Previsão da Temperatura Máxima Diária

Uma simulação com Monte Carlo

Tailine J. S. Nonato

2023-12-21

1 Introdução

A introdução deste estudo sobre simulação de Monte Carlo para estimativa de temperatura máxima se concentra na importância da compreensão e previsão dos padrões climáticos. Explorando a técnica estatística de Monte Carlo para simular 100 amostras de 365 dias de temperatura máxima, esse estudo busca examinar a variação sazonal, distribuição e comportamento temporal das médias estimadas. A introdução destaca a relevância de prever as variações climáticas e como a simulação de Monte Carlo oferece uma abordagem poderosa para modelar e compreender esses fenômenos.

2 Metodologia

O estudo parte da definição de parâmetros, seguida pela geração de `n` amostras de `m` dias de temperatura máxima utilizando a técnica de Monte Carlo. Nesse momento, parte-se de algumas informações preexistentes:

- A variação de temperatura anual pode ser representada por função senoidal, onde cria-se uma sequência de números de 1 até a quantidade de dias a ser prevista no estudo e divide-se essa sequência por 365 para normalizá-la em um período de 1 ano, conforme descrito a seguir:
- Uma função senoidal gera valores que variam entre -1 e 1 ao longo do ano, representando variações sazonais que se assemelham aos padrões climáticos reais.

A função criada para gerar os dados climáticos simula séries temporais de temperatura máxima diária ao longo de um período dada uma média de temperatura, a amplitude da variação de temperatura e o desvio padrão dos erros/ruídos, ou seja, a parte componente aleatória presente nos dados.

3 Resultados

3.1 Geração dos dados da simulação

Para iniciar os processos, define-se arbitrariamente o período ao qual deseja-se prever, para este estudo definiu-se um período de 365 dias (1 ano), a partir de 1 de janeiro de 2024. Logo depois, define-se a função que tem como parâmetros:

- `n`: número de amostras a serem geradas
- `dias`: número de dias para cada amostra
- μ: média da temperatura máxima
- `amplitude`: amplitude da variação sazonal (representada por uma função senoidal)
- σ: desvio padrão do ruído

3.2 Parâmetros da simulação

Com a função pronta, define-se então os parâmetros da simulação. Serão geradas na simulação 100 amostras utilizando os dados do Boletim Climatológico do INMET do versão 2022/2023 de Brasília, onde a temperatura média é de 21,9° C, a amplitude de variação de acordo com a média da temperatura mínima (18,2°C) e a média da máxima (27,4°C), é de 9,2 e o desvio padrão dos ruídos arbitrário, será fixado em 5. Ou seja:

$$\mu = 22,9$$

$$\sigma = 5$$

$$\text{Amplitude} = 9,2$$

3.3 Exploração dos dados gerados

Criando um banco de dados com 365 datas a partir de 01 de janeiro, a médias por data das temperaturas geradas em cada uma das 100 amostras

Variáveis
indice
datas
temp_max_1
...
temp_max_100
media

3.4 Medidas resumo

Observando as medidas resumo da média de temperaturas máximas, tem-se que:

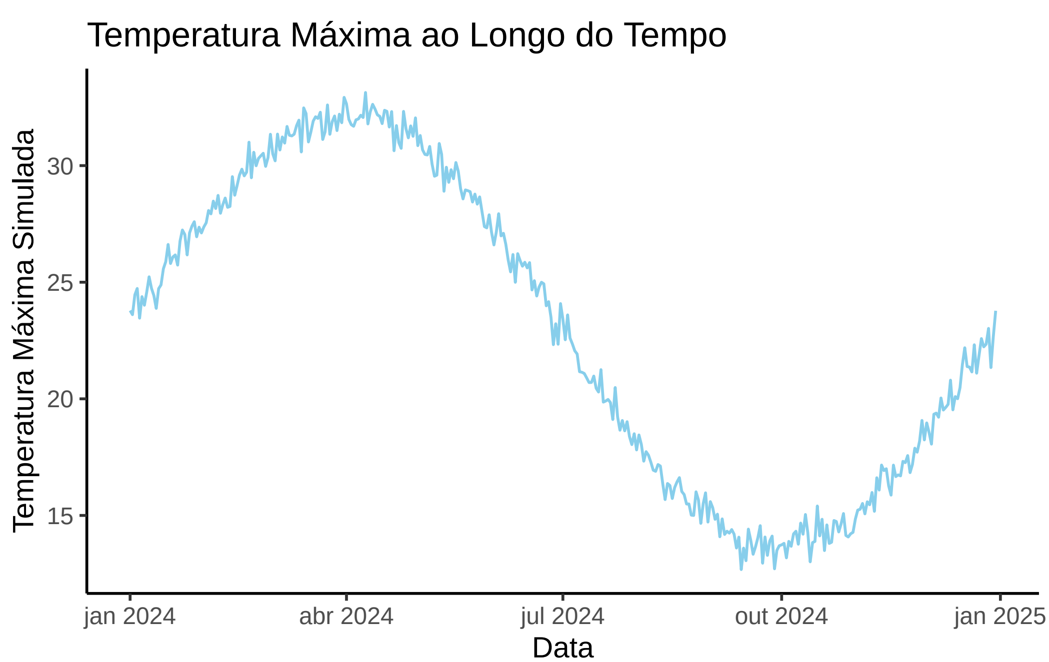
minimum	q1	median	mean	q3	maximum
12.68	16.44	23.22	22.95	29.44	33.13

sd
6.54804

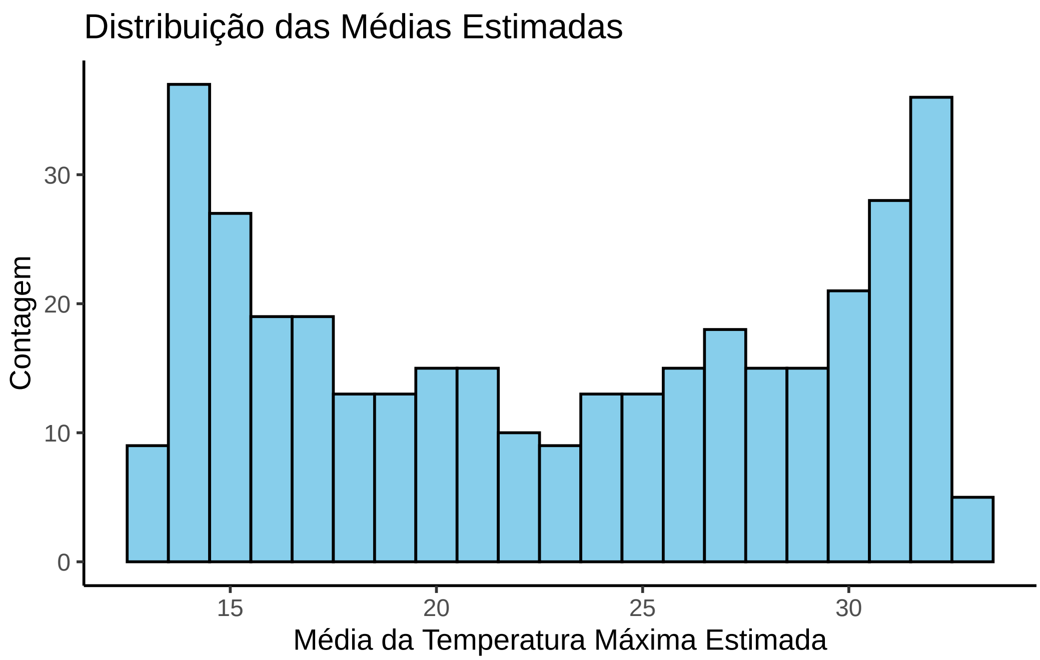
Dado o contexto brasiliense, apesar de ter a média conforme o parâmetro inicial, uma temperatura máxima de 12,8°C pode ser baixa demais, assim como uma máxima 33°C pode ser pouco, já que, dadas as últimas ondas de calor, foram registradas no DF temperaturas mais altas que isso. Mas é importante ressaltar que tudo isso está sendo influenciado pelos parâmetros pré-definidos.

Realiza-se a seguir representações gráficas para complementar as medidas resumo e visualizar propriamente os dados gerados pela simulação.

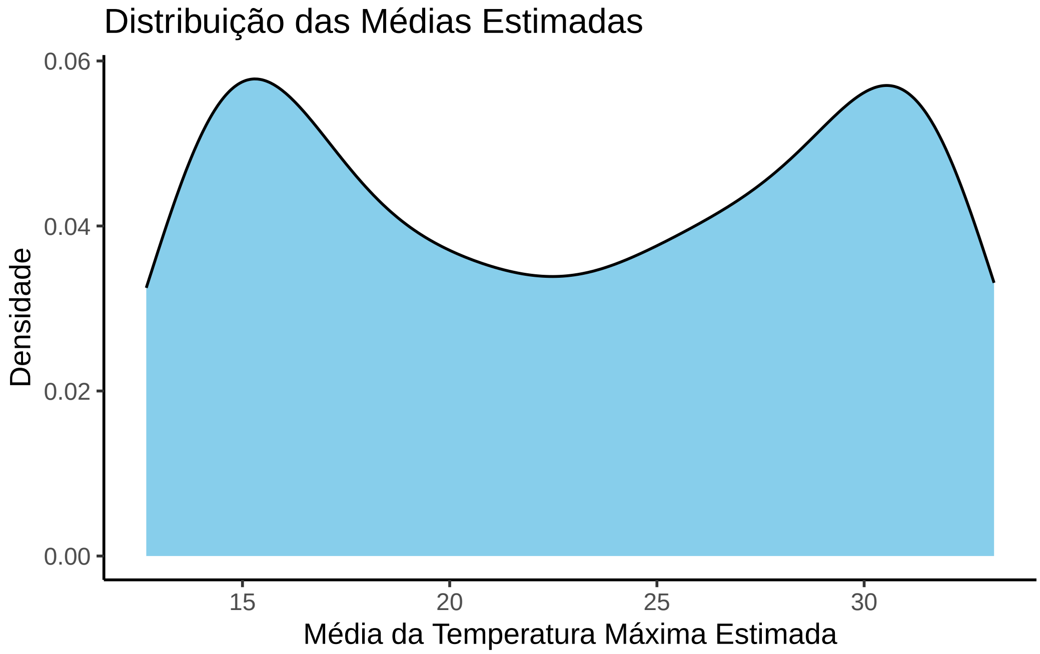
3.4.1 Série temporal



3.4.2 Histograma das médias



3.4.3 Densidade das médias

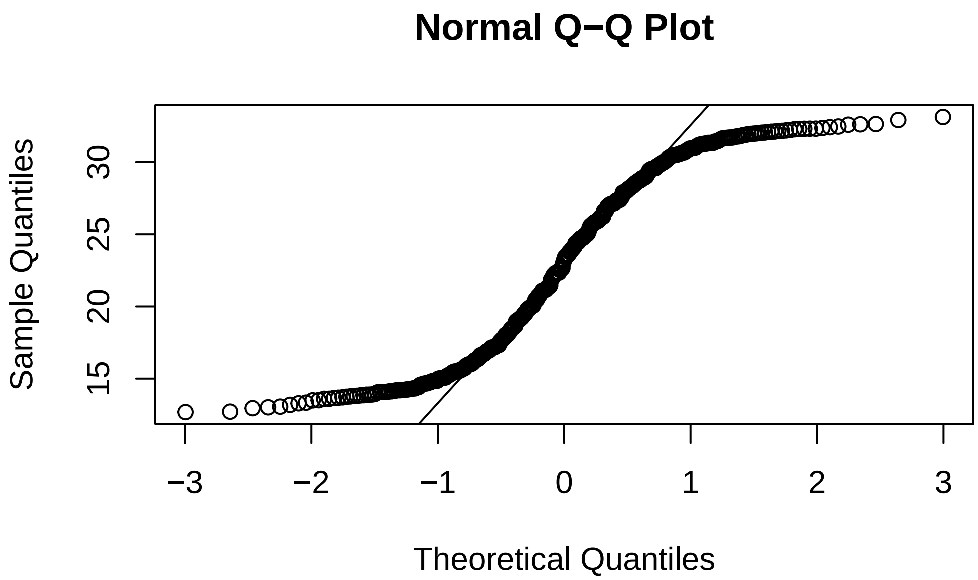


Observa-se nas 3 representações gráficas a variação sazonal por meio da função senoidal de forma que de janeiro a abril as temperaturas sobem e atingem o pico do ano e a partir de maio, aproximadamente, começam a cair até atingir as menores temperaturas em outubro, onde logo após começam a subir mais uma vez.

3.4.4 Teste de Normalidade

Observando os dados, tem-se como hipóteses:

$$\begin{cases} H_0 : \text{Os dados seguem distribuição Normal} \\ H_1 : \text{Os dados não seguem distribuição Normal} \end{cases}$$



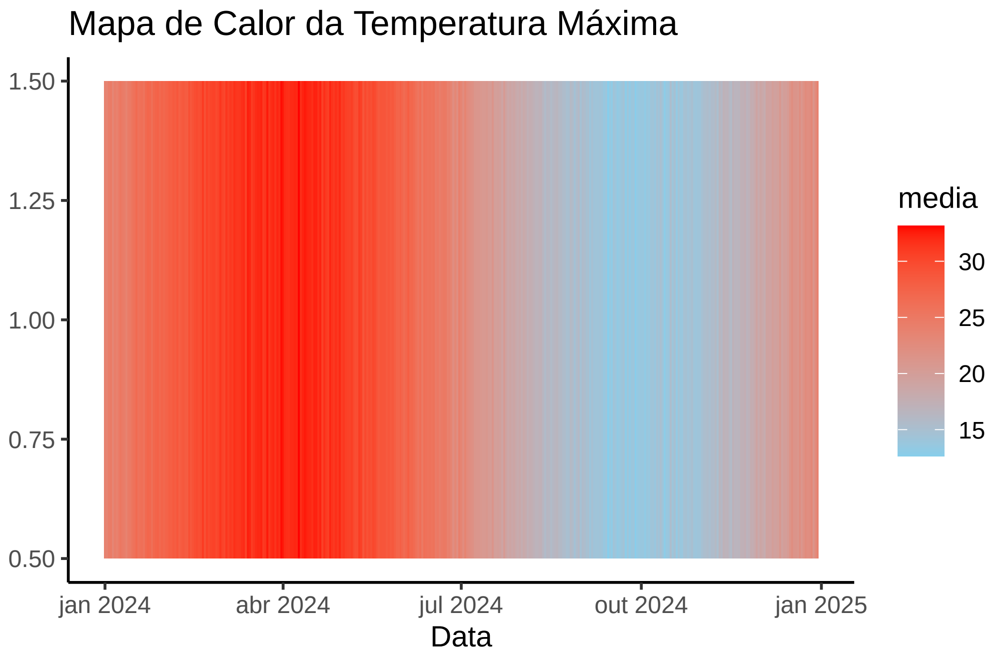

```
    Shapiro-Wilk normality test

data:  dados_simulados$media
W = 0.90989, p-value = 6.114e-14
```

A análise gráfica e os resultados do teste Shapiro-Wilk apontam para rejieção de H_0, constatando a não normalidade univariada dos dados.

3.4.5 Mapa de calor

Para finalizar, realiza-se então, um mapa de calor simulando um distribuição espacial:



Embora não apresente coordenadas geográficas reais, proporciona uma visualização da distribuição espacial fictícia das temperaturas máximas médias estimadas.

4 Conclusões

Ao realizar 100 simulações de 365 dias cada é possível observar uma clara variação sazonal nas médias estimadas da temperatura máxima ao longo do tempo. Com temperaturas mais elevadas no mês de abril e mais baixas no mês de outubro, o que pode não ser tão compatível com a realidade local.

Oportunidades:

- É válido realizar um estudo mais aprofundado para constatar o comportamento das médias quanto a distribuição Normal utilizando técnicas Multivariadas.
- Também é válido vincular essa simulação com dados espaciais reais de Latitude e Longetude para observar o mapa de calor.

5 Apêndices

Código em R:

```r
# Carregamento de pacotes
if (!require(pacman)) install.packages('pacman')
pacman::p_load(tidyverse,knitr,broom)

# Função Método Monte Carlo
gerar_dados_climaticos <- function(n, dias, mu, amplitude, sigma){
  temperaturas <- matrix(NA, nrow = dias, ncol = n)
  for (i in 1:n) {
    variacao_anual <- sin(2 * pi * (1:dias) / 365) * amplitude
    ruido <- rnorm(dias, mean = 0, sd = sigma)
    temperatura_maxima <- mu + variacao_anual + ruido
    temperaturas[, i] <- temperatura_maxima}
  return(temperaturas)}

# Parâmetros da simulação
n <- 100
dias <- 365
mu <- 22.9
amplitude <- 9.2
sigma <- 5

# Simulação
simulacao_monte_carlo <- gerar_dados_climaticos(n, dias, mu, amplitude, sigma)

# Criando um banco de dados com a simulação e adicionando informações
indices <- c(1:365)
datas <- seq(as.Date("2024-01-01"), by = "day", length.out = dias)
medias_temperatura_maxima <- rowMeans(simulacao_monte_carlo)
dados_simulados <- data.frame(indice = indices,
                  datas = datas,
                  temperatura_maxima = simulacao_monte_carlo,
                  media = medias_temperatura_maxima)
kable(names(dados_simulados),align='c')

# Medidas resumo
kable(tidy(round(summary(dados_simulados$media),2)),align='c')
```

```r
kable(data.frame(sd = sd(dados_simulados$media)),align='c')

# Série temporal
ggplot(dados_simulados, aes(x = datas, y = media)) +
  geom_line(color = "skyblue") +
  labs(title = "Temperatura Máxima ao Longo do Tempo",
       x = "Data", y = "Temperatura Máxima Simulada")+
     theme_classic()

# Histograma das médias
ggplot(dados_simulados, aes(x = media)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Distribuição das Médias Estimadas",
       x = "Média da Temperatura Máxima Estimada", y = "Contagem")+
     theme_classic()

# Densidade
ggplot(dados_simulados, aes(x = media)) +
  geom_density(fill = "skyblue", color = "black") +
  labs(title = "Distribuição das Médias Estimadas",
       x = "Média da Temperatura Máxima Estimada", y = "Densidade")+
     theme_classic()

# Normalidade
qqnorm(dados_simulados$media)
qqline(dados_simulados$media)
shapiro.test(dados_simulados$media)

# Mapa de calor
ggplot(dados_simulados, aes(x = indice, y = 1, fill = media)) +
  geom_tile() +
  scale_fill_gradient(low = "skyblue", high = "red") +
  labs(title = "Mapa de Calor da Temperatura Máxima",
       x = "Índice", y = "")+
     theme_classic()
```