

Relatório Técnico

Fátima P. Santos Pinho¹, Ismael R. de Oliveira Neto², Rebeca Helen B. Amorim³

¹Curso de Sistemas de Informação – Centro Universitário de Excelência (UNEX)
Av. Artêmia Pires Freitas – s/n – Sim – 44085-370 – Feira de Santana, BA – Brasil

fatmapinh0@gmail.com,
rebecabatista165@gmail.com, ismaelneto2003@gmail.com

1. Introdução

O conhecimento em Estatística é importante em diversas áreas de atuação. Desenvolvedores de software lidam diariamente com algoritmos voltados à análise de dados. A Estatística oferece ferramentas capazes de auxiliar no monitoramento e interpretação dessas informações, sejam elas provenientes de pequenos ou grandes volumes de dados. Nesse contexto, percebe-se que a Estatística está diretamente ligada à mineração de dados e à obtenção de resultados relevantes, e compreender essas métricas é essencial antes de aplicar algoritmos de Machine Learning.

2. Fundamentação Teórica

2.1 Média Aritmética

A Média Aritmética é calculada pela soma de todos os elementos do conjunto dividida pela quantidade de elementos do conjunto.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Figura 1. Fórmula média aritmética

2.2 Mediana

A Mediana é o valor que ocupa a posição central dos valores de um conjunto ordenado. Para encontrá-la, é possível listar os termos em ordem crescente ou decrescente e encontrar o termo que ocupa a posição central.

Exemplo: $A = \{1,63; 1,65; 1,72; \underline{1,75}; 1,79; 1,80; 1,82\}$

Como $7 / 2 = 3,5$, arredondando, a mediana é a 4ª posição do conjunto, sendo assim a mediana é 1,75.

2.3 Moda

A Moda é encontrada através da observação de um conjunto, o resultado mais recorrente no conjunto, ou seja, com maior frequência absoluta é a moda.

Exemplo: $N = \{35, \underline{37}, 36, 38, 35, \underline{37}, \underline{37}, 33, 36, \underline{37}, \underline{37}, 34, 33, \underline{37}, \underline{37}\}$

No conjunto N a moda é o número 37, pois é ele que se repete com maior frequência.

2.4 Desvio Padrão

O Desvio Padrão é chamado especificamente de desvio de padrão amostral. Um baixo desvio indica que os dados estão próximos da média ou do valor esperado. Já um alto desvio padrão, indica que os dados estão espalhados por uma ampla gama de valores.

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

Figura 2. Fórmula desvio padrão

2.5 Variância Populacional

O cálculo da Variância Populacional é obtido através da soma dos quadrados da diferença entre cada valor e a média aritmética, dividida pela quantidade de elementos observados.

$$var = \frac{(segunda - média\ aritmética)^2 + (terça - média\ aritmética)^2 + \dots + (sexta - média\ aritmética)^2}{quantidade\ de\ dias}$$

Figura 3. Fórmula variância populacional

2.6 Covariância

Quando duas variáveis aleatórias X e Y não são independentes, geralmente é de interesse avaliar quão fortemente estão relacionadas uma com a outra. A covariância dá uma ideia da dispersão dos valores da variável bidimensional (X,Y) em relação ao ponto (E(X),E(Y)).

Definição 13.1 Seja (X, Y) uma variável aleatória bidimensional. A covariância de X e Y que denotaremos $Cov(X, Y)$ é definida por:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Figura 4. Fórmula covariância

2.7 Item Único

Dado um conjunto de dados, através do uso de um algoritmo é possível observá-la e extrair apenas os valores únicos daquele conjunto. Ou seja, os valores que aparecem uma única vez, sem repetição.

Exemplo: C = {1, 2, 2, 3, 4, 4} Itens únicos = {1,3}

2.8 Frequência Absoluta

A frequência absoluta é o número de vezes que um mesmo elemento se repetiu em um conjunto de dados. Exemplo: o objetivo de Ana Júlia era descobrir quais seriam os sabores favoritos dos seus colegas, ela realizou a contagem do número de vezes que um mesmo sabor apareceu e construiu a tabela frequência. A quantidade de vezes que um mesmo sabor apareceu é a sua frequência absoluta.

Sabor	Frequência absoluta
Brigadeiro	16
Ninho	12
Churros	6
Paçoca	7
Casadinho	5
Moranginho	4
Total	50

Figura 5. Tabela de Frequência Absoluta

2.9 Frequência Relativa

Conhecemos como frequência relativa a divisão entre a frequência absoluta e o número de dados coletados para um determinado conjunto. Como o nome sugere, a frequência relativa mostra a frequência que um determinado dado tem em relação ao todo, por isso, é bastante comum que ela seja representada como uma porcentagem.

Exemplo 2: Com a intenção de compreender melhor o fluxo de correntes ao decorrer de uma semana, o número de clientes que uma empresa atendeu nesse período foi anotado na lista a seguir: Segunda-feira: 10 clientes, Terça-feira: 11 clientes, Quarta-feira: 8 clientes, Quinta-feira: 16 clientes, Sexta-feira: 25 clientes e Sábado: 30 clientes.

Dia da semana	Frequência absoluta	Frequência relativa	Frequência relativa (%)
Segunda-feira	16	$16 : 150 = 0,11$	11%
Terça-feira	15	$15 : 150 = 0,10$	10%
Quarta-feira	12	$12 : 150 = 0,08$	8%
Quinta-feira	20	$20 : 150 = 0,13$	13%
Sexta-feira	37	$37 : 150 = 0,25$	25%
Sábado	50	$50 : 150 = 0,33$	33%
Total	150	$100 : 100 = 1$	100%

Figura 6. Tabela de Frequência Absoluta e Relativa

2.10 Frequência absoluta acumulada

A frequência absoluta acumulada é a soma das frequências absolutas ao decorrer das linhas da tabela. Essa frequência é bastante útil para obter alguns dados de determinada tabela.

Na tabela a seguir, a quantidade de funcionários nada mais é que a frequência absoluta, pois é o número de vezes que um salário se repete na folha de pagamento. Para construir a frequência acumulada na primeira linha, apenas copiamos a frequência absoluta, depois da primeira linha, somando a frequência absoluta da linha anterior com a frequência absoluta da linha atual.

Salário	Frequência absoluta	Frequência absoluta acumulada
R\$ 1155,55	24	24
R\$ 1520,00	13	$24 + 13 = 37$
R\$ 2000,00	10	$37 + 10 = 47$
R\$ 2500,00	18	$47 + 18 = 65$
R\$ 5000,00	5	$65 + 5 = 70$
Total	70	

Figura 7. Tabela de frequência absoluta acumulada

2.11 Frequência relativa acumulada

A frequência relativa acumulada é o acúmulo da frequência relativa. Para encontrar a frequência relativa acumulada, acrescentamos uma nova coluna à tabela. Copiamos a primeira frequência relativa na primeira linha, a segunda linha será a soma da frequência relativa da linha com a frequência acumulada da linha anterior, e assim sucessivamente. Veja um exemplo a seguir:

Dia da semana	Frequência absoluta	Frequência relativa	Frequência relativa acumulada
Segunda-feira	16	0,11	0,11
Terça-feira	15	0,10	0,11 + 0,10 = 0,21
Quarta-feira	12	0,08	0,21 + 0,08 = 0,29
Quinta-feira	20	0,13	0,29 + 0,13 = 0,42
Sexta-feira	37	0,25	0,42 + 0,25 = 0,67
Sábado	50	0,33	0,67 + 0,33 = 1,00
Total	150	1	

Figura 8. Tabela de frequência relativa acumulada

2.12 Probabilidade condicional

A probabilidade condicional é a chance de um determinado evento acontecer tendo como base que um evento aconteceu anteriormente; ambos os eventos possuem o mesmo espaço amostral. Esse evento que ocorreu anteriormente é conhecido como condicionante.

Dizemos que a probabilidade de o evento A acontecer, sabendo que o evento B aconteceu, é conhecida como probabilidade condicional do evento A dado B, representada por $P(A|B)$ (lê-se: probabilidade de A dado B). Essa probabilidade só existe se os eventos possuírem o mesmo espaço amostral (diferente do vazio) e se B não for um evento impossível.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Figura 9. Fórmula Probabilidade condicional

Exemplo:

Durante o lançamento de dois dados, queremos calcular a probabilidade de o resultado da soma das faces superiores ser igual a 6, sabendo que o resultado do lançamento dos dados são dois números pares.

Perceba que há dois eventos:

$B \rightarrow$ As duas faces são números pares.

$A \rightarrow$ A soma das faces superiores é igual a 6.

Então, queremos a probabilidade $P(A|B)$: $P(\text{A soma das faces superiores é 6} \mid \text{As duas faces são números pares})$.

3. Metodologia

Neste projeto, foi desenvolvida uma classe em Python para calcular métricas estatísticas diretamente a partir de um dicionário de listas, representando um dataset simples. A proposta foi construir todas as funcionalidades manualmente, sem recorrer a bibliotecas externas, garantindo que cada cálculo fosse compreendido na prática.

Primeiramente, A classe **Statistics** verifica se o dataset fornecido está no formato correto, onde cada coluna deve ser uma lista e todas as colunas devem ter o mesmo tamanho. Garantindo que os cálculos sejam consistentes e evitando erros durante a execução das funções.

Cada métrica estatística foi implementada como um método específico, seguindo conceitos clássicos da estatística. Criamos funções para calcular média, mediana, moda,

variância, desvio-padrão, covariância, frequências (absoluta, relativa e acumulada), conjuntos de valores únicos e probabilidade condicional. Cada método foi projetado para lidar com diferentes tipos de dados e cenários.

Cada função trabalha de forma independente, o que facilita a leitura, manutenção e teste do código. Além disso, incluímos verificações de erro na inicialização para entradas inválidas, garantindo que o usuário recebesse mensagens claras caso fornecesse dados inconsistentes. E por fim, realizamos testes utilizando conjuntos de dados simples e conhecidos fornecidos pelo Professor da disciplina. Os quais confirmaram que cada métrica estava correta e se comportava conforme esperado. A validação manual dos resultados permitiu identificar ajustes necessários e garantir que todas as funções da classe entregassem valores confiáveis.

4. Resultados

Após a execução dos testes iniciais, identificamos que algumas métricas precisavam de ajustes para garantir resultados consistentes. Realizamos as alterações necessárias no código, revisando cálculos e lógica de implementação. Com essas correções, conseguimos processar o dataset corretamente e validar que todas as métricas retornavam resultados confiáveis, atendendo aos objetivos do projeto.

5. Considerações Finais

Um dos maiores desafios foi relacionar o cálculo das fórmulas com a lógica de programação. Além disso, nenhum dos integrantes da equipe tinha familiaridade prévia com a linguagem Python, o que tornou o desafio ainda mais intenso.

Foi muito interessante perceber a importância da Estatística no desenvolvimento de software, especialmente na Mineração de Dados, onde os algoritmos precisam ser analisados e compreendidos detalhadamente.

O principal aprendizado foi que testes frequentes são essenciais. Não realizá-los durante o desenvolvimento inicial levou a ajustes significativos após a execução dos testes finais.

6. Referências

A IMPORTÂNCIA da estatística para a ciência de dados. [S. l.], 14 set. 2020. Disponível em:

<https://iaexpert.academy/2020/09/14/importancia-da-estatistica-para-ciencia-de-dados>.

Acesso em: 23 ago. 2025.

DOCUMENTAÇÃO Python. [S. l.], [2025]. Disponível em:

<https://docs.python.org/pt-br/3.13/contents.html>. Acesso em: 23 ago. 2025.

OLIVEIRA, Raul Rodrigues de. Moda, média e mediana. Brasil Escola. Disponível em:

<https://brasilecola.uol.com.br/matematica/moda-media-mediana.htm>. Acesso em: 23

ago. 2025.

VARIÂNCIA e desvio padrão. [S. l.], [2025]. Disponível em:

<https://mundoeducacao.uol.com.br/matematica/variancia-desvio-padrao.htm>. Acesso em:

23 ago. 2025.

DESVIO padrão. [S. l.], 27 fev. 2019. Disponível em:

<https://www.educamaisbrasil.com.br/enem/matematica/desvio-padrao>. Acesso em: 23

ago. 2025.

LIBERAL, Tarciana. Aula 13 – Covariância e Correlação. Probabilidade II. Universidade Federal da Paraíba, Departamento de Estatística. Disponível em:

<http://www.de.ufpb.br/~tarciana/Probabilidade2/Aula13.pdf>. Acesso em: 23 ago. 2025.

OLIVEIRA, Raul Rodrigues de. Frequência absoluta: o que é e como calcular? Mundo Educação. Disponível em:

<https://mundoeducacao.uol.com.br/matematica/frequencia-absoluta.htm>. Acesso em: 23

ago. 2025.

OLIVEIRA, Raul Rodrigues de. Frequência relativa: como calcular e exercícios. Mundo Educação. Disponível em:

<https://mundoeducacao.uol.com.br/matematica/frequencia-relativa.htm>. Acesso em: 23

ago. 2025.

ASTH, Rafael. Probabilidade condicional. Toda Matéria, [s.d.]. Disponível em:

<https://www.todamateria.com.br/probabilidade-condicional/>. Acesso em: 23 ago. 2025.