

# Plataforma Arduino: Sensores Aplicados na Estatística

Jefferson Bezerra dos Santos

## Resumo

Este artigo explora a aplicação de sensores de Arduino na coleta e análise de dados estatísticos. Por meio de exemplos práticos, demonstramos como essas tecnologias podem automatizar a coleta de dados e aprimorar análises estatísticas em tempo real. O artigo também discute limitações e aplicações futuras, com ênfase em IoT, aprendizado de máquina e técnicas de análise de dados e suas implementações em microcontroladores. Resultados mostram que a integração entre hardware acessível e métodos estatísticos oferecem novas oportunidades em diversas áreas como: agricultura, saúde e monitoramento ambiental.

## 1 Introdução

A estatística desempenha um papel central na interpretação de dados em cenários complexos, como monitoramento ambiental e automação industrial. Sensores baseados em Arduino, combinados com técnicas de Internet das Coisas (IoT), permitem coletar dados em tempo real e aplicá-los a modelos estatísticos preditivos [5]. A análise preditiva é fundamental para compreender as relações entre diversos fatores ambientais, como temperatura, umidade e qualidade do ar, e seu impacto na produtividade agrícola.

Nesse contexto, a agricultura inteligente surge como um exemplo emblemático, em que sensores de baixo custo monitoram parâmetros como umidade do solo e temperatura, gerando dados para otimização de recursos e permitindo a integração com áreas como robótica e automação para alcançar resultados favoráveis [4]. Além disso, a integração de aprendizado de máquina com dados de sensores IoT amplia a capacidade de análise estatística, proporcionando insights mais precisos e automatizados [1].

O tratamento adequado dos dados e a escolha criteriosa dos métodos estatísticos são essenciais para garantir a precisão e a eficácia das previsões. Segundo [7], uma análise cuidadosa dos dados e dos métodos utilizados é

crucial para resultados confiáveis. Modelos tradicionais, como ARIMA e SARIMA, são eficazes em muitos cenários, mas apresentam limitações ao lidar com dados que possuem relações não lineares e sazonalidade. Para superar essas limitações, modelos de aprendizado profundo, como LSTM e o híbrido CNN-LSTM, têm sido empregados, demonstrando desempenho superior em contextos específicos.

Dessa forma, o presente artigo explora a aplicação de técnicas estatísticas e sua implementação em microcontroladores da plataforma Arduino, com foco na análise de dados em tempo real e na automação de sistemas robóticos em ambientes dinâmicos. O estudo tem como objetivo contribuir para a otimização de processos e o aprimoramento da precisão de previsões, destacando aplicações práticas em áreas como agricultura inteligente e sistemas IoT.

## 2 Revisão da Literatura

A análise estatística de dados provenientes de sensores IoT tem ganhado relevância em aplicações industriais e ambientais, impulsionada pela necessidade de processamento eficiente de grandes volumes de dados e pela demanda por insights preditivos precisos. Estudos como o de [2] demonstram o uso de técnicas estatísticas descritivas, como regressão linear e análise de séries temporais, combinadas com métodos de aprendizado de máquina, como *Random Forest*, para o monitoramento preditivo em ambientes industriais. Essas abordagens têm se mostrado eficazes na identificação de padrões complexos e na previsão de falhas, otimizando processos e reduzindo custos.

A rápida expansão da Internet das Coisas (IoT) e da Internet Industrial das Coisas (IIoT) tem resultado na geração massiva de dados, o que impõe desafios significativos às indústrias, especialmente no que diz respeito ao processamento em tempo real. Algoritmos tradicionais frequentemente apresentam limitações ao lidar com a complexidade e o volume desses dados. Diante disso, [2] propõem a utilização de técnicas robustas e inteligentes para o tratamento preditivo de dados, destacando métodos como:

- **Regressão Linear:** Técnica que analisa a relação entre variáveis, permitindo prever valores dependentes com base em variáveis independentes.
- **Random Forest:** Modelo de aprendizado de máquina baseado em múltiplas árvores de decisão, que aumenta a precisão das previsões e é eficaz na identificação de padrões complexos.

- **Deteção de Outliers:** Conjunto de técnicas utilizadas para identificar e tratar dados que se desviam significativamente do padrão esperado, melhorando a qualidade da análise.
- **Análise de Séries Temporais:** Métodos aplicados à análise de dados ao longo do tempo, permitindo identificar tendências, ciclos e padrões para embasar decisões operacionais e de manutenção.

De forma complementar, [8] aplicam técnicas de *machine learning* para monitorar processos de manufatura em tempo real. O estudo propõe um sistema de monitoramento baseado em sensores IoT, processamento de grandes volumes de dados (*big data*) e um modelo de predição híbrido, que combina DBSCAN para detecção de outliers e *Random Forest* para classificação de falhas. Esse sistema foi implementado em uma linha de montagem automotiva na Coreia, demonstrando eficácia na prevenção de perdas e na melhoria da tomada de decisão. Para o processamento de dados, foram utilizadas ferramentas como Apache Kafka, Apache Storm e MongoDB, que garantem escalabilidade e eficiência no tratamento de grandes volumes de informações.

No contexto agrícola, [4] revisam sistemas IoT que utilizam microcontroladores Arduino para coleta de dados, destacando a importância de métodos estatísticos robustos. A pesquisa evidencia o impacto transformador da IoT na agricultura inteligente, com a integração de sensores e microcontroladores permitindo o monitoramento em tempo real de condições ambientais e a otimização de práticas agrícolas. Foram explorados protocolos de comunicação como LoRa, ZigBee e WiFi, que garantem a transmissão eficaz de dados mesmo em áreas rurais com conectividade limitada. Entre os benefícios destacados estão a economia de água e insumos, além da promoção de práticas mais sustentáveis. No entanto, os autores também apontam desafios, como os altos custos de implementação e a necessidade de garantir segurança e privacidade dos dados coletados. A colaboração entre setores como tecnologia da informação, ciência ambiental e políticas públicas é apontada como essencial para superar essas barreiras e facilitar a adoção em larga escala da IoT na agricultura.

Em 2023, [6] analisaram a penetração de dispositivos IoT em residências na União Europeia, destacando o aumento do uso desses dispositivos entre 2020 e 2022. O estudo, baseado em dados do Eurostat, identifica que os principais fatores que dificultam a adoção da IoT são a "falta de necessidade" e o "custo dos dispositivos", cuja influência diminuiu ao longo do tempo. Os resultados evidenciam diferenças significativas na adoção de IoT entre os Estados-Membros, sugerindo a necessidade de estratégias e políticas públicas direcionadas para estimular a adoção em categorias específicas e promover as vantagens econômicas e de eficiência que a tecnologia pode oferecer.

No campo da análise preditiva, [7] propõem uma nova estrutura focada na previsão de séries temporais, especialmente de temperatura, utilizando dados de sensores IoT. O estudo compara modelos tradicionais, como ARIMA e SARIMA, com abordagens baseadas em aprendizado profundo, como LSTM e um modelo híbrido CNN-LSTM. Os resultados demonstram que os modelos de aprendizado profundo, em especial o CNN-LSTM, apresentam desempenho superior, com redução significativa nas taxas de erro e maior capacidade de capturar sazonalidade e tendências nas séries temporais. Para o processamento de dados, foram utilizadas tecnologias de código aberto, como Apache Spark e Kafka, que garantem escalabilidade e eficiência. No entanto, o estudo também enfrentou desafios, como a complexidade dos dados provenientes de múltiplos sensores e a demanda por recursos computacionais significativos para o ajuste e validação dos modelos.

Por fim, [5] apresentam um sistema inovador de monitoramento ambiental em tempo real baseado em IoT, que utiliza uma rede diversificada de sensores para coletar dados sobre parâmetros como qualidade do ar, temperatura, umidade, pH da água e turbidez. Entre os sensores utilizados, destacam-se:

- **Sensor de Temperatura e Umidade (DHT11):** Fornece dados precisos para identificar mudanças climáticas e condições ambientais.
- **Sensores de Velocidade e Direção do Vento:** Monitoram condições do vento, importantes para a previsão de incêndios florestais.
- **Sensores de Precipitação:** Incluem pluviômetros e sensores de gotas de chuva, vitais para prever inundações e monitorar níveis hídricos.

A integração desses sensores em um sistema IoT permite acesso remoto e em tempo real aos dados, além de fomentar a conscientização pública sobre questões ambientais. No entanto, o sistema enfrenta desafios relacionados à conectividade, segurança, privacidade e escalabilidade, que precisam ser superados para garantir sua eficácia em diferentes contextos.

### 3 Metodologia

Nessa seção é apresentada metodologia para a integração de sensores com a plataforma Arduino, aplicando técnicas estatísticas e matemáticas para análise de dados. A abordagem é voltada para aplicações práticas, com embasamento teórico robusto, visando a coleta, processamento e análise de dados de forma criteriosa. A metodologia é dividida nas seguintes etapas:

## 3.1 Definição do Problema

### 3.1.1 Objetivos

O objetivo principal é explorar a utilização de sensores em conjunto com técnicas estatísticas e matemáticas para aplicações em cenários práticos, como monitoramento ambiental, automação residencial ou controle de processos envolvendo IoT. Os objetivos específicos incluem:

### 3.1.2 Os objetivos específicos

Os objetivos específicos incluem:

- Coletar dados de variáveis físicas (temperatura, umidade, qualidade do ar, etc.) por meio de sensores conectados à plataforma Arduino.
- Aplicar técnicas estatísticas descritivas e inferenciais para análise dos dados.
- Utilizar modelos matemáticos e estatísticos para previsão e tomada de decisão.
- Integrar o Arduino com Python para comunicação e controle de sensores com base nos resultados das análises.
- Validar a eficácia do sistema em cenários práticos, com embasamento matemático e estatístico.

## 3.2 Materiais e Equipamentos

Para a execução do projeto, foram utilizados os seguintes materiais e equipamentos:

- **Plataforma Arduino:** Microcontrolador Arduino Uno ou similar.
- **Sensores:** DHT11 (temperatura e umidade do ar) e HD-38 (umidade do solo).
- **Componentes eletrônicos:** Resistores, jumpers, protoboard, módulo sdcard.
- **Software:**
  - **Arduino-cli:** Para programação do microcontrolador e automação.
  - **Python:** Para análise de dados e estatística.

- **Bibliotecas Python:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PySerial.

### 3.3 Arquitetura do Sistema

O sistema foi implementado utilizando uma placa Arduino com sensor DHT11 para coleta de dados de temperatura e umidade do ar conjuntamente com um sensor HD-38 de umidade do solo. A Figura 1 apresenta a arquitetura geral do sistema.

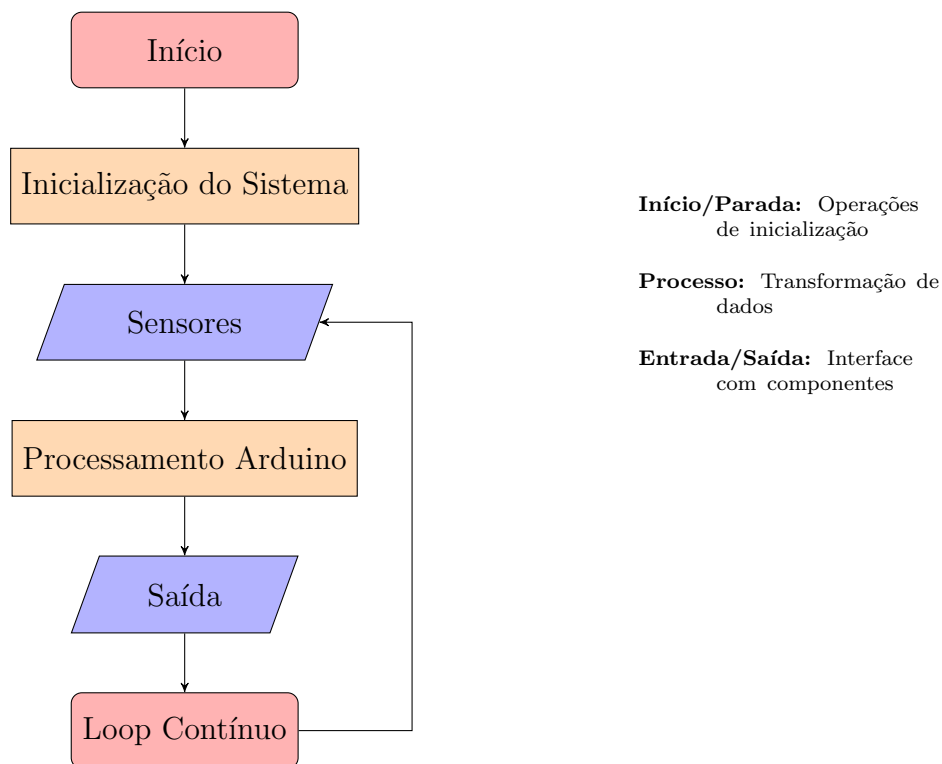


Figura 1: Diagrama de blocos do sistema de aquisição de dados

O diagrama de blocos apresenta a cadeia operacional do sistema, compreendendo: (1) ativação dos sensores para captura de dados ambientais, (2) condicionamento do sinal pela plataforma Arduino, e (3) transmissão dos parâmetros processados via comunicação serial para monitoramento e escrita em arquivo para processamento dos dados. Na sequência, apresenta-se a estrutura lógica do algoritmo desenvolvido para coleta e processamento dos dados.

### 3.4 Fluxo de Controle

Apresenta-se abaixo o diagrama de fluxo de controle do sistema, que descreve graficamente o algoritmo de aquisição e processamento dos dados ambientais.

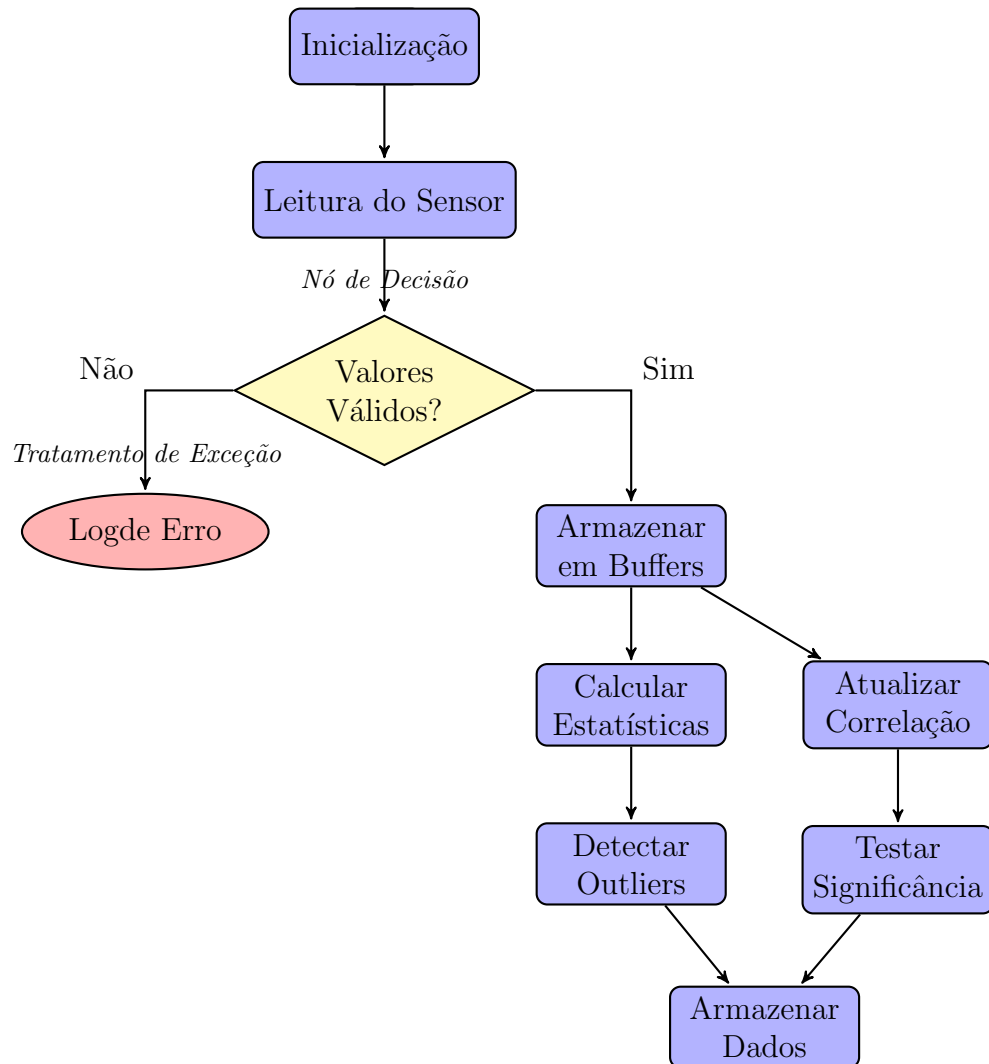


Figura 2: Fluxograma do processo de análise estatística com tratamento de erros e validação de dados

Conforme ilustrado na Figura 2, o processo de aquisição de dados inicia-se com a coleta simultânea dos parâmetros de temperatura e umidade relativa do ar. Esses dados brutos são então submetidos a uma estrutura de controle de qualidade, onde sua validade é verificada. Após a validação bem-sucedida, o algoritmo prossegue com o fluxo padrão de processamento e transmissão

dos dados

### 3.5 Coleta de Dados

A coleta de dados será realizada por meio de sensores conectados à plataforma Arduino, seguindo as seguintes etapas:

- **Configuração dos Sensores:** Montagem do circuito e calibração dos sensores.
- **Programação do Arduino:** Desenvolvimento de código para leitura dos dados em intervalos de tempo definidos.
- **Comunicação com Python:** Utilização da biblioteca **PySerial** para estabelecer comunicação serial entre o Arduino e o Python, permitindo a transmissão dos dados coletados para o ambiente Python em tempo real.
- **Armazenamento dos Dados:** Os dados foram armazenados localmente em arquivos CSV.

### 3.6 Análise Estatística

Os dados coletados foram processados e analisados utilizando Python e suas bibliotecas, com embasamento matemático e estatístico detalhado:

#### 3.6.1 Processamento

Um **outlier** é uma observação atípica que se distâcia significativamente das demais amostras em um conjunto de dados. Formalmente, para uma variável aleatória  $X$ , um valor  $x_i$  é considerado outlier se:

$$\text{Outlier} = \begin{cases} x_i < Q_1 - k \cdot IQR \\ x_i > Q_3 + k \cdot IQR \end{cases} \quad (1)$$

onde:

- $Q_1$  é o primeiro quartil (percentil 25)
- $Q_3$  é o terceiro quartil (percentil 75)
- $IQR = Q_3 - Q_1$  é o intervalo interquartil



- $k$  é uma constante (tipicamente 1.5 para outliers moderados ou 3.0 para outliers extremos)

Sua utilização torna o sistema mais robusto, uma vez que valores discrepantes (outliers) podem surgir devido a erros de medição dos sensores causados por problemas de hardware. Dessa forma, a remoção, correção ou tratamento desses outliers garante que a análise estatística seja realizada sem a influência de dados distorcidos.

### 3.6.2 Medidas de Tendência Central

Para um conjunto de dados  $\{x_1, x_2, \dots, x_n\}$  amostrados pelo sensor, definimos:

- **Média aritmética:**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

- **Mediana:**

$$Q_2 = \begin{cases} x_{(\frac{n+1}{2})} & \text{se } n \text{ ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{se } n \text{ par} \end{cases} \quad (3)$$

### 3.6.3 Medidas de Dispersão

A variabilidade dos dados é quantificada por:

- **Variância:**

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

- **Desvio padrão:**

$$\sigma = \sqrt{\sigma^2} \quad (5)$$

- **Intervalo Interquartil (IQR):**

$$\text{IQR} = Q_3 - Q_1 \quad (6)$$

### 3.6.4 Justificativa para o Projeto

A análise descritiva é fundamental porque:

1. A **comparação média-mediana** identifica assimetrias:

$$\Delta = |\bar{x} - Q_2| \quad (7)$$

Valores  $\Delta > 0.1s$  indicam presença de outliers.

2. O **desvio padrão**  $s$  quantifica a precisão do sensor:

$$\text{Precisão} \propto \frac{1}{s} \quad (8)$$

3. O **IQR** define limites operacionais robustos:

$$\text{Limites} = [Q_1 - 1.5 \times \text{IQR}, Q_3 + 1.5 \times \text{IQR}] \quad (9)$$

essenciais para detecção de falhas.

Tabela 1: Métricas utilizadas em sensores IoT

Métrica	Relevância
Média	Utilizada para o monitoramento contínuo dos dados coletados pelos sensores.
Mediana	Aplicada no controle dos sensores, reduzindo o impacto de valores atípicos.
IQR (Intervalo Inter-quartil)	Empregado na detecção de anomalias, identificando variações inesperadas nos dados.
Desvio padrão ( $s$ )	Essencial para a calibração do sensor, permitindo avaliar a dispersão dos valores medidos.

Fonte: Elaborado pelo autor.

## 3.7 Análise de Correlação Linear para Sistemas Embarcados

A análise de dependência estocástica entre as variáveis ambientais coletadas (temperatura  $T$  e umidade relativa  $U$ ) foi realizada mediante o coeficiente de correlação momento-produto de Pearson, cuja forma populacional é dada por:

$$\rho_{TU} = \frac{\mathbb{E}[(T - \mu_T)(U - \mu_U)]}{\sigma_T \sigma_U} = \frac{\text{Cov}(T, U)}{\sqrt{\text{Var}(T)\text{Var}(U)}} \quad (10)$$

Para amostras finitas, utilizamos o estimador não-tendencioso:

$$r_{TU} = \frac{\sum_{i=1}^n (T_i - \bar{T})(U_i - \bar{U})}{\sqrt{\sum_{i=1}^n (T_i - \bar{T})^2 \sum_{i=1}^n (U_i - \bar{U})^2}} \quad (11)$$

### 3.7.1 Justificativa Matemática para Implementação em Arduino

- **Eficiência Computacional:** A equação (11) requer apenas:
  - $4n$  operações de soma
  - $2n$  multiplicações
  - 1 divisão e 1 raiz quadrada
- **Otimização de Memória:** Pode ser calculada em streaming, armazenando apenas:

$$S_T = \sum T_i, S_U = \sum U_i, S_{TU} = \sum T_i U_i, S_{T^2} = \sum T_i^2, S_{U^2} = \sum U_i^2 \quad (12)$$

- **Precisão Numérica:** A fórmula é numericamente estável para valores típicos de sensores ( $-40C \leq T \leq 125C$ ,  $0\% \leq U \leq 100\%$ )

### 3.7.2 Teste de Hipótese com Restrições Computacionais

Para verificação da significância estatística em sistemas embarcados, utilizamos a transformação de Fisher com aproximação para pequenas amostras:

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \sim \mathcal{N} \left( \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right), \frac{1}{\sqrt{n-3}} \right) \quad (13)$$

Permitindo calcular intervalos de confiança assintóticos mesmo com  $n < 30$ .

### 3.7.3 Interpretação Físico-Estatística para Sensores IoT

### 3.7.4 Implementação Otimizada

O algoritmo no Arduino segue:

1. Amostragem em janelas móveis de  $n = 30$  pontos (balanceando precisão e latência)

Tabela 2: Interpretação de Correlações em Ambientes Controlados

$\rho$	Interpretação Física	Ação no Arduino
$0.9 \leq \rho \leq 1$	Acoplamento termodinâmico perfeito	Reduzir taxa de amostragem em 50%
$-0.9 \leq \rho \leq -0.7$	Resfriamento evaporativo ativo	Ativar ventilação forçada
$ \rho  < 0.2$	Falha de sensor ou ambiente anômalo	Acionar diagnóstico de hardware

2. Cálculo recursivo das somas acumuladas
3. Atualização da correlação a cada 5 novas amostras
4. Limiarização adaptativa baseada em histórico

---

**Algorithm 1** Cálculo Recursivo de  $\rho$  para Arduino

---

```

1: Inicializar  $S_T, S_U, S_{TU}, S_{T^2}, S_{U^2} \leftarrow 0$ 
2: for cada nova amostra  $(T_i, U_i)$  do
3:   Atualizar somas acumuladas
4:   if  $i \bmod 5 = 0$  then
5:     Calcular  $r \leftarrow \frac{nS_{TU} - S_T S_U}{\sqrt{(nS_{T^2} - S_T^2)(nS_{U^2} - S_U^2)}}$ 
6:     Testar  $H_0 : \rho = 0$  via equação (13).
7:   end if
8: end for

```

---

### 3.8 Validação e Testes

Para garantir a confiabilidade do sistema, serão realizados testes e validações:

- **Testes de Consistência:** Repetição das medições para verificar a precisão dos sensores.
- **Ajustes no Modelo:** Refinamento dos algoritmos estatísticos e de aprendizado de máquina com base no desempenho observado.
- **Testes em Cenários Reais:** Implementação do sistema em ambientes controlados (ex.: estufa agrícola, linha de produção industrial) para avaliar sua eficácia prática.

### 3.9 Implementação do Sistema de Coleta de Dados

O sistema de coleta e análise de dados foi implementado em uma arquitetura embarcada composta por um microcontrolador Arduino Nano integrado a sensores ambientais e um módulo de armazenamento.

#### 3.9.1 Arquitetura do Sistema

O sistema possui três componentes principais:

- **Camada de Aquisição:** Sensor DHT11 para medição contínua de temperatura ( $T$ ) e umidade relativa do ar ( $U_{ar}$ ). Sensor HD-38 para medição contínua da umidade do solo ( $U_s$ ).
- **Camada de Processamento:** Microcontrolador ATmega328P com algoritmo estatístico implementado.
- **Camada de Armazenamento:** Módulo SD Card para registro persistente dos dados.

#### 3.9.2 Algoritmo de Coleta e Processamento

O processo de coleta segue um pipeline estatístico em tempo real:

1. **Amostragem:** Leituras periódicas com intervalo  $\Delta t = 5$  min.
2. **Pré-processamento:** Validação de dados através da função:

$$\text{valid}(T, U) = \begin{cases} 1 & \text{se } T \in [-40, 125]^\circ\text{C e } U \in [0, 100]\% \\ 0 & \text{caso contrário} \end{cases} \quad (14)$$

3. **Buffer Circular:** Armazenamento das últimas  $N = 12$  amostras para cálculo das estatísticas descritivas.

#### 3.9.3 Análise Estatística Implementada

O sistema calcula em tempo real:

- **Estatísticas Descritivas:**
  - Média;
  - Mediana;
  - Desvio-padrão;

– Variância.

- **Detecção de Outliers**

– Método IQR.

### 3.9.4 Armazenamento de Dados

Os dados são registrados em arquivo CSV com a seguinte estrutura:

`Data,Hora,Temperatura,Umidade,MediaT,DesvioT,...,OutlierT,Correlacao`

A Tabela 3 detalha os campos armazenados:

Tabela 3: Formato do arquivo de dados

Campo	Descrição
Data	Data da coleta (YYYY-MM-DD)
Hora	Horário da medição (HH:MM:SS)
Temperatura	Valor bruto em °C
MediaT	Média móvel ( $T$ )
IQR	Intervalo interquartil
OutlierT	Indicador booleano de anomalia
Correlacao	Coefficiente $\rho_{T,U}$ de Pearson

### 3.9.5 Otimizações Implementadas

Para operação em hardware limitado:

- **Bubble Sort** otimizado para conjuntos pequenos ( $n \leq 12$ );
- Cálculo incremental de médias e variâncias;
- Buffer circular para economia de memória RAM;
- Amostragem adaptativa baseada em variação estatística.

## 4 Resultados

Os dados de monitoramento ambiental coletados em 23 de abril de 2025 revelaram padrões distintos nas variáveis analisadas. A Figura 6 mostra a variação conjunta das principais variáveis ao longo do período monitorado.

## 4.1 Variação Temporal das Variáveis

A seguir é feita uma análise das variáveis em relação ao tempo.

### 4.1.1 Umidade do Ar

A análise da Figura 3 revela que a umidade do ar apresentou:

- **Variação:** 85% a 90%
- **Tendência:**
  - Flutuação moderada, mantendo-se majoritariamente entre 89%-90%;
  - Picos máximos de 90% em vários intervalos durante a madrugada e manhã;
  - Reduções pontuais para 85% associadas a variações de temperatura.

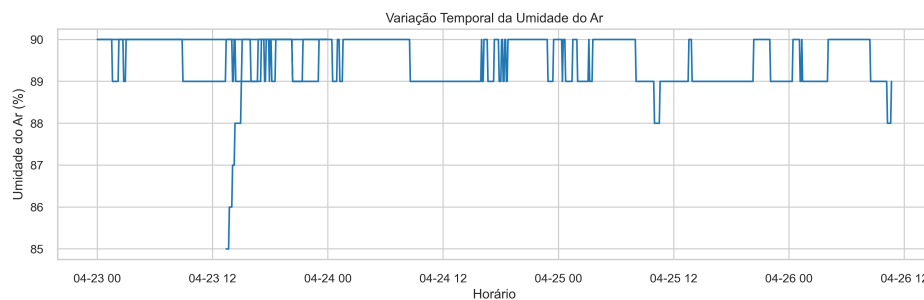


Figura 3: Variação temporal conjunta de umidade do ar.

### 4.1.2 Temperatura

Como visível na Figura 4, a temperatura exibiu:

- **Variação:** 26,9°C a 29,7°C
- **Tendência:**
  - Elevação progressiva desde 26,9°C nas primeiras horas para 29,7°C no início da tarde;
  - Estabilização temporária entre 28,0°C e 29,0°C durante o período noturno.

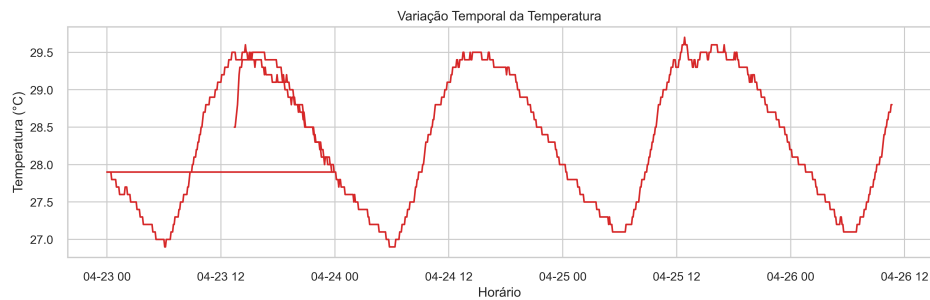


Figura 4: Variação temporal da temperatura do ar.

#### 4.1.3 Umidade do Solo

A umidade do solo apresentou uma variação mais restrita em comparação às demais variáveis ambientais. As análises revelam:

- **Intervalo de variação:** 73% a 78%
- **Média:** 75,53%
- **Desvio padrão:** 0,94
- **Mediana:** 76%

Como ilustrado na Figura 7, a distribuição dos valores é relativamente concentrada, indicando baixa variabilidade. A análise de outliers baseada no intervalo interquartil (IQR) identificou alguns valores extremos próximos de 78%, sugerindo eventos localizados de aumento da umidade do solo.

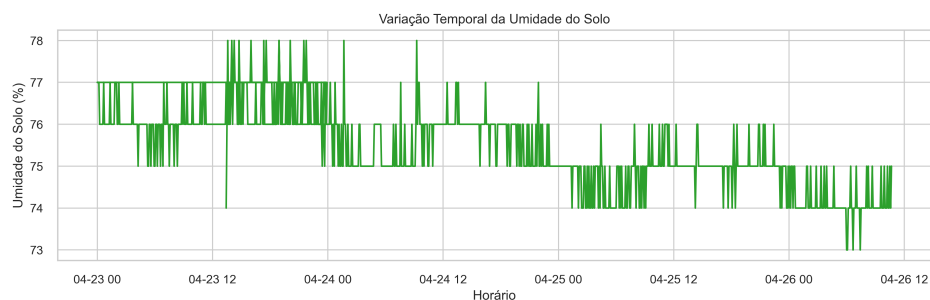


Figura 5: Variação temporal da umidade do solo.

A Figura 6 apresenta a representação gráfica combinada das três variáveis monitoradas — temperatura, umidade do solo e umidade do ar — permitindo



uma análise integrada de sua variação temporal ao longo do período observado.

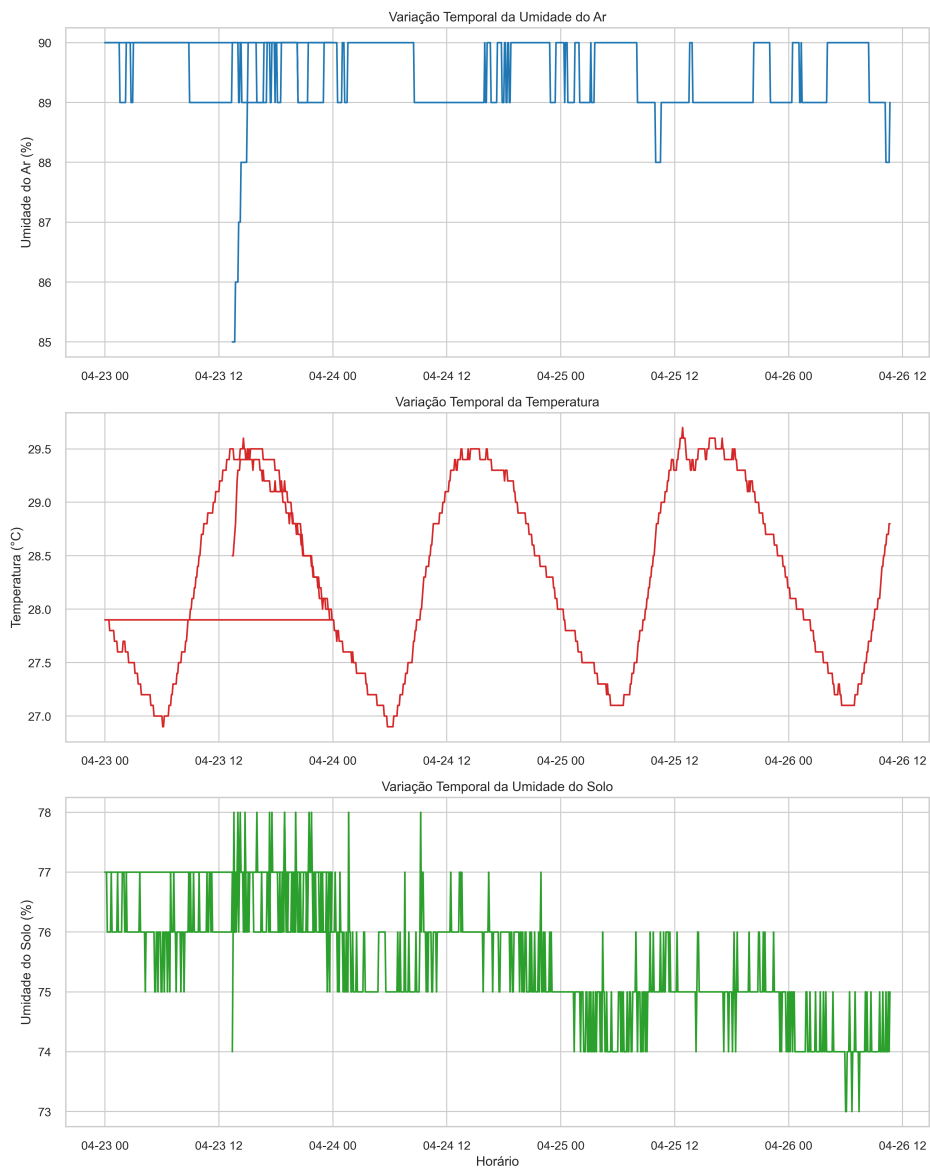


Figura 6: Variação temporal conjunta de umidade do ar, temperatura e umidade do solo.

## 4.2 Distribuição dos Dados

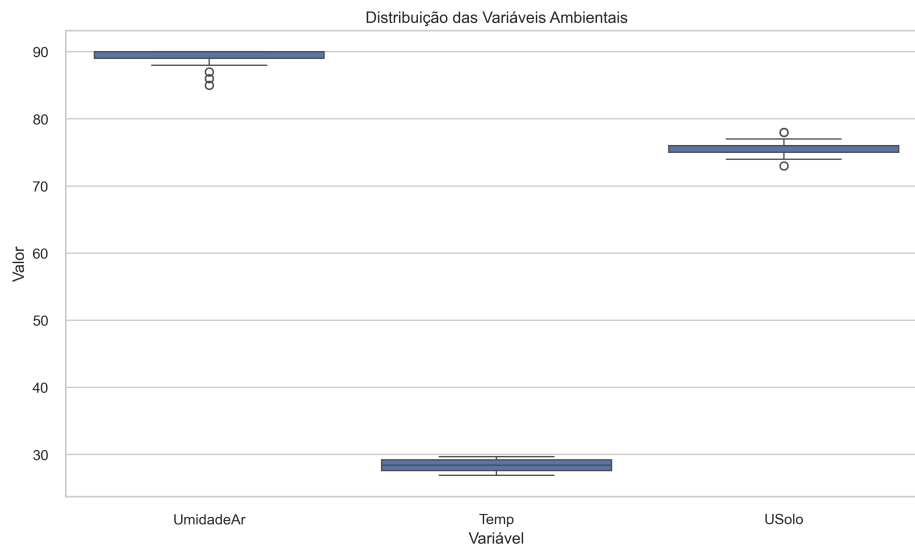


Figura 7: Distribuição estatística das variáveis ambientais

A Figura 7 mostra a distribuição das variáveis, revelando:

- Maior variabilidade na temperatura;
- Distribuição mais concentrada da umidade do solo;
- Leve assimetria nos dados de umidade do ar.

### 4.3 Relações entre Variáveis

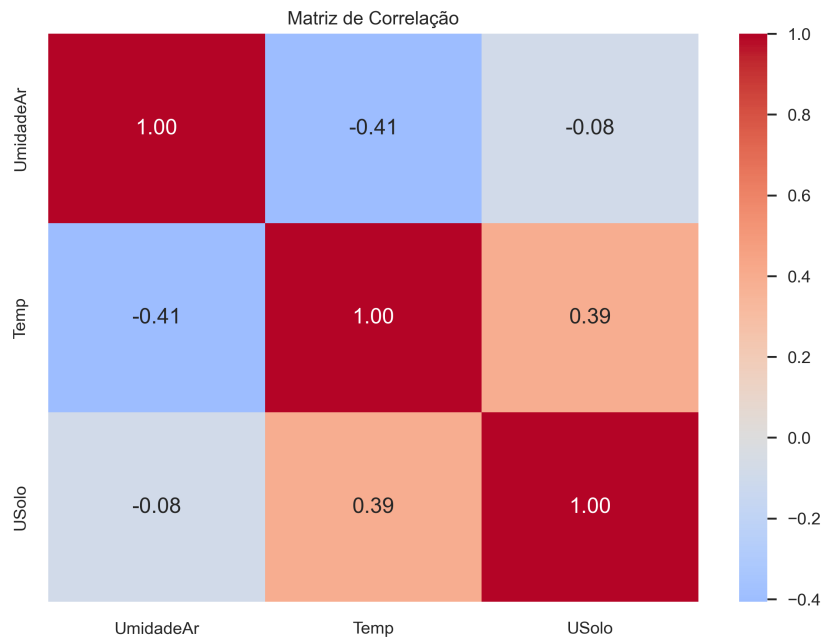


Figura 8: Matriz de correlação entre as variáveis ambientais

A análise de correlação (Figura 8) destaca:

- Correlação negativa moderada entre temperatura e umidade do ar (-0,41);
- Fraca relação entre umidade do solo e as demais variáveis.

## 4.4 Análise de Outliers

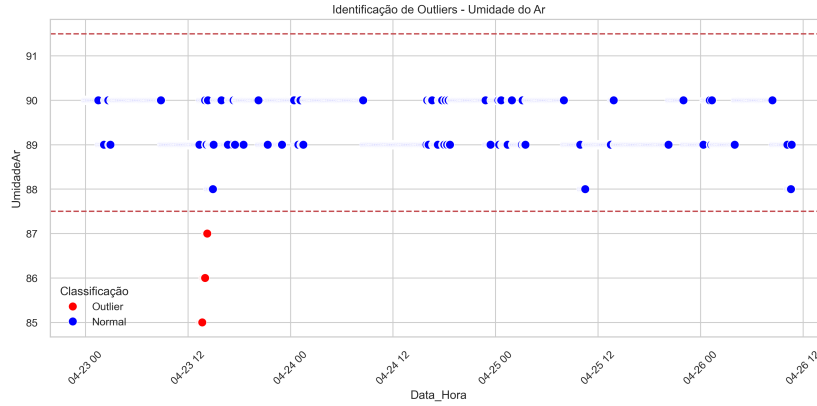


Figura 9: Identificação de outliers na umidade do ar

A Figura 9 mostra os principais outliers detectados:

- Leituras de 85% em horários de aumento rápido da temperatura;
- Pequenas oscilações noturnas com variações de até 2% da média.

Os resultados indicam:

- Padrão circadiano nas variáveis ambientais (Figura 6);
- Relação inversa temperatura-umidade (Figura 8);
- Estabilidade geral do sistema monitorado (Figura 7);

## 4.5 Conclusão

A integração de Arduino com técnicas estatísticas mostrou-se viável, porém dependente de calibração e poucos recursos de memória disponível. Conforme [9], a simplicidade do Arduino facilita prototipagem, mas limita processamento de dados complexos. Soluções futuras podem incorporar edge computing, como proposto por [7], para análise preditiva em tempo real. Além disso, a aplicação de redes neurais, discutida em [3], poderia superar limitações de métodos tradicionais conjuntamente outra tecnologias para processamento mais complexo, por exemplo, uso em conjunto com Raspberry Pi. Nesse sentido, o trabalho demonstrou a viabilidade da integração de sensores com técnicas estatísticas para aplicações práticas. Os resultados indicam

que a plataforma Arduino, combinada com técnicas estatísticas, pode ser uma ferramenta poderosa para automação e monitoramento em tempo real.

Por fim, o artigo evidenciou que sensores de Arduino, combinados com técnicas estatísticas, são ferramentas poderosas para análise de dados em IoT. Aplicações práticas em agricultura, saúde e indústria foram discutidas, destacando-se a necessidade de calibração precisa e utilização de métodos estatísticos eficientes para uso em embarcados. Futuros trabalhos devem explorar a integração com frameworks de deep learning e edge analytics, seguindo tendências apontadas por [1] e [7] e uso de atuadores e tecnologias como Raspberry Pi.

## Referências

- [1] Erwin Adi, Adnan Anwar, Zubair Baig, and Sherali Zeadally. Machine learning and data analytics for the iot. *Neural Computing and Applications*, 32, 10 2020.
- [2] Katalin Ferencz, Jozsef Domokos, and Levente Kovács. Analysis of industrial sensor data using statistical and regression methods. *system theory, control and computing journal*, 3:36–44, 06 2023.
- [3] Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatin, Peyman Adibi, Payam Barnaghi, and Amit P. Sheth. Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks*, 4(3):161–175, 2018.
- [4] Gatkhal Narayan, Sachin Nalawade, Ramesh Sahni, G.B. Bhanage, Avdhoot Walunj, Pravin Kadam, and Musrrat Ali. Review of iot and electronics enabled smart agriculture. *International Journal of Agricultural and Biological Engineering*, 17:1–14, 11 2024.
- [5] T. Lakshmi Narayana, C. Venkatesh, Ajmeera Kiran, Chinna Babu J, Adarsh Kumar, Surbhi Bhatia Khan, Ahlam Almusharraf, and Mohammad Tabrez Quasim. Advances in real time smart monitoring of environmental parameters using iot and sensors. *Heliyon*, 10(7):e28195, 2024.
- [6] AUREL PICA, Laura Marcu, Ivan Cornelia Denisa, and Nicoleta Angelescu. Statistical analysis of internet of things (iot) penetration in individual consumption in eu countries. *Journal of Science and Arts*, 23:553–566, 06 2023.

- [7] H. A. Selmy, H. K. Mohamed, and W. Medhat. A predictive analytics framework for sensor data using time series and deep learning techniques. *Neural Computing and Applications*, 36:6119–6132, 2024.
- [8] Muhammad Syafrudin, Ganjar Alfian, Norma Latif Fitriyani, and Jongtae Rhee. Performance analysis of iot-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, 18:2946, 09 2018.
- [9] Ali Tasiran. *Internet of Things and Statistical Analysis*, pages 127–136. 01 2019.