleatórias com distribuição Normal Padrão a partir de duas variáveis aleatórias independentes com distribuição Uniforme no intervalo $(0, 1)$:

$$U_1, U_2 \sim \text{Uniforme}(0, 1).$$

A ideia fundamental é transformar duas variáveis uniformes em duas variáveis normais independentes. Isso é feito utilizando a seguinte transformação:

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

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

As variáveis resultantes Z_1 e Z_2 são independentes e seguem distribuição Normal Padrão:

$$Z_1, Z_2 \sim \mathcal{N}(0, 1)$$

Respostas dos itens da questão 3:

3.1 Resposta:

Criamos a função *boxMuller* que transforma 2 variáveis aleatórias uniformes em 2 variáveis com distribuições normais padrão, após isso temos a função *disT* que aplica a distribuição da temperatura na CPU para uma variável Z normalizada, e por fim temos a função que une as duas anteriores e gera $2k$ de variáveis aleatórias já que as funções anteriores tem 2 entradas e 2 saídas.

```
# Questao 3.1
rm(list = ls())

boxMuller <- function(Unif1, Unif2){
  Z1 <- sqrt(-2*log(Unif1)) * cos(2*pi*Unif2);
  Z2 <- sqrt(-2*log(Unif1)) * sin(2*pi*Unif2);

  vetorZ <- c(Z1,Z2);

  return(vetorZ);
}

disT <- function(Z){
  return(62+(3.5*Z))
}

GeradorT <- function(k){
  vetor_total <- c()

  for (i in 1:k) {
    Unif1 <- runif(1,0,1);
    Unif2 <- runif(1,0,1);

    vetorZ <- boxMuller(Unif1,Unif2);
    vetorT <- disT(vetorZ);

    vetor_total <- c(vetor_total, vetorT)
  }
}
```

```
    return(vetor_total);
}
```

3.2 Resposta:

Estamos gerando 1000 valores da temperatura da CPU utilizando função feita no item anterior (*GeradorT*) e armazenando no vetor *simTemperaturaCPU₁*, após isso geramos 1000 valores da temperatura da CPU utilizando função do R com média $\mu = 62C$ e desvio padrão $\sigma = 3.5C$ e armazenando no vetor *simTemperaturaCPU₂*.

```
# Questao 3.2
rm(list = ls())

simTemperaturaCPU_1 <- GeradorT(500);

simTemperaturaCPU_2 <- rnorm(1000,62,3.5);
```

3.3 Resposta:

Para a simulação utilizando a nossa função *boxMuller* será referenciado pelo nome *SimTemperaturaCPU_1* e para a simulação feita pela função do R utilizaremos a seguinte referência: *SimTemperaturaCPU_2*.

Item a) Utilizamos funções do próprio R, *mean* para calcular a média amostral.

```
mediaSimTemperaturaCPU_1 <- mean(simTemperaturaCPU_1)
mediaSimTemperaturaCPU_2 <- mean(simTemperaturaCPU_2)
```

Resultados obtidos:

```
> mediaSimTemperaturaCPU_1
[1] 61.94655
> mediaSimTemperaturaCPU_2
[1] 61.98488
```

Item b) Utilizamos funções do próprio R, *sd* para calcular o desvio padrão.

```
desvioPadraoSimTemperaturaCPU_1 <- sd(simTemperaturaCPU_1)
desvioPadraoSimTemperaturaCPU_2 <- sd(simTemperaturaCPU_2)
```

Resultados obtidos:

```
> desvioPadraoSimTemperaturaCPU_1
[1] 3.620178
> desvioPadraoSimTemperaturaCPU_2
[1] 3.541967
```

Item c) Utilizamos funções do próprio R, *min* e *max* para calcular o máximo e mínimo amostral.

```
temperaturaMinSimTemperaturaCPU_1 <- min(simTemperaturaCPU_1)
temperaturaMinSimTemperaturaCPU_2 <- min(simTemperaturaCPU_2)

temperaturaMaxSimTemperaturaCPU_1 <- max(simTemperaturaCPU_1)
temperaturaMaxSimTemperaturaCPU_2 <- max(simTemperaturaCPU_2)
```

Resultados obtidos:

```
> temperaturaMinSimTemperaturaCPU_1
[1] 50.85293
> temperaturaMinSimTemperaturaCPU_2
[1] 48.33085
> temperaturaMaxSimTemperaturaCPU_1
[1] 73.13087
> temperaturaMaxSimTemperaturaCPU_2
[1] 73.73896
```

Item d) Utilizamos o R para calcular a probabilidade empírica $P(T > 68)$, se dá pelo somatório das temperaturas simuladas que sejam maiores do que 68 dividida pelo total de amostras (1000).

```
#Empírica:

probEmpSimTemperaturaCPU_1 <- sum(simTemperaturaCPU_1 > 68) / length(
  simTemperaturaCPU_1)
probEmpSimTemperaturaCPU_2 <- sum(simTemperaturaCPU_2 > 68) / length(
  simTemperaturaCPU_2)
```

Resultados obtidos da probabilidade empírica de $P(T > 68)$:

```
> probEmpSimTemperaturaCPU_1
[1] 0.04
> probEmpSimTemperaturaCPU_2
[1] 0.053
```

Diferentemente da empírica na teórica não dependemos de dados coletados, somente da função de distribuição da temperatura, com ela podemos aplicar a padronização Z-score e utilizar a tabela das integrais para descobrir qual a probabilidade de $T > 68$.

Função Z-score 3.1.4 aplicado à função T:

$$Z = \frac{68 - 62}{3.5}$$

Para $P(T > 68)$:

$$P(Z > \frac{68 - 62}{3.5}) = P(Z > 1.7143)$$

Diminuindo 1 da área acumulada até valor 68, para sabermos a probabilidade da área que é maior que o valor 68.

$$1 - \Phi(1.71)$$

Olhando na tabela 3.1.4:

$$1 - 0.9564 = 0.0436$$

Ou seja, a probabilidade teórica de $T > 68$ é de 4.36%. Agora calculamos o mesmo valor usando o R:

```
#Teorica:
probTeoSimTemperaturaCPU_1 <- 1 - pnorm(68, mean = 62, sd = 3.5)
probTeoSimTemperaturaCPU_2 <- 1 - pnorm(68, mean = 62, sd = 3.5)
```

Resultados obtidos da probabilidade teórica de $P(T > 68)$:

```
> probTeoSimTemperaturaCPU_1
[1] 0.04323813
> probTeoSimTemperaturaCPU_2
[1] 0.04323813
```

Podemos calcular o erro obtido:

$$Erro = \left| \frac{0.04323813 - 0.0436}{0.04323813} \right| \cdot 100 = 0.837\%$$

Item e) Utilizamos o R para calcular a probabilidade empírica $P(60 < T < 65)$, se dá pelo somatório das temperaturas simuladas que estejam entre 60 e 65 dividida pelo total de amostras (1000).

```
#Empirica:
probEmpSimTemperaturaCPU_1 <- sum(60 < simTemperaturaCPU_1 &
  simTemperaturaCPU_1 < 65) / length(simTemperaturaCPU_1)
probEmpSimTemperaturaCPU_2 <- sum(60 < simTemperaturaCPU_2 &
  simTemperaturaCPU_2 < 65) / length(simTemperaturaCPU_2)

#Teorica:
probTeoSimTemperaturaCPU_1 <- pnorm(65, mean = 62, sd = 3.5) - pnorm(60,
  mean = 62, sd = 3.5)
probTeoSimTemperaturaCPU_2 <- pnorm(65, mean = 62, sd = 3.5) - pnorm(60,
  mean = 62, sd = 3.5)
```

Resultados obtidos da probabilidade empírica de $P(60 < T < 65)$:

```
> probEmpSimTemperaturaCPU_1
[1] 0.511
> probEmpSimTemperaturaCPU_2
[1] 0.495
```

Agora Para a probabilidade teórica podemos obter utilizando a mesma estratégia de utilizar a padronização Z-score, mas desta vez subtraímos o maior limite de $Z_2 = 65$ pelo menor limite de $Z_1 = 60$.

Calculando Z_1 pela fórmula 3.1.4.

$$Z_1 = \frac{60 - 62}{3.5}$$

$$Z_1 = -0.5714$$

Calculando Z_2 pela fórmula 3.1.4.

$$Z_2 = \frac{65 - 62}{3.5}$$

$$Z_2 = 0.8571$$

Agora podemos calcular $P(60 < T < 65)$ subtraindo $\Phi(Z_2)$ e $\Phi(Z_1)$:

$$P(60 < T < 65) = \Phi(0.8571) - \Phi(-0.5714)$$

como Z_1 está negativo usamos uma solução de transforma-lo em positivo, subtraímos o positivo dele por 1.

$$P(60 < T < 65) = \Phi(0.8571) - (1 - \Phi(0.5714))$$

Abrindo os parênteses:

$$P(60 < T < 65) = \Phi(0.8571) - 1 + \Phi(0.5714)$$

Reordenando e mudando deixando com 2 algarismos significativos:

$$P(60 < T < 65) = \Phi(0.85) + \Phi(0.57) - 1$$

Aplicando os valores da tabela 3.1.4

$$P(60 < T < 65) = 0.8023 + 0.7157 - 1$$

Resultado final:

$$P(60 < T < 65) = 0.518$$

Pelos cálculos teóricos a probabilidade de $(60 < T < 65)$ é de aproximadamente 51.8%.

Resultados obtidos da probabilidade teórica de $P(60 < T < 65)$ utilizando R:

```
probTeoSimTemperaturaCPU_1 <- pnorm(65, mean = 62, sd = 3.5) - pnorm(60,
  mean = 62, sd = 3.5)
probTeoSimTemperaturaCPU_2 <- pnorm(65, mean = 62, sd = 3.5) - pnorm(60,
  mean = 62, sd = 3.5)
```

Resultados obtidos:

```
> probTeoSimTemperaturaCPU_1
[1] 0.5204624
> probTeoSimTemperaturaCPU_2
[1] 0.5204624
```


Fazendo o cálculo do erro:

$$Erro = \left| \frac{0.5204624 - 0.518}{0.5204624} \right| \cdot 100 = 0.47\%$$

Item f) Utilizamos o R para calcular a probabilidade teórica $P(T > 75)$, se dá pelo somatório das temperaturas simuladas que sejam maiores do que 75 de amostras (1000).

```
probTeo_T_maior_75 <- 1 - pnorm(75, mean = 62, sd = 3.5)
```

Resultado

```
> probTeoSimTemperaturaCPU  
[1] 0.0001018892
```

Verificando se existe alguma acima de 75

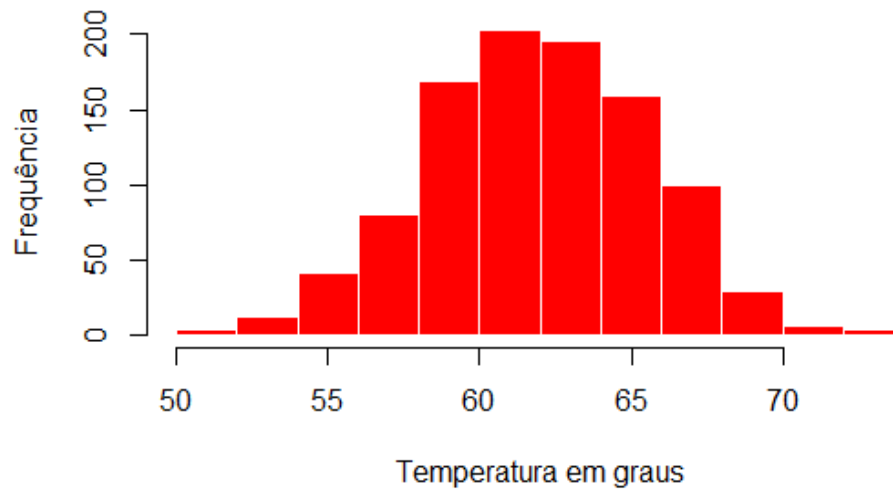
```
> any(simTemperaturaCPU_1 > 75)  
[1] FALSE  
> sum(simTemperaturaCPU_1 > 75)  
[1] 0  
> any(simTemperaturaCPU_2 > 75)  
[1] FALSE  
> sum(simTemperaturaCPU_2 > 75)  
[1] 0
```

Nota-se que, embora exista uma probabilidade teórica diferente de zero de a temperatura exceder 75 °C, esse valor é extremamente pequeno. Em um conjunto de apenas 1.000 amostras, é perfeitamente possível que nenhum valor acima desse limiar seja observado. Para que eventos tão raros apareçam nos dados simulados, é necessário um volume de amostras muito maior, pois somente com um número elevado de observações a frequência empírica desses eventos se aproxima da probabilidade teórica. Isso ocorre porque eventos de baixa probabilidade tendem a se manifestar apenas em amostras grandes, conforme previsto pela Lei dos Grandes Números.

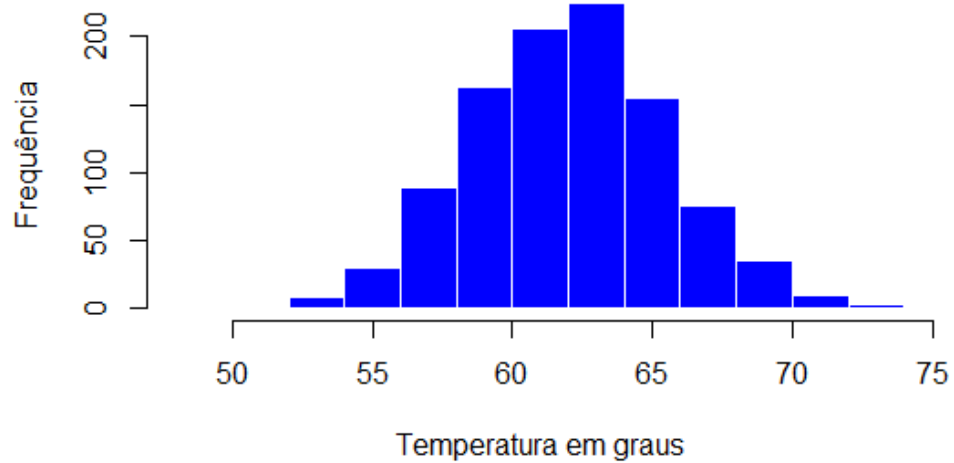
3.4 Resposta:

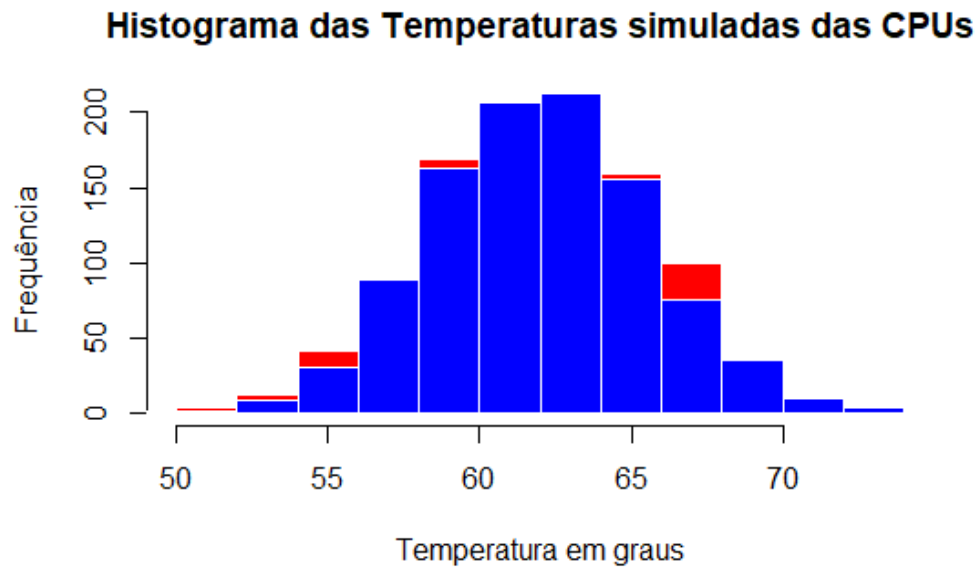
Item a) Utilizamos a linguagem R para a criação dos histogramas.

Histograma da Temperatura simulada da CPU 1



Histograma da Temperatura simulada da CPU 2



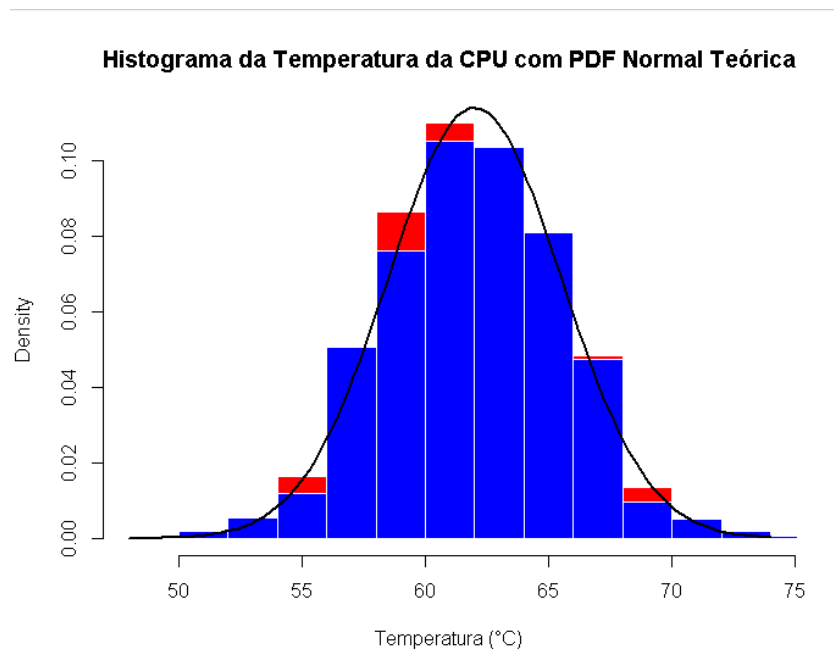


Código para criar os gráficos:

```
hist(simTemperaturaCPU_1,
     main = "Histograma das Temperaturas simuladas das CPUs",
     xlab = "Temperatura em graus",
     ylab = "Frequência",
     col = "red",
     border = "white"
)

hist(simTemperaturaCPU_2,
     main = "Histograma da Temperatura simulada da CPU 2",
     xlab = "Temperatura em graus",
     ylab = "Frequência",
     col = "blue",
     border = "white",
     add = TRUE
)
```

Item b)



Código para criar o gráfico:

```
# Histograma com densidade
hist(simTemperaturaCPU_1,
     probability = TRUE,    # transforma em densidade
     col = "red",
     border = "white",
     main = "Histograma da Temperatura da CPU com PDF Normal Teórica",
     xlab = "Temperatura ( C )")

hist(simTemperaturaCPU_2,
     probability = TRUE,    # transforma em densidade
     col = "blue",
     border = "white",
     add = TRUE)

# Curva da PDF normal teórica
curve(dnorm(x, mean = 62, sd = 3.5),
      col = "black",
      lwd = 2,
      add = TRUE)
```

3.5 Resposta:

As distribuições empíricas das temperaturas da CPU apresentam boa concordância com a curva normal teórica, exibindo formato aproximadamente simétrico e unimodal. As diferenças observadas, principalmente nas caudas, são compatíveis com a variabilidade esperada em amostras finitas.

A média amostral e o desvio-padrão amostral obtidos nas simulações mostram-se próximos dos valores teóricos de 62 °C e 3,5 °C, com pequenas discrepâncias atribuídas ao caráter aleatório do processo de geração. Tais diferenças tendem a diminuir com o aumento do tamanho amostral.

Não foram observadas diferenças relevantes entre os dados gerados pelo RNG (Random Number Generation) manual e pelo gerador embutido do R, sendo ambos compatíveis com a distribuição normal teórica. Entretanto, o RNG do R apresenta maior robustez estatística.

A simulação auxilia na avaliação de estratégias de resfriamento e de escalonamento dinâmico de clock ao permitir a análise probabilística de temperaturas elevadas sob diferentes condições operacionais. Geradores uniformes são a base dos sistemas de RNG, pois possibilitam a geração de distribuições mais complexas por meio de transformações matemáticas adequadas.