

Documento Explicativo dos Cálculos Estatísticos - Airlytics

Introdução

O **Airlytics** é uma solução IoT que monitora dados ambientais, como temperatura, umidade e CO₂, com o objetivo de fornecer informações precisas e em tempo real. Além da visualização dos dados, a aplicação oferece análises estatísticas para auxiliar na interpretação e tomada de decisão.

Este documento apresenta os cálculos estatísticos utilizados na solução, detalha sua implementação e justifica sua aplicação com base no escopo do projeto.

Cálculos Estatísticos Utilizados

1. Média (Mean)

- **Definição:** Soma de todos os valores dividida pela quantidade total.
- **Motivação:** A média fornece uma visão geral do comportamento central dos dados coletados. É um indicador simples, porém eficaz, para monitorar tendências gerais de temperatura, umidade e CO₂ ao longo do tempo.
- **Fórmula:**

$$Média = \frac{\sum_{i=1}^n (PM_i \times Fi)}{\sum_{i=1}^n Fi}$$

2. Moda (Mode)

- **Definição:** O valor mais frequente em um conjunto de dados.
- **Motivação:** Identificar a moda ajuda a compreender valores mais comuns, como temperaturas ou níveis de umidade frequentemente registrados.

- **Implementação:**

O cálculo foi implementado com um contador de frequência para determinar o valor mais repetido no conjunto de dados.

3. Mediana (Median)

- **Definição:** O valor central de um conjunto de dados ordenados.
- **Motivação:** A mediana é útil para representar a "centralidade" dos dados quando há outliers que distorcem a média.
- **Fórmula:**
 - Para N ímpar:
 - Para N par:

Mediana

Divide o conjunto ao meio (“valor central”).

Como calcular:

- ordenar os valores do conjunto
- Se n for ímpar:

$$Md = x_{\left(\frac{n+1}{2}\right)}$$

- Se n for par:

$$Md = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

4. Desvio Padrão (Standard Deviation)

- **Definição:** Medida da dispersão dos dados em relação à média.
- **Motivação:** Avaliar o grau de variabilidade nos valores de temperatura, umidade e CO₂, ajudando a identificar possíveis oscilações no ambiente monitorado.
- **Fórmula:**

Desvio Padrão (Dp)

$$Dp = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

x_i = valor individual

\bar{x} = média dos valores

n = número de valores

5. Assimetria (Skewness)

- **Definição:** Mede a simetria dos dados em relação à distribuição normal.
- **Motivação:** Identificar se os dados possuem valores concentrados em uma das extremidades, o que pode indicar comportamentos atípicos ou variações sazonais.
- **Fórmula Implementada:**

$$Assimetria = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

onde:

n = é o número de observações,

x_i = representa cada valor individual no conjunto de dados,

\bar{x} = é a média do conjunto de dados,

s = é o desvio padrão do conjunto de dados.

6. Curtose (Kurtosis)

- **Definição:** Mede a "altura" e "largura" da distribuição dos dados, indicando a propensão a outliers.
- **Motivação:** Identificar a concentração de valores extremos que possam impactar a análise.
- **Fórmula Implementada:**

$$Curtose = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

7. Projeção Futura (Future Projection)

- **Definição:** Estima a inclinação da tendência (slope) com base na regressão linear.
- **Motivação:** Antecipar variações futuras nas condições ambientais com base no comportamento histórico.
- **Fórmula:**

A projeção futura foi obtida utilizando a equação da reta:

$$Y = mx + b$$

Onde:

- m : inclinação da regressão linear (slope).
- b : intercepto da regressão.

8. Probabilidade

- **Definição:** Proporção de valores que excedem a média, calculada como:

Probabilidade = Quantidade de valores acima da média

Total de valores

- **Motivação:** Avaliar a frequência com que condições mais elevadas ocorrem, como umidade ou temperatura acima da média, para detectar padrões importantes.

9. Distribuição

- **Definição:** Identifica se os dados se aproximam de uma distribuição normal ou não.
- **Motivação:** Classificar os dados para entender comportamentos esperados e possíveis anomalias.
 - **Critério:** Dados com assimetria (*Skewness*) menor que 0.5 são considerados "normalmente distribuídos".

Justificativa dos Cálculos

Os indicadores selecionados atendem aos requisitos do projeto, permitindo:

1. **Monitoramento centralizado:** Média, moda e mediana mostram tendências gerais.
2. **Identificação de variabilidade:** Desvio padrão, assimetria e curtose revelam padrões e anomalias.
3. **Análise preditiva:** Projeção futura utiliza regressão linear para antecipar tendências.
4. **Análise probabilística:** Probabilidade fornece insights sobre eventos significativos.
5. **Classificação de dados:** Determinar a distribuição permite adaptar ações conforme os padrões observados.

Conclusão

Os cálculos estatísticos do **Airlytics** foram projetados para atender à necessidade de monitoramento e análise de dados ambientais de forma robusta e eficiente. Esses indicadores oferecem um panorama abrangente, permitindo insights detalhados sobre as condições monitoradas, além de prever comportamentos futuros com base nos dados históricos.