

# Statistical Analysis Of Environmental Monitoring Data

Williams Andrade  
Universidade Federal de Pernambuco  
Recife, Brazil  
wasf@cin.ufpe.br

Gabriel Walisson  
Universidade Federal de Pernambuco  
Recife, Brazil  
gwam@cin.ufpe.br

Gleybson Ricardo  
Universidade Federal de Pernambuco  
Recife, Brazil  
gras2@cin.ufpe.br

Victor Aurelio  
Universidade Federal de Pernambuco  
Recife, Brazil  
vags@cin.ufpe.br

Fabio Filho  
Universidade Federal de Pernambuco  
Recife, Brazil  
fjcff@cin.ufpe.br

## ABSTRACT

In the Fernando de Noronha Marine National Park, a fierce competition takes place between the Natural Innovation Center (CIn) and the Catholic University (UC) to demonstrate the superiority of their environmental monitoring technologies. Both institutions have deployed sophisticated sensor networks and are engaged in detailed statistical analysis to assess the reliability of their data transmissions. While CIn highlights its cutting-edge technology, UC emphasizes its expertise in statistical analysis. This report seeks not only to present numbers and graphs, but also to provide a detailed understanding of the factors that influence the reliability of data transmissions in the sensor network of the Fernando de Noronha Marine National Park. Through a comprehensive statistical approach, we hope to provide meaningful insights that can inform strategies to optimize the efficiency and reliability of this vital environmental monitoring system.

### ACM Reference Format:

Williams Andrade, Gabriel Walisson, Gleybson Ricardo, Victor Aurelio, and Fabio Filho. 2024. Statistical Analysis Of Environmental Monitoring Data. In *Proceedings of* . ACM, New York, NY, USA, 4 pages.

## 1 INTRODUÇÃO

No Parque Nacional Marinho de Fernando de Noronha, um santuário intocado onde a vida selvagem prospera, uma rivalidade acirrada está em curso entre duas instituições de prestígio: o Centro de Inovação Natural (CIN) e a Universidade Católica (UC). Ambas competem para demonstrar a superioridade de suas tecnologias de monitoramento ambiental, cada uma buscando provar que sua abordagem é mais precisa e confiável do que a da outra. O CIN implantou uma rede de sensores de última geração, utilizando tecnologia avançada para capturar uma ampla gama de variáveis ambientais. Eles afirmam que sua abordagem é a mais eficaz, fornecendo dados cruciais para a gestão e preservação da reserva. Enquanto isso, a UC desenvolveu sua própria rede de sensores sofisticada, destacando

sua expertise em análise estatística para garantir transmissões de dados confiáveis. Determinados a provar a qualidade de sua tecnologia, eles estão empenhados em demonstrar que sua abordagem rivaliza com a do CIN, se não a supera. Ambas as instituições estão envolvidas em uma análise estatística detalhada para avaliar a confiabilidade de suas transmissões de dados. Calculando medidas de tendência central, analisando distribuições de dados e identificando possíveis problemas, tanto o CIN quanto a UC estão determinados a demonstrar a excelência de suas respectivas tecnologias. Além disso, a rivalidade se estende à análise qualitativa, onde fatores como interferência de sinal e condições climáticas são minuciosamente examinados. Ambas as equipes estão comprometidas em explorar todos os aspectos que possam afetar a confiabilidade das transmissões, buscando soluções para otimizar seus sistemas de monitoramento ambiental.

## 2 MÉTODO

### 2.1 Coleta de Dados

A coleta de dados é o primeiro passo essencial para qualquer análise estatística. Nesta fase, coletamos informações sobre a taxa de sucesso das transmissões de dados na rede de sensores do Parque Nacional Marinho de Fernando de Noronha. Esses dados são a base de nossa análise e determinam a qualidade e confiabilidade de nossos resultados.

### 2.2 Pré-processamento de Dados

O pré-processamento de dados é crucial para garantir a qualidade e integridade dos dados utilizados na análise. Nesta etapa, lidamos com valores ausentes, inconsistências e outliers que podem distorcer nossos resultados. A limpeza e tratamento adequado dos dados garantem que nossa análise seja precisa e confiável.

### 2.3 Análise Descritiva e Verificação da Distribuição dos Dados

Após a coleta e pré-processamento dos dados, é crucial realizar uma análise descritiva detalhada para compreender a distribuição da taxa de sucesso das transmissões de dados na rede de sensores do Parque Nacional Marinho de Fernando de Noronha. Utilizaremos medidas de centralidade, como média, mediana, moda, quartis, desvio padrão e variância, para explorar diferentes aspectos dos dados e identificar possíveis padrões ou anomalias.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Association for Computing Machinery.  
ACM ISBN 978-1-4503-X/18/06...\$15.00

## 2.4 Análise Descritiva e Verificação da Distribuição dos Dados

Após a coleta e pré-processamento dos dados, é crucial realizar uma análise descritiva detalhada para compreender a distribuição da taxa de sucesso das transmissões de dados na rede de sensores do Parque Nacional Marinho de Fernando de Noronha. Utilizaremos medidas de centralidade, como média, mediana, moda, quartis, desvio padrão e variância, para explorar diferentes aspectos dos dados e identificar possíveis padrões ou anomalias.

2.3.1 - Média: A média nos fornecerá uma medida do valor típico da taxa de sucesso das transmissões de dados. Seu cálculo nos permitirá entender o desempenho médio do sistema de transmissão ao longo do período de observação.

2.3.2 - Mediana: A mediana nos ajudará a entender melhor a distribuição dos valores da taxa de sucesso. Ao calcular a mediana, poderemos identificar se existe uma tendência central nos dados ou se há uma distribuição assimétrica.

2.3.3 - Moda: A moda representará o valor mais frequente da taxa de sucesso das transmissões de dados. Sua identificação nos permitirá identificar se há um valor que se destaca como o mais comum nas transmissões.

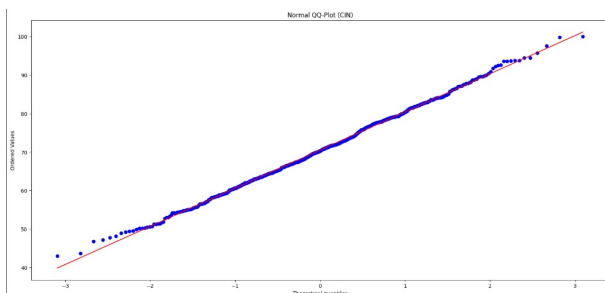
2.3.4 - Quartis: A análise dos quartis nos fornecerá informações sobre a dispersão dos dados. Os quartis nos ajudarão a identificar se existem diferenças significativas entre os percentis dos dados e se há uma distribuição uniforme ou concentrada em determinadas faixas de valores.

2.3.5 - Desvio Padrão e Variância: O desvio padrão e a variância nos fornecerão uma medida da dispersão dos dados em torno da média. Ao calcular essas medidas, poderemos determinar a variabilidade das transmissões de dados e avaliar a consistência do sistema de transmissão ao longo do tempo.

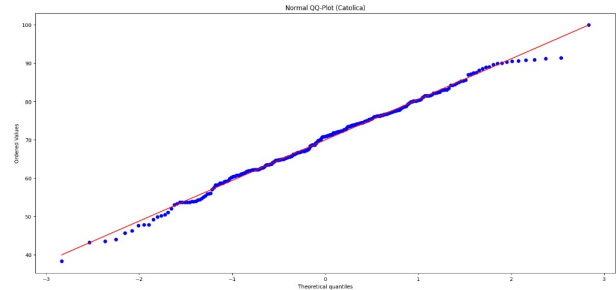
Além disso, utilizaremos gráficos como histogramas e boxplots para visualizar a distribuição dos dados e identificar possíveis padrões ou tendências. Essas visualizações nos ajudarão a compreender melhor a variabilidade dos dados e a detectar eventuais outliers que possam influenciar nossa análise.

## 2.5 Teste de Hipótese e Análise Qualitativa - Quando Aplicável

2.5.1 *Teste de Hipótese (H0)*. Não há diferença significativa na confiabilidade das transmissões de dados entre os sensores implantados pelo Centro de Inovação Natural (CIN) e os sensores implantados pela Universidade Católica (UC) no Parque Nacional Marinho de Fernando de Noronha.



2.5.2 *Hipótese Alternativa (H1)*. Existe uma diferença significativa na confiabilidade das transmissões de dados entre os sensores implantados pelo CIN e os sensores implantados pela UC no Parque Nacional Marinho de Fernando de Noronha.



2.5.3 *Método*: 1. Selecione aleatoriamente uma amostra de transmissões de dados dos sensores implantados pelo CIN e outra amostra dos sensores implantados pela UC.

2. Calcule medidas de tendência central (média, mediana) e de dispersão (desvio padrão, amplitude interquartil) para ambas as amostras.

3. Realize um teste estatístico apropriado para comparar as médias das transmissões de dados entre os dois grupos, como o teste t de Student ou um teste de Mann-Whitney U para dados não paramétricos, dependendo da distribuição dos dados.

4. Defina um nível de significância (alfa) para o teste, por exemplo,  $\alpha = 0,05$ .

5. Compare o valor-p resultante do teste com o nível de significância escolhido.

## 2.6 Interpretação dos Resultados

Após realizar o teste estatístico, avaliaremos o valor-p resultante. Se o valor-p for menor que 0,05, rejeitaremos a hipótese nula e concluiremos que há evidências estatísticas suficientes para suportar a hipótese alternativa. Isso indicaria que a localização dos sensores tem, de fato, um impacto significativo na taxa de sucesso das transmissões de dados na reserva. A realização deste teste de hipótese nos permitirá entender melhor como a localização dos sensores influencia a eficácia do sistema de monitoramento no Parque Nacional Marinho de Fernando de Noronha. Essa compreensão é crucial para aprimorar a configuração e distribuição dos sensores, garantindo uma coleta de dados mais precisa e confiável para apoiar os esforços de conservação e gestão ambiental na reserva. Demonstrando que é rejeitado a hipótese nula. Mostrando que evidências suficientes para suportar a hipótese alternativa

## 2.7 Explorando Tendências Temporais

Investigar as tendências temporais na taxa de sucesso das transmissões de dados ao longo do tempo é fundamental para entender padrões sazonais ou tendências de longo prazo. Gráficos de linha nos permitem visualizar visualmente como a taxa de sucesso varia ao longo do tempo, o que nos ajuda a identificar períodos de maior ou menor eficiência na transmissão de dados.

## 2.8 Análise de Correlação

A análise de correlação nos permite investigar a relação entre a taxa de sucesso das transmissões de dados e outras variáveis, como condições climáticas e localização dos sensores. Ao quantificar essas relações, podemos identificar fatores que influenciam a confiabilidade das transmissões de dados e direcionar esforços para otimizar o sistema.

## 2.9 Análise Qualitativa

A análise qualitativa complementa nossa análise quantitativa, permitindo compreender entender os fatores contextuais que podem afetar a confiabilidade das transmissões de dados. Isso pode incluir problemas técnicos, condições ambientais específicas ou problemas operacionais. Essa compreensão mais holística nos ajuda a interpretar os resultados da análise quantitativa de forma mais completa e informada.

## 3 RESULTADOS

### 3.1 Medidas de Dispersão

Os dados obtidos são variáveis quantitativas contínuas, que representam os valores da medição apresentando as leituras realizadas pelos sensores de ambas as universidades o CIN a universidade federal e (UC) universidade católica abrangendo todos os dados obtidos.

|                         |   |
|-------------------------|---|
| Média                   | 70.34926999999999                                   |
| Mediana                 | 70.515  |
| Moda                    | 67.46 3<br>68.72 3<br>71.15 3<br>75.90 3<br>79.33 3 |
| Quarti 1                | 63.5075   |
| Quarti 2                | 70.515  |
| Quarti 3                | 77.435  |
| Intervalo entre Quartis | 13.927500000000002                                  |
| Variância               | 101.59869136709999                                  |
| Coeficiente de variação | 14.335132705574487                                  |

Figure 1: Utilizando como base os valores encontrados e tabulados acima foram gerados gráficos com o intuito de aprofundar o entendimento dos dados demonstrados

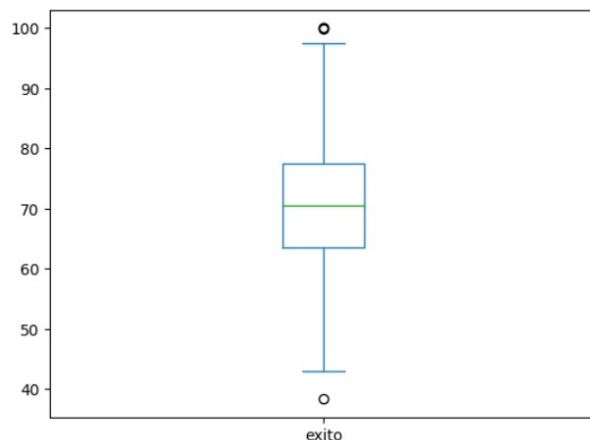


Figure 2: Boxplot

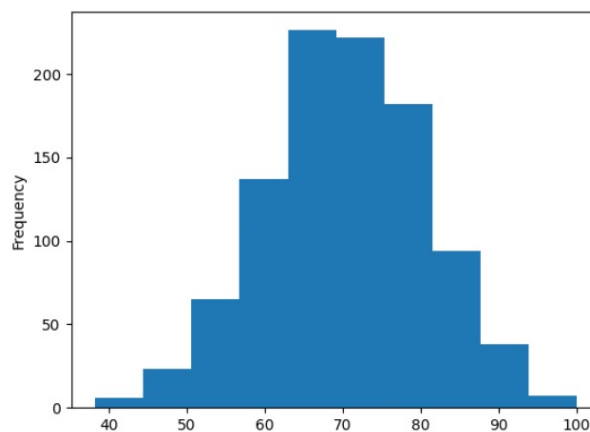


Figure 3: Histograma

Na figura 2 mostram os outliers dentro das amostras informadas apresentando os quartis e o desvio padrão. E na figura 3 fornece indícios dos dados que estão seguindo uma distribuição normal, demonstrando uma simetria em seu volume analisado e com a maioria dos dados no centro e uma diminuição nas caudas, demonstrando assim que podemos seguir o primeiro teste de hipótese.

### 3.2 Teste de Hipótese

Utilizamos primeiramente distribuição normal para os casos do CIN e da Universidade católica e como foi encontrado o p-valor de 0,4388, que é maior do que o alfa (0,05). Assim pode-se inferir que a normalidade da distribuição dos dados, Aceitando a hipótese nula. Mostrado na imagem do gráfico de distribuição mostrando o valor dos valores críticos, o "z" é menor do que o valor crítico inferior sendo assim a Demonstrando que é rejeitado a hipótese nula. Mostrando que evidências suficientes para suportar a hipótese alternativa.

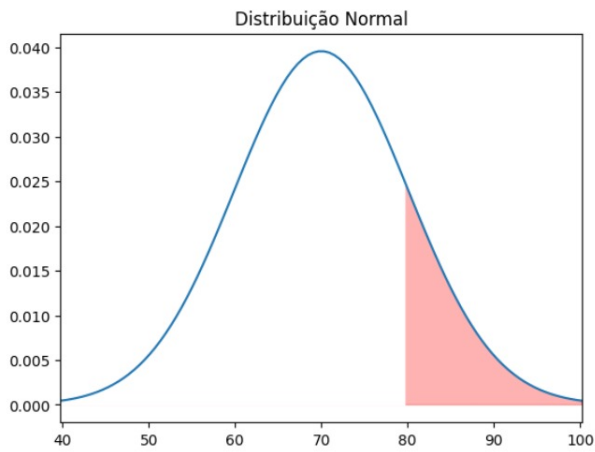


Figure 4: Gráfico do Teste de Hipótese

## 4 CONCLUSÃO

Portanto, concluímos que com os dados apresentados após o processamento tivemos o histograma mostrando ser uma distribuição normal assim seguindo para a Hipótese Alternativa em que confiabilidade das transmissões de dados entre os sensores implantados pela UC tem mais assertividade do que os do CIN como tendo mais evidência para suportá-la.

## 5 REFERÊNCIA

[1] <https://github.com/FabinhoFarias/projeto-estatistica-ufpe?tab=readme-ov-file>