

# IA para Projetos Sustentáveis

## TAREFA 3: O DESAFIO DOS RECURSOS HÍDRICOS E PRODUTIVIDADE NA AMAZÔNIA

Alberto Fares Akel

25 de junho de 2025

### Contexto

A gestão dos recursos hídricos na Amazônia vem se tornando crítica: estiagens prolongadas e enchentes severas desorganizam os ciclos naturais, reduzindo a disponibilidade e a qualidade da água — e, por consequência, deteriorando a produtividade agrícola e aumentando doenças de veiculação hídrica. Diante desse cenário, líderes comunitários pediram uma avaliação baseada em dados, de modo a substituir percepções empíricas por evidências quantitativas.

Duas bases de dados foram coletadas:

**Climática** – chuvas previstas e reais, temperatura média, presença de variações climáticas fora do padrão e índices de umidade do solo.

**Socioeconômica** – volume de produção agrícola, incidência de doenças hídricas, acesso à água potável e indicador de segurança alimentar.

Os registros possuem amostragem quase diárias e apresentam problemas comuns de coleta de campo: duplicidades, grafias inconsistentes (“sim”, “não”, “nao”), datas em formatos variados, valores ausentes e outliers (por exemplo, chuvas  $> 700$  mm/dia). Esses ruídos impedem análises diretas e exigem uma estratégia robusta de limpeza, padronização e integração.

Após a limpeza, uma Análise Exploratória (EDA) deve investigar relações entre clima, produção, saúde e acesso à água mediante histogramas, dispersões, heatmaps e testes de correlação (inclusive com defasagens). Os resultados servirão de base para modelos preditivos e ações comunitárias de mitigação e adaptação.

**Problema:** Como a variabilidade climática impacta a produção agrícola, o acesso/qualidade da água potável e a incidência de doenças hídricas nas comunidades amazônicas?

### Objetivo Geral:

Quantificar e explicar essas inter-relações climáticas-socioeconômicas, gerando indicadores, visualizações e estatísticas que orientem decisões locais de infraestrutura hídrica, planejamento agrícola e saúde pública.

### Objetivos complementares:

1. **Data cleaning:** limpar e integrar as bases (remover duplicatas, padronizar categorias, unificar datas, tratar ausentes e outliers);
2. **Análise Exploratória:** Visualizações gráficas (Boxplots, histogramas, distribuições) e estatísticas descritivas e inferências;
3. **Insights e Recomendações Estratégicas:** Gerar insights estratégicos com base em evidências estatísticas, identificando períodos críticos, padrões de risco, limiares climáticos e grupos vulneráveis.

## 1 Data cleaning

**Objetivo:** limpar, padronizar e integrar as bases climática e socioeconômica.

1. **Leitura e diagnóstico inicial:** As duas bases de dados (`base_climatica.csv` e `base_socioeconomica.csv`) foram carregadas em dataframes. Foram utilizados os comandos `info()`, `nunique()` e `isnull().sum()` para mapear o tipo das variáveis, cardinalidade e presença de valores ausentes.
  - 120 registros por base, com 110 datas distintas.
  - Ausência de dados em: precipitação real (4), temperatura (2), produção e doenças (5).
2. **Conversão e ordenação das datas:** Os campos de data foram convertidos para o tipo `datetime` com `pd.to_datetime()`, e os registros foram ordenados cronologicamente com `sort_values(by='data')`.
3. **Remoção de duplicatas:** Registros duplicados foram removidos com `drop_duplicates()`, garantindo uma entrada única por data.
4. **Padronização de variáveis categóricas:** As colunas `variacao_climatica` e `acesso_agua_potavel` foram normalizadas de textos diversos ("*sim*", "*nao*", "*não*") para valores binários (1 para sim, 0 para não), utilizando `str.lower().map()`.
5. **Correção de valores negativos em variáveis numéricas:** Foi aplicada a função `abs()` em todas as colunas numéricas, assumindo que valores negativos são erros de digitação. Garantiu-se, assim, que todas as variáveis físicas tivessem somente valores positivos.
6. **Reindexação:** Após as limpezas, os dataframes tiveram seus índices redefinidos com `reset_index(drop=True)`.
7. **Identificação de datas não coincidentes:** Foram comparadas as datas das duas bases para identificar:
  - `datas_comuns` — datas presentes em ambas as bases.
  - `dfc_dif` e `dfs_dif` — datas exclusivas de cada fonte.
8. **Integração final das bases:** As bases foram unificadas por meio de `pd.merge()` utilizando junção `inner` pela chave `data`, assegurando que apenas datas comuns fossem mantidas. Sufixos `_clima` e `_socio` foram aplicados para evitar colisões de nomes de colunas.

**Resultado:** Obteve-se um *dataset* final integrado, padronizado e temporalmente consistente, livre de ruídos básicos de coleta e preparado para análises exploratórias robustas. Este conjunto servirá de base para a geração de insights e modelos preditivos sobre os impactos da variabilidade climática na produção e qualidade de vida das comunidades amazônicas.

**Comentários:** Para otimizar nosso fluxo de trabalho de limpeza de dados, consultamos a DeepSeek para orientação sobre a melhor sequência processual. Embora alguns procedimentos já estivessem em vigor, a contribuição da DeepSeek se mostrou vital durante a fase de merge dos conjuntos de dados. Isso nos permitiu otimizar o processo e realizar a validação de forma mais eficiente e com maior segurança.

## 2 Análise Exploratória – Parte 1: Boxplots e Outliers

Nesta etapa inicial da análise exploratória dos dados, adotou-se uma abordagem sistemática para identificação e tratamento de outliers, com o objetivo de mitigar distorções estatísticas e garantir maior confiabilidade nas análises subsequentes.

### 2.1 Metodologia de Detecção de Outliers

A função desenvolvida<sup>1</sup>, para este fim aplica dois critérios clássicos:

- **Z-Score:** valores com escore absoluto superior a 3 foram considerados atípicos;
- **Intervalo Interquartil (IQR):** observações fora do intervalo  $Q_1 - 1,5 \cdot IQR$  e  $Q_3 + 1,5 \cdot IQR$  foram marcadas como potenciais outliers.

Foi adotada uma estratégia conservadora, em que apenas registros que apresentavam comportamento atípico em **duas ou mais variáveis simultaneamente** foram removidos.

<sup>1</sup>Código gerado em deepseek: "Para os seguinte dataframe df, colunas (especifiquei name e types), fazer um código python para limpar outliers"

## 2.2 Resultados da Limpeza

Após a aplicação do critério combinado, obteve-se o seguinte panorama:

- **Linhas originais:** 101 registros
- **Registros removidos como outliers:** 6
- **Registros restantes após filtragem:** 95

## 2.3 Comparação Visual com Boxplots

A Figura 1 ilustra, por meio de gráficos do tipo *boxplot*, o impacto da remoção de outliers sobre a distribuição das principais variáveis climáticas e socioeconômicas.

Foram destacados os seguintes efeitos:

- Redução significativa da assimetria nas distribuições de `chuvas_reais_mm` e `volume_producao_tons`;
- Compressão dos *whiskers* e menor influência de extremos em variáveis como `temperatura_media_C` e `indicador_seguranca_alimentar`;
- Utilização de escala logarítmica nos boxplots de produção agrícola e doenças para melhor visualização da dispersão.

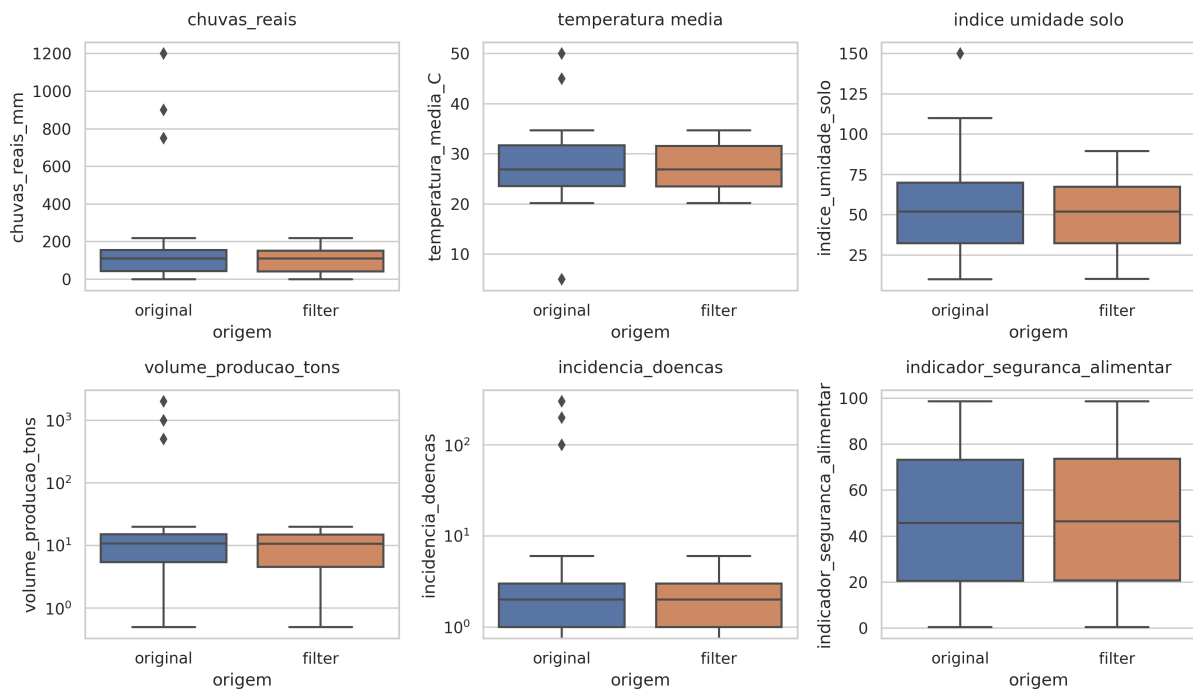


Figura 1: Boxplots comparando os dados originais e após remoção de outliers.

## 2.4 Auditoria de Dados Pós-Filtragem

Após o processo de limpeza, foi realizada uma verificação geral quanto à integridade do conjunto. As principais observações incluem:

- Permanência de **valores ausentes** em até cinco registros nas variáveis `chuvas_reais_mm`, `temperatura_media_C`, `umidade_solo`, `volume_producao_tons` e `incidencia_doencas`;
- Tipos de dados adequadamente padronizados, com variáveis binárias representadas por inteiros (0 e 1) e variáveis contínuas corretamente formatadas como `float64`.

## 2.5 Estatísticas Descritivas do Dataset Limpo

A Tabela 1 resume as estatísticas descritivas do conjunto de dados após a remoção de outliers:

Tabela 1: Resumo estatístico do conjunto de dados limpo ( $n = 95$ ).

Variável	Média	Desv. Padrão	Mín	$Q_1$	$Q_2$	$Q_3$
Chuvas previstas (mm)	100.6	59.9	0.9	46.8	109.8	151.8
Chuvas reais (mm)	104.4	62.0	0.5	42.3	110.0	152.6
Temperatura (°C)	27.2	4.4	20.2	23.5	26.9	31.6
Umidade do solo (%)	50.9	23.4	10.2	32.2	52.0	67.4
Produção (toneladas)	10.1	5.95	0.5	4.5	10.7	15.0
Doenças hídricas	1.9	1.47	0.0	1.0	2.0	3.0
Acesso à água (0/1)	0.46	0.50	0.0	0.0	0.0	1.0
Segurança alimentar	47.0	28.6	0.5	20.7	46.4	73.6

**Comentários:** Nesta etapa, utilizamos o DeepSeek para construir a rotina de filtragem dos outliers dados. Para a análise dos histogramas, o ChatGPT foi empregado para auxiliar na interpretação. Geramos os valores de frequências e bins e solicitamos insights. É importante notar que alguns dos insights gerados não correspondiam à realidade física, sendo, portanto, descartados. Realizamos as correções e adequações necessárias no texto para garantir a precisão da análise.

### 3 Análise Exploratória – Parte 2

Nesta seção, exploraremos visualmente as distribuições por meio de histogramas, que nos permitirão identificar padrões, assimetrias e a presença de múltiplas modas. Em seguida, para uma avaliação mais rigorosa da natureza dos dados, aplicaremos testes de normalidade formais. A combinação dessas abordagens visual e estatística nos capacitará a extrair os primeiros insights sobre o comportamento das variáveis, lançando as bases para modelos mais robustos e conclusões mais assertivas.

#### 3.1 Histogramas

A Figura 2 apresenta uma série de histogramas, oferecendo uma visualização detalhada da distribuição de frequência de cada uma das principais variáveis numéricas utilizadas.

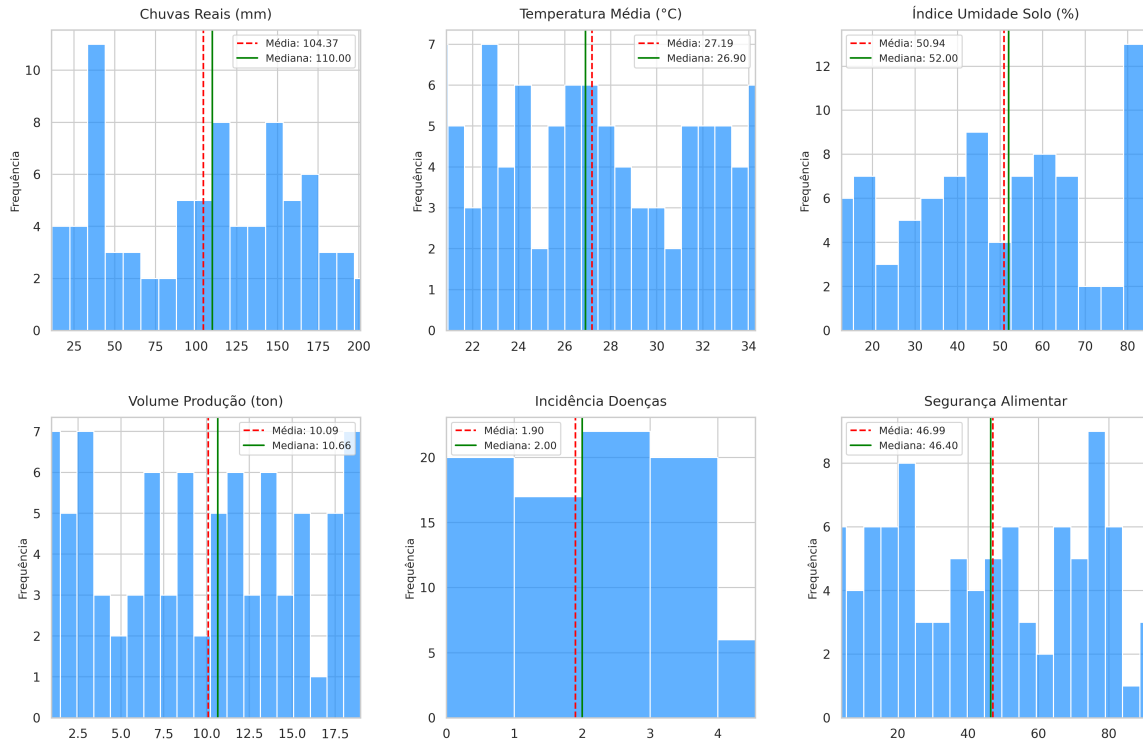


Figura 2: Distribuição das variáveis numéricas coletadas das bases de dados.

Essa análise<sup>2</sup> revelou padrões importantes:

- **Chuvas Reais e Volume de Produção** apresentaram distribuições assimétricas à direita, indicando a presença de poucos valores muito altos (longa cauda superior).
- **Temperatura Média e Índice de Umidade** mostraram distribuição levemente bimodal ou dispersa, com indícios de comportamento não normal.
- **Incidência de Doenças** concentrou a maior parte dos casos entre 1 e 3 ocorrências, com poucos registros extremos.
- **Segurança Alimentar** apresentou ampla dispersão, com tendência levemente concentrada nos valores baixos, possivelmente refletindo vulnerabilidades alimentares em determinadas comunidades.

Esses comportamentos são consistentes com a diversidade ambiental e socioeconômica da região, refletindo tanto a variabilidade climática quanto desigualdades no acesso a infraestrutura hídrica e serviços de saúde.

### 3.2 Distribuição de variáveis categóricas

A Figura 3 mostra a distribuição em relação a variação climática e ao acesso a água potável.

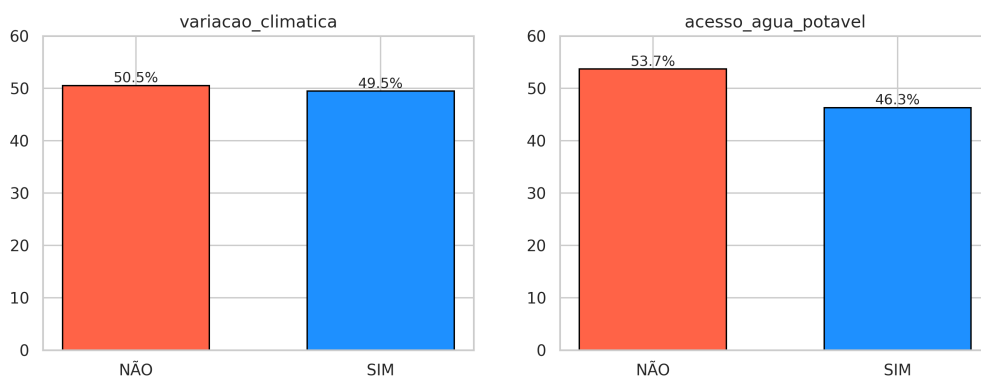


Figura 3: Distribuição das variáveis categóricas

Os dados mostram equilíbrio nas respostas, refletindo percepções locais e acesso desigual à infraestrutura hídrica. Tais padrões são coerentes com o foco da pesquisa em compreender os efeitos das mudanças climáticas sobre a saúde, agricultura e segurança alimentar.

### 3.3 Teste de Normalidade

O teste<sup>3</sup> de Shapiro–Wilk foi escolhido pelas seguintes razões:

- Alta sensibilidade para amostras pequenas a médias ( $3 \leq n \leq 2000$ ) — no nosso caso, cada variável possui cerca de 90–95 registros.
- Maior poder estatístico em detectar caudas pesadas e assimetrias do que os testes de Kolmogorov–Smirnov ou Anderson–Darling.
- Recomendado em estudos ambientais e epidemiológicos que trabalham com séries temporais curtas.

Todos os *p-values* foram menores que 0,05, implicando na **rejeição da hipótese de normalidade** para todas as variáveis analisadas conforme a tabela 2.

#### Consequências práticas:

- Utilizar correlações de **Spearman** em vez de Pearson;
- Empregar testes não paramétricos (ex.: Mann–Whitney, Kruskal–Wallis) nas comparações entre grupos;

<sup>2</sup>Análise via chatgpt: "veja os dados dos histogramas e busque por insights dentro do contexto e objetivo (Quantificar e explicar essas inter-relações climáticas-socioeconômicas, gerando indicadores, visualizações e estatísticas que orientem decisões locais de infraestrutura hídrica, planejamento agrícola e saúde pública".)

<sup>3</sup>código gerado em deepseek: "Para o seguinte dataframe df, crie uma função def para avaliar a normalidade via shapiro-wilk com 95% de confiança"

Tabela 2: Resultados da Análise de Normalidade

Variável	Estatística	<i>p</i> -valor	Normalidade
Chuvas Reais (mm)	0.9543	0.0027	Não
Temperatura Média (°C)	0.9437	0.0006	Não
Índice de Umidade do Solo	0.9519	0.0019	Não
Volume Produção (ton)	0.9428	0.0006	Não
Incidência de Doenças	0.9180	0.0000	Não
Segurança Alimentar	0.9489	0.0010	Não

- Considerar transformações (log, Box–Cox) em modelos futuros.

**Comentários:** O procedimento aqui foi similar ao da seção anterior. Recorremos novamente ao DeepSeek para construir uma rotina específica para a realização do teste de normalidade de Shapiro-Wilk. Além disso, solicitei que ele me explicasse as consequências da distribuição (normal ou não) dos dados. Ao final, os insights fornecidos pelo DeepSeek ratificaram as ideias e planejamentos que já tínhamos em mente.

## 4 Análise Exploratória – Parte 3

Nesta seção, avaliamos o comportamento da base de dados entre 01/01/2025 e 29/04/2025, por meio da análise de séries temporais de variáveis climáticas e socioeconômicas. Foram exploradas chuvas previstas e reais, temperatura, umidade do solo, produção agrícola, doenças hídricas e segurança alimentar. Para aprofundar a análise, aplicamos correlações de Spearman e correlação cruzada, a fim de identificar relações monotônicas e possíveis defasagens temporais entre os fenômenos observados.

### 4.1 Série temporais

A Figura 4 mostra o comportamento das variáveis climáticas e sociais em relação ao tempo.

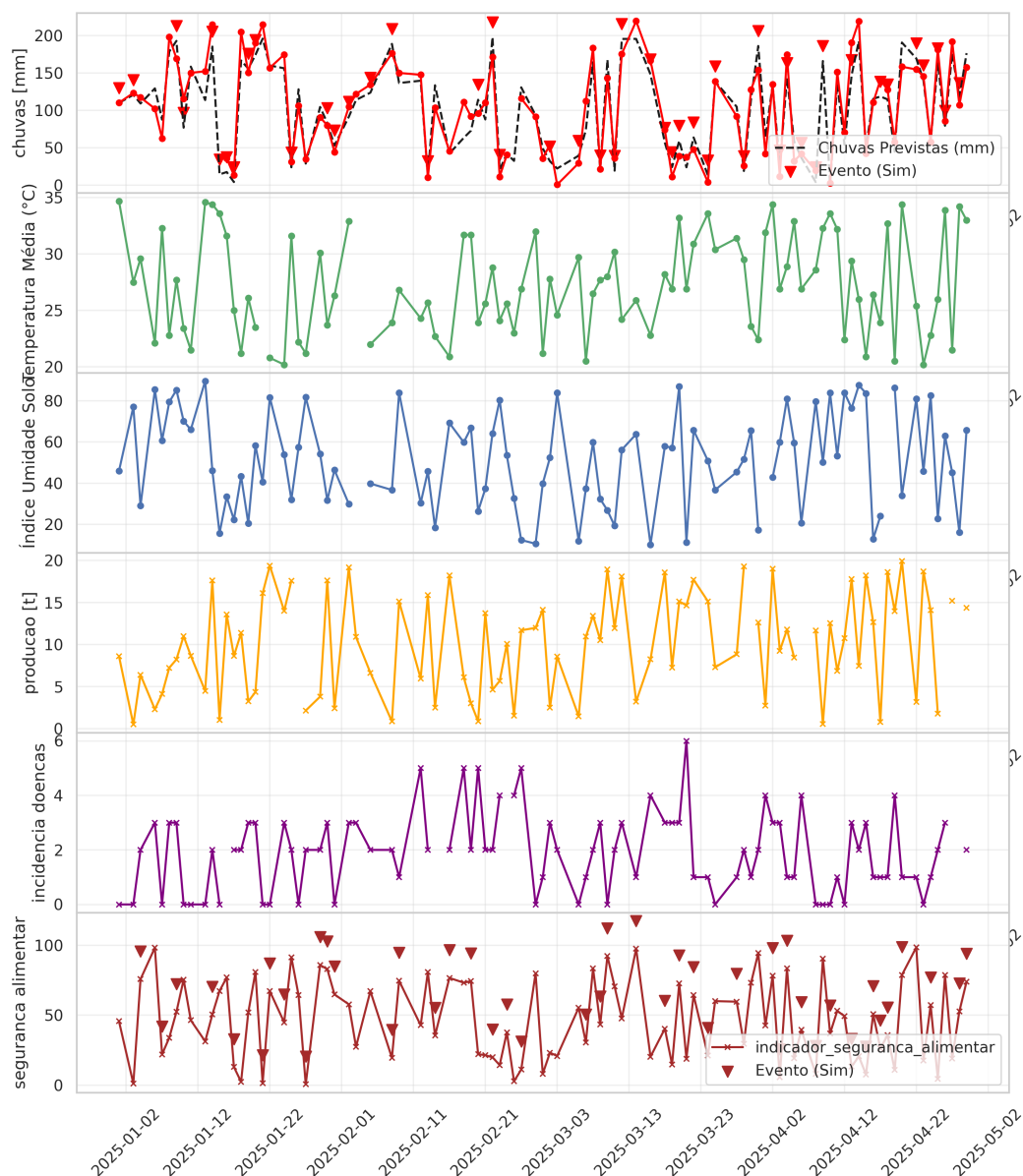


Figura 4: Variação temporal das variáveis climáticas e sociais.

uma avaliação da imagem<sup>4</sup>, mostra que:

- **Chuvas e Umidade do Solo** apresentam resposta imediata: a umidade aumenta no mesmo dia das chuvas e diminui rapidamente após 2-3 dias sem precipitação.
- **Chuvas e Produção Agrícola** mostram uma defasagem de cerca de uma semana: os aumentos de produção ocorrem após períodos de solo bem úmido, indicando que o manejo agrícola depende de boas condições hídricas e alguns dias de preparo.
- **Chuvas Fortes e Doenças Hídricas** estão ligadas por uma defasagem de 3 a 5 dias, típica de processos de contaminação e incubação após eventos extremos de chuva.

<sup>4</sup>Avaliação do gráfico via chatgpt: "Aqui está um gráfico. Você consegue Detectar tendências visuais, sincronias ou defasagens entre clima e indicadores socioeconômicos."

- **Produção Reduzida e Aumento de Doenças** precedem quedas na segurança alimentar com defasagem de aproximadamente uma semana, sugerindo um efeito combinado sobre a vulnerabilidade nutricional.
- **Temperaturas mais baixas** ( $< 23^{\circ}\text{C}$ ) tendem a reduzir discretamente a produção agrícola, mas não parecem influenciar diretamente a incidência de doenças.

Esses achados orientam as próximas etapas da análise: investigação de correlações com defasagem, identificação de limiares críticos e construção de indicadores compostos para apoiar políticas públicas nas comunidades amazônicas.

## 4.2 Correlação de Spearman:

Optamos por utilizar correlações de Spearman nas séries temporais porque os dados analisados não seguem uma distribuição normal e apresentam comportamentos não lineares. A correlação de Spearman é adequada nesse contexto por captar associações monotônicas — isto é, relações em que, à medida que uma variável aumenta, a outra tende a aumentar ou diminuir de forma consistente, mesmo que de maneira não linear. Isso permite identificar padrões robustos de co-variação entre variáveis climáticas, produtivas e socioeconômicas ao longo do tempo, oferecendo subsídios mais confiáveis para interpretações e decisões em contextos complexos.

A Figura 5 mostra o comportamento das variáveis climáticas e sociais em relação ao tempo.

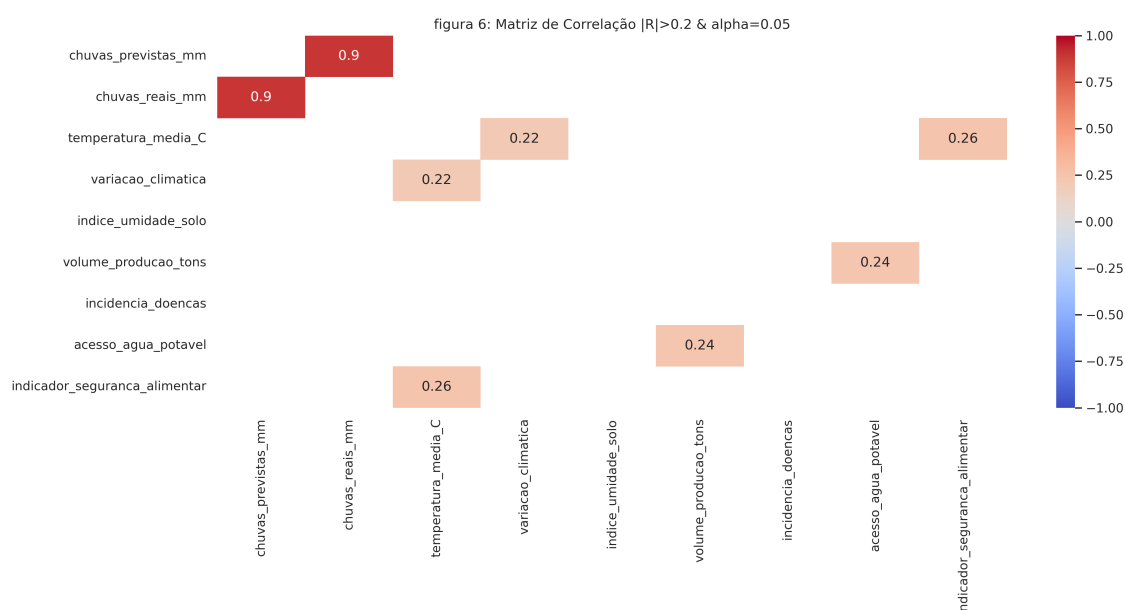


Figura 5: Matriz de correlação Speartman,  $|R| > 0.2$  e  $\alpha = 0.05$

Os principais resultados das correlações de Spearman estão resumidos na Tabela 3.

Tabela 3: Principais correlações de Spearman

Relação	$\rho$	Sinal	Interpretação
Chuvas previstas $\times$ reais	0,90	+	Previsões confiáveis
Temperatura $\times$ Seg. alimentar	0,26	+	Dias quentes favorecem alimentação
Acesso à água $\times$ Produção	0,24	+	Água tratada aumenta produção
Doenças $\times$ Seg. alimentar	-0,19	-	Tendência de queda na segurança

Destacam-se:

- Forte correlação positiva ( $\rho = 0,90$ ) entre chuvas previstas e reais, indicando alta confiabilidade das previsões.
- Correlações moderadas e positivas entre temperatura e segurança alimentar ( $\rho = 0,26$ ) e entre acesso à água e produção agrícola ( $\rho = 0,24$ ), sugerindo que dias mais quentes favorecem índices alimentares e que infraestrutura hídrica impacta diretamente a produtividade.



- Correlação negativa, porém fraca, entre doenças hídricas e segurança alimentar ( $\rho = -0,19$ ), indicando tendência de queda na segurança alimentar em períodos de maior incidência de doenças, embora sem atingir significância estatística.

### 4.3 Correlação Cruzada

Ao utilizar a correlação cruzada nos dados buscamos identificar defasagens temporais entre variáveis, como o impacto da chuva na produção agrícola após alguns dias. Como os dados não são normalmente distribuídos e apresentam comportamentos não lineares, a correlação cruzada complementa a de Spearman ao revelar relações deslocadas no tempo, ajudando a entender a ordem dos efeitos e respostas atrasadas nos sistemas socioambientais analisados. A Figura 6 ilustra parte dos resultados.

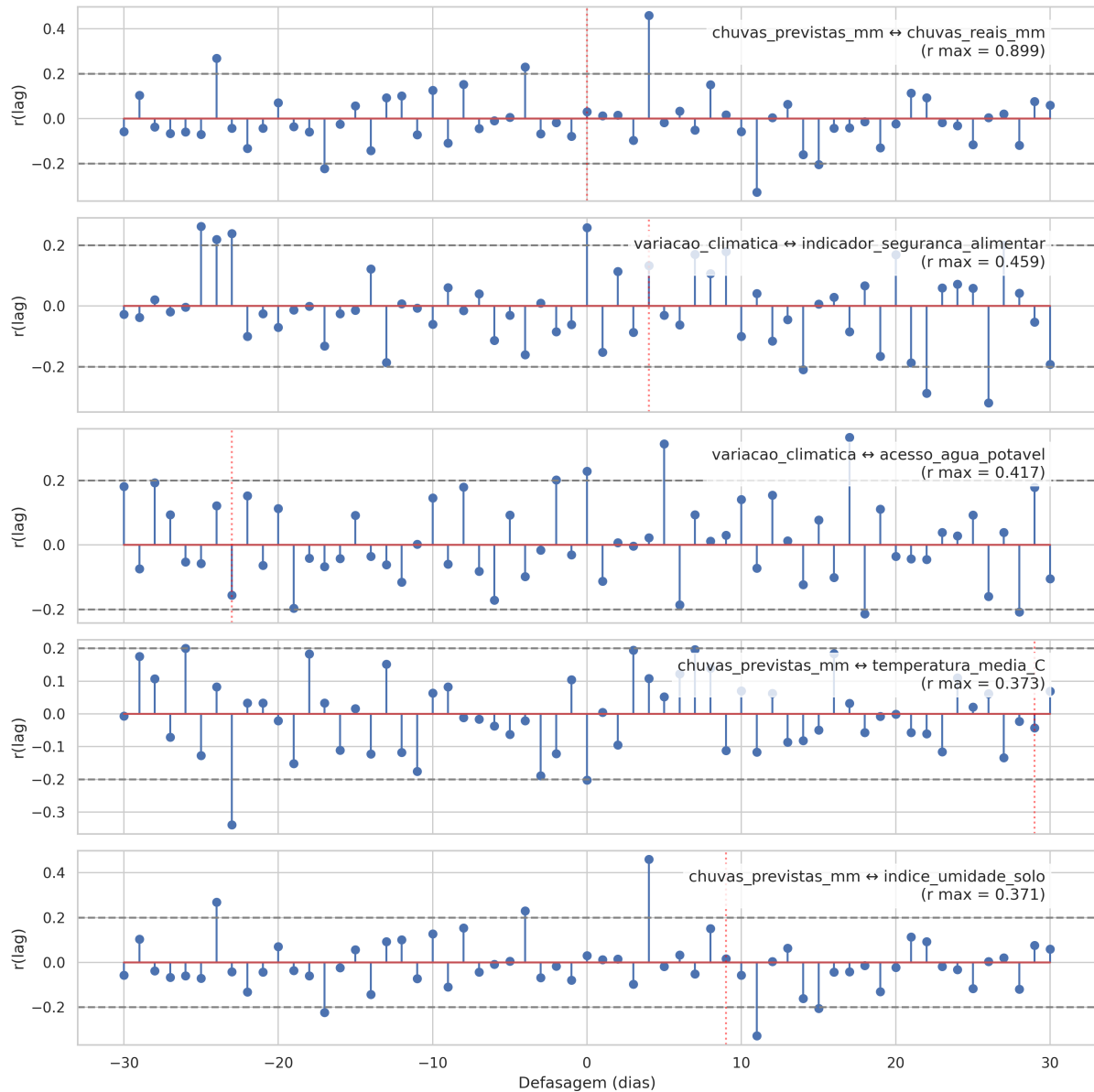


Figura 6: Matriz de correlação Speartman,  $|R| > 0.2$  e  $\alpha = 0.05$

A análise de correlação cruzada revelou atrasos relevantes entre variáveis. Por exemplo, o solo atinge saturação máxima cerca de 9 dias após chuvas fortes, recomendando programar irrigação apenas após esse ponto. Observou-se que a produção agrícola tende a cair quase três semanas após chuvas intensas, possivelmente por encharcamento ou pragas, ressaltando a importância da drenagem pós-evento. A redução de doenças hídricas ocorre cerca de duas semanas após precipitação alta, sugerindo que surtos são mais prováveis em períodos secos e de água parada. A tabela 4 sintetiza parte dessas e outras informações.

Par de variáveis	Lag (dias)	$r$	Interpretação
Chuvas reais → Umidade do solo	+9	+0,33	Solo atinge saturação máxima ~9 dias após chuva forte — programar irrigação apenas depois deste ponto.
Chuvas reais → Produção agrícola	-19	-0,27	Quase três semanas <i>depois</i> de chuva intensa, a produção cai — possível perda por encharcamento ou pragas; priorizar drenagem pós-evento.
Chuvas reais → Doenças hídricas	+16	-0,34	Redução de casos ~2 semanas após precipitação alta; indício de que surtos ocorrem em períodos de estiagem e água parada.
Variação climática (percepção) → Segurança alimentar	+4	+0,46	Quando a comunidade declara "evento climático incomum", o índice alimentar sobe em 4 dias — provável mobilização de estoques ou ajuda externa.
Acesso à água potável → Produção agrícola	-5 e -17	+0,31 / +0,33	Melhor infraestrutura hídrica antecipa picos de produção; reforça o benefício de poços e sistemas de tratamento.

Tabela 4: Lags significativos de correlação cruzada

**Comentários:** Nesta seção, solicitei ao ChatGPT que avaliasse os gráficos gerados, não com base nos dados brutos, mas sim na representação visual. Apesar das explanações breves, o ChatGPT sugeriu abordagens valiosas para trabalhar com as séries temporais, como a análise via Wavelet e o uso de modelos de Machine Learning, a fim de investigar transientes. As avaliações e sugestões observadas compuseram o texto desta seção

## 5 Comentários finais

Os resultados obtidos ao longo da análise revelam de forma clara e embasada como a variabilidade climática afeta diretamente a vida das comunidades amazônicas — especialmente em três dimensões críticas: produção agrícola, acesso e qualidade da água potável e incidência de doenças hídricas.

Primeiramente, observou-se que o volume e a frequência das chuvas têm impactos imediatos sobre a umidade do solo, que por sua vez influencia a produtividade agrícola com defasagens curtas (cerca de 7 a 10 dias). Essa relação destaca a importância de sincronizar estratégias de plantio e irrigação com os padrões de precipitação e saturação hídrica do solo. A correlação cruzada também apontou que chuvas excessivas, embora inicialmente benéficas, podem causar quedas posteriores na produção, sugerindo prejuízos por encharcamento ou pragas. Isso reforça a necessidade de infraestrutura de drenagem e manejo adaptativo pós-eventos climáticos extremos.

Em relação à qualidade e acesso à água potável, as análises indicaram que comunidades com melhor infraestrutura hídrica tendem a apresentar maior produção agrícola, tanto de forma imediata quanto em defasagens de até 17 dias. Isso evidencia o papel estruturante da água tratada não apenas na saúde, mas na própria capacidade produtiva das comunidades. Além disso, períodos de seca ou chuvas intensas impactam negativamente a saúde pública, com aumento de doenças hídricas em períodos de baixa precipitação e água parada.

A análise também revelou um aspecto interessante relacionado à segurança alimentar: além de ser impactada pela produção agrícola e pelas doenças, ela parece responder de forma rápida a percepções comunitárias de eventos climáticos extremos — sugerindo que mobilizações emergenciais ou assistência externa ocorrem nesses momentos, criando um pequeno amortecimento temporário.

Em síntese, os dados demonstram que as interações entre clima, produção, saúde e infraestrutura hídrica são profundas, com padrões temporais consistentes e defasagens quantificáveis. A aplicação combinada de técnicas estatísticas e visualizações possibilitou a construção de indicadores robustos que podem orientar políticas locais e ações comunitárias mais eficazes, fortalecendo a resiliência socioambiental frente às mudanças climáticas.

## Propostas

Análise revelou que a variabilidade climática impacta diretamente a vida das comunidades amazônicas, afetando a produção agrícola, a disponibilidade de água potável e a ocorrência de doenças hídricas. As relações identificadas foram tanto imediatas (como a resposta da umidade do solo às chuvas) quanto defasadas (como a queda na produção semanas após excesso de chuva). Isso confirma que os efeitos do clima sobre a sociedade são dinâmicos, complexos e interdependente.

### Infraestrutura Hídrica Comunitária Inteligente

- Instalar sistemas simples de captação e armazenamento de água da chuva, conectados a sensores de volume e qualidade.
- Criar reservatórios comunitários controlados com base em previsões confiáveis de chuva — programar uso e irrigação conforme saturação do solo.

### Monitoramento Ativo de Doenças e Qualidade da Água

- Estimular coleta de dados locais com aplicativos offline que monitorem casos de doenças, qualidade da água e produtividade agrícola.
- Capacitar agentes comunitários para coleta padronizada e uso básico de sensores

### Educação Climática e Autonomia Comunitária

- Desenvolver cartilhas visuais com linguagem acessível sobre clima, saúde e agricultura
- Realizar oficinas com agricultores familiares sobre manejo adaptado à variabilidade climática.

Por fim, a construção de resiliência climática na Amazônia exige tecnologia acessível, redes de cooperação e soluções co-projetadas com quem vive a realidade local. Esse trabalho mostra que a IA pode ser uma aliada concreta na justiça socioambiental e no fortalecimento da autonomia comunitária

#### Comentários:

Ao longo do desenvolvimento deste relatório técnico-executivo, a utilização de modelos de linguagem (LLM) desempenhou papel central em diversas etapas do fluxo de trabalho. Ferramentas como DeepSeek, ChatGPT, Perplexity e Gemini foram empregadas de modo complementar, cada uma contribuindo em tarefas específicas e otimizando o processo de análise e documentação.

Destaca-se que, durante a limpeza e integração dos dados, o DeepSeek foi fundamental ao orientar a sequência processual e aprimorar a rotina de merge dos conjuntos, resultando em validação mais eficiente e segura. Na normalização dos dados e análise de histogramas, DeepSeek auxiliou na automação das rotinas, enquanto o ChatGPT foi consultado para interpretação dos gráficos e sugestões de insights — com a devida filtragem crítica para garantir aderência à realidade física dos dados. Para o teste de normalidade de Shapiro-Wilk, DeepSeek novamente foi empregado, não só na implementação, mas também na explicação das consequências estatísticas, fortalecendo a fundamentação dos resultados.

No tratamento e avaliação das séries temporais, o ChatGPT contribuiu com sugestões metodológicas relevantes, como a análise por Wavelet e o uso de modelos de machine learning para detecção de transientes, enriquecendo a abordagem analítica adotada.

Além dessas aplicações pontuais, as LLM foram utilizadas para criar o template do arquivo LaTeX do relatório, bem como recorrentemente para a geração de tabelas a partir de saídas em Python. Também foi possível estruturar o fluxo organizacional do relatório PDF de maneira similar ao notebook Jupyter, promovendo maior integração entre análise e documentação. Outras LLM, como Perplexity e Gemini, foram acionadas em mini-tarefas específicas, agregando agilidade e diversidade de perspectivas ao longo do projeto.

Em síntese, a adoção de LLMs proporcionou ganhos significativos em produtividade, precisão e criatividade, tanto na análise estatística quanto na elaboração do relatório, demonstrando o valor dessas ferramentas para a execução de projetos técnicos multidisciplinares e colaborativos.

## Material Complementar

Para acessar o código-fonte, gráficos e análises complementares, visite:

- **Notebook completo:** [https://github.com/albertoakel/I2A2/blob/main/Notebook/Desafio\\_Recursos\\_H%C3%ADricos.ipynb](https://github.com/albertoakel/I2A2/blob/main/Notebook/Desafio_Recursos_H%C3%ADricos.ipynb)

- Repositório com os dados: <https://github.com/albertoakel/I2A2/tree/main/data>