

# Metodos Quantitativos - A9

Marcelo Wzorek Filho, Rafael Leal Machado

March 15, 2024

## Abstract

Neste trabalho, implementamos e analisamos um [Conjunto de Dados](#) de nossa escolha, além de criar um algoritmo em Python que visa a criação de tabelas de distribuição de frequência de cada uma das variáveis previamente selecionadas.

## 1 Introdução

O conjunto de dados que utilizo em meu trabalho é composto por informações sobre carros fabricados em 1985. Este conjunto de dados é cru e desorganizado, exigindo uma análise cuidadosa para extrair informações.

A escolha desse conjunto de dados foi feita com base em meu interesse pessoal por carros antigos. Por se tratar de um conjunto de dados pequeno, com apenas 205 instâncias, não foi necessário fazer uma amostragem. Além disso, o tema dos dados oferece várias oportunidades para análises.

Para uma análise estatística mais precisa, foi implementado um programa em Python para a criação de Tabelas de Distribuição de Frequência para os dados disponíveis. A implementação do algoritmo utilizado está disponível no [GitHub](#).

## 2 Descrição do Conjunto de Dados

O conjunto de dados que estamos explorando oferece uma visão abrangente das características dos veículos fabricados em 1985. Com um total de 26 atributos, tornando este conjunto de dados interessante e abrangente. Ele contém informações detalhadas sobre diversos aspectos dos carros, como marca, tipo de combustível, estilo de carroceria, tipo de motor e sistema de alimentação de combustível, entre outros. Essa variedade de atributos oferece diversas possibilidades para análise e exploração.

Além disso, a inclusão de atributos como classificação de risco de seguro e perdas normalizadas permite uma análise mais aprofundada sobre a relação entre as características dos veículos. Isso pode ser especialmente relevante para profissionais da indústria automotiva, seguradoras e entusiastas que desejam entender melhor as tendências e padrões relacionados aos veículos produzidos na década de 1980.

Confira algumas análises deste conjunto feitas por outros contribuintes da plataforma Kaggle: [Model Development and Evaluation with Python](#).  
[Exploratory Data Analysis with Python](#).  
[Data Wrangling](#).

## 3 Descrição dos Atributos

### 3.1 classificação-de-risco:

Este atributo é ordinal, pois os valores -3, -2, -1, 0, 1, 2, 3 representam categorias com uma ordem específica, indicando o grau de risco associado ao veículo.

### 3.2 perdas-normalizadas:

Este é um atributo contínuo, pois representa valores numéricos que podem assumir uma gama infinita de valores dentro do intervalo definido.

### 3.3 marca:

Este atributo é nominal, pois representa categorias sem uma ordem específica, apenas identificando a marca do carro.

### 3.4 tipo-combustivel:

Outro atributo nominal, pois apenas categoriza o tipo de combustível usado, sem ordem específica.

### 3.5 aspiracao:

Também é nominal, distinguindo entre o tipo de sistema de aspiração do motor.

### 3.6 num-de-portas:

Este é um atributo discreto, pois representa o número de portas dos veículos em valores discretos (four, two).

### 3.7 estilo-carroceria:

Atributo nominal que descreve o estilo da carroceria, sem ordem específica.

### 3.8 tração-rodas:

Outro atributo nominal, que categoriza o tipo de tração das rodas.

### 3.9 localizacao-motor:

Atributo nominal que indica a localização do motor, na frente ou na traseira do veículo.

### 3.10 distancia-entre-eixos, comprimento, largura, altura, peso-padrao, tamanho-motor, diâmetro, curso, taxa-compressão, potência, pico-rpm, consumo-cidade, consumo-estrada, preço:

Todos esses atributos são contínuos, pois representam valores numéricos que podem variar de forma contínua dentro de um intervalo específico.

## 4 Descrição dos Dados Escolhidos

Os dados escolhidos para análise incluem **distância entre eixos**, um indicador importante da proporção do veículo e pode influenciar no conforto e estabilidade durante a condução.

A **altura** e o **peso total padrão** são atributos que oferecem informações sobre a robustez do veículo. O **tamanho do motor** e a **potência do motor** são características ligadas diretamente ao desempenho e a capacidade de resposta do veículo.

As **rotações máximas por minuto** fornecem informações sobre o alcance de operação ideal do motor, enquanto a **eficiência de combustível na cidade e na rodovia** ajudam a entender o consumo de combustível em diferentes condições de direção. Por fim, o **preço** é um fator decisivo para muitos consumidores ao considerar a compra de um veículo, e analisar sua relação com os demais atributos pode revelar uma relação importantes entre o valor percebido e real do carro.

## 5 Tabela de Distribuição de Frequência

A Tabela de Distribuição de Frequência é uma ferramenta estatística fundamental que organiza e resume os dados de uma variável em categorias ou intervalos, exibindo a frequência com que cada categoria ocorre. Ela é utilizada para visualizar e compreender a distribuição dos dados, fornecendo dados sobre padrões, tendências e comportamentos presentes nos dados coletados.

Essa tabela é importante porque permite uma análise sistemática e estruturada dos dados, facilitando a identificação de valores atípicos, a compreensão da concentração dos dados em determinadas faixas e a detecção de possíveis padrões ou relações entre variáveis. Além disso, a Tabela de Distribuição de Frequência é frequentemente utilizada como base para a construção de gráficos, como histogramas e gráficos de barras, que auxiliam na visualização e interpretação dos dados, o que ajuda a extrair insights valiosos e fundamentando processos de análise e tomada de decisão.

## 6 Implementação

A implementação do código python para análise estatística e visualização das Tabelas de Distribuição de Frequência do conjunto de dados escolhidos consiste na leitura de um arquivo CSV contendo os dados a serem analisados, extração dos campos relevantes e armazenamento dos mesmos em listas correspondentes. Depois, chama a função de criação da tabela para a lista de variáveis escolhida pelo usuário, gerando tabelas de distribuição de frequências, como distância entre eixos, altura, peso total padrão, etc.

Para calcular e plotar as tabelas de distribuição de frequências, o código implementa uma série de funções que organizam os dados em intervalos, calculam as frequências de ocorrência em cada intervalo, e os apresentam de forma tabular. Além disso, inclui uma série de funções auxiliares para ordenar os dados, calcular amplitude, determinar o número de classes, calcular intervalos, ponto médio e frequência acumulada.

Por fim, o código utiliza a biblioteca matplotlib para plotar a tabela de distribuição de frequência, exibindo os intervalos, os pontos médios, as frequências e as frequências acumuladas em cada intervalo. Essa visualização é essencial para compreender a distribuição dos dados e identificar padrões ou tendências presentes nos mesmos.

## 7 Resultados

### 7.1

Tabela da Distância entre Eixos

| Intervalo      | Ponto Médio | Frequência | Frequência Acumulada |
|----------------|-------------|------------|----------------------|
| (86.6, 90.6)   | 88.6        | 8          | 8                    |
| (90.6, 94.6)   | 92.6        | 49         | 57                   |
| (94.6, 98.6)   | 96.6        | 65         | 122                  |
| (98.6, 102.6)  | 100.6       | 32         | 154                  |
| (102.6, 106.6) | 104.6       | 21         | 175                  |
| (106.6, 110.6) | 108.6       | 16         | 191                  |
| (110.6, 114.6) | 112.6       | 7          | 198                  |
| (114.6, 118.6) | 116.6       | 2          | 200                  |
| (118.6, 122.6) | 120.6       | 1          | 201                  |

### 7.2

Tabela da Altura

| Intervalo    | Ponto Médio | Frequência | Frequência Acumulada |
|--------------|-------------|------------|----------------------|
| (47.8, 49.8) | 48.8        | 12         | 12                   |
| (49.8, 51.8) | 50.8        | 36         | 48                   |
| (51.8, 53.8) | 52.8        | 46         | 94                   |
| (53.8, 55.8) | 54.8        | 70         | 164                  |
| (55.8, 57.8) | 56.8        | 27         | 191                  |
| (57.8, 59.8) | 58.8        | 8          | 199                  |
| (59.8, 61.8) | 60.8        | 2          | 201                  |
| (61.8, 63.8) | 62.8        | 0          | 201                  |
| (63.8, 65.8) | 64.8        | 0          | 201                  |

### 7.3

Tabela do Peso Padrão

| Intervalo        | Ponto Médio | Frequência | Frequência Acumulada |
|------------------|-------------|------------|----------------------|
| (1488.0, 1775.0) | 1631.5      | 2          | 2                    |
| (1775.0, 2062.0) | 1918.5      | 38         | 40                   |
| (2062.0, 2349.0) | 2205.5      | 43         | 83                   |
| (2349.0, 2636.0) | 2492.5      | 39         | 122                  |
| (2636.0, 2923.0) | 2779.5      | 28         | 150                  |
| (2923.0, 3210.0) | 3066.5      | 29         | 179                  |
| (3210.0, 3497.0) | 3353.5      | 11         | 190                  |
| (3497.0, 3784.0) | 3640.5      | 7          | 197                  |
| (3784.0, 4071.0) | 3927.5      | 4          | 201                  |

### 7.4

Tabela do Tamanho do Motor

| Intervalo      | Ponto Médio | Frequência | Frequência Acumulada |
|----------------|-------------|------------|----------------------|
| (61.0, 91.0)   | 76.0        | 16         | 16                   |
| (91.0, 121.0)  | 106.0       | 91         | 107                  |
| (121.0, 151.0) | 136.0       | 51         | 158                  |
| (151.0, 181.0) | 166.0       | 20         | 178                  |
| (181.0, 211.0) | 196.0       | 16         | 194                  |
| (211.0, 241.0) | 226.0       | 2          | 196                  |
| (241.0, 271.0) | 256.0       | 2          | 198                  |
| (271.0, 301.0) | 286.0       | 0          | 198                  |
| (301.0, 331.0) | 316.0       | 3          | 201                  |

## 7.5

Tabela da Potência do Motor

| Intervalo      | Ponto Médio | Frequência | Frequência Acumulada |
|----------------|-------------|------------|----------------------|
| (48.0, 72.0)   | 60.0        | 53         | 53                   |
| (72.0, 96.0)   | 84.0        | 49         | 102                  |
| (96.0, 120.0)  | 108.0       | 51         | 153                  |
| (120.0, 144.0) | 132.0       | 13         | 166                  |
| (144.0, 168.0) | 156.0       | 22         | 188                  |
| (168.0, 192.0) | 180.0       | 8          | 196                  |
| (192.0, 216.0) | 204.0       | 4          | 200                  |
| (216.0, 240.0) | 228.0       | 0          | 200                  |
| (240.0, 264.0) | 252.0       | 1          | 201                  |

## 7.6

Tabela de Rotação Máxima por Minuto

| Intervalo        | Ponto Médio | Frequência | Frequência Acumulada |
|------------------|-------------|------------|----------------------|
| (4150.0, 4423.0) | 4286.5      | 20         | 20                   |
| (4423.0, 4696.0) | 4559.5      | 8          | 28                   |
| (4696.0, 4969.0) | 4832.5      | 41         | 69                   |
| (4969.0, 5242.0) | 5105.5      | 55         | 124                  |
| (5242.0, 5515.0) | 5378.5      | 55         | 179                  |
| (5515.0, 5788.0) | 5651.5      | 1          | 180                  |
| (5788.0, 6061.0) | 5924.5      | 19         | 199                  |
| (6061.0, 6334.0) | 6197.5      | 0          | 199                  |
| (6334.0, 6607.0) | 6470.5      | 2          | 201                  |

## 7.7

Tabela de Eficiência na Cidade

| Intervalo    | Ponto Médio | Frequência | Frequência Acumulada |
|--------------|-------------|------------|----------------------|
| (13.0, 17.0) | 15.0        | 11         | 11                   |
| (17.0, 21.0) | 19.0        | 45         | 56                   |
| (21.0, 25.0) | 23.0        | 46         | 102                  |
| (25.0, 29.0) | 27.0        | 41         | 143                  |
| (29.0, 33.0) | 31.0        | 40         | 183                  |
| (33.0, 37.0) | 35.0        | 4          | 187                  |
| (37.0, 41.0) | 39.0        | 11         | 198                  |
| (41.0, 45.0) | 43.0        | 0          | 198                  |
| (45.0, 49.0) | 47.0        | 2          | 200                  |

## 7.8

Tabela de Eficiência na Rodovia

| Intervalo    | Ponto Médio | Frequência | Frequência Acumulada |
|--------------|-------------|------------|----------------------|
| (16.0, 21.0) | 18.5        | 9          | 9                    |
| (21.0, 26.0) | 23.5        | 50         | 59                   |
| (26.0, 31.0) | 28.5        | 46         | 105                  |
| (31.0, 36.0) | 33.5        | 47         | 152                  |
| (36.0, 41.0) | 38.5        | 34         | 186                  |
| (41.0, 46.0) | 43.5        | 8          | 194                  |
| (46.0, 51.0) | 48.5        | 5          | 199                  |
| (51.0, 56.0) | 53.5        | 2          | 201                  |
| (56.0, 61.0) | 58.5        | 0          | 201                  |

## 7.9

Tabela de Preço

| Intervalo          | Ponto Médio | Frequência | Frequência Acumulada |
|--------------------|-------------|------------|----------------------|
| (5118.0, 9594.0)   | 7356.0      | 90         | 90                   |
| (9594.0, 14070.0)  | 11832.0     | 46         | 136                  |
| (14070.0, 18546.0) | 16308.0     | 35         | 171                  |
| (18546.0, 23022.0) | 20784.0     | 11         | 182                  |
| (23022.0, 27498.0) | 25260.0     | 3          | 185                  |
| (27498.0, 31974.0) | 29736.0     | 4          | 189                  |
| (31974.0, 36450.0) | 34212.0     | 7          | 196                  |
| (36450.0, 40926.0) | 38688.0     | 2          | 198                  |
| (40926.0, 45402.0) | 43164.0     | 3          | 201                  |

## 8 Conclusão e Considerações Finais

Após uma análise abrangente dos dados sobre carros do ano de 1985, podemos tirar várias conclusões importantes sobre as características e especificações desses veículos.

### 8.1 Identificação de Padrões:

Ao analisar as tabelas de distribuição de frequência para cada variável, podemos identificar padrões e tendências nos dados dos carros de 1985. Por exemplo, podemos observar se há uma faixa de distância entre eixos, altura, peso, tamanho do motor, potência do motor, rotação máxima por minuto, eficiência na cidade e na rodovia, e preço que é mais comum entre os carros desse ano. Podemos identificar se existem clusters ou agrupamentos de valores em determinadas faixas, o que pode sugerir preferências de design, características técnicas dominantes ou tendências de mercado na época, o que pode ser observado nas tabelas de **Altura** e **Distância entre Eixos**.

### 8.2 Medidas de Tendência Central:

A partir das tabelas de distribuição de frequência, podemos calcular medidas de tendência central, como a média, mediana e moda, para cada variável, o que permite entender onde os dados tendem a se agrupar .

### 8.3 Medidas de Dispersão:

Além das medidas de tendência central, também podemos calcular medidas de dispersão, como a variância e o desvio padrão, para entender o quão dispersos estão os dados em relação à média. Isso nos fornece insights sobre a variabilidade nos dados.

### 8.4 Identificação de Valores Atípicos:

Ao examinar as tabelas e visualizar os dados, podemos identificar valores atípicos em cada variável, o que podem indicar carros com características incomuns ou excepcionais em relação aos padrões



gerais da época, o que pode ser importante para compreender o mercado automotivo de 1985. Um exemplo bastante concreto disso é retratado na **Tabela de Eficiência na Cidade**, onde uma dos dados simplesmente não foi capaz de abranger nenhum dos intervalos.

## 8.5 Tomada de Decisão:

Com base nos padrões identificados podemos tomar decisões com uma grande variedade de contextos.

Por exemplo, os dados obtidos podem ser usados por fabricantes de carros para entender as preferências dos consumidores e orientar o desenvolvimento de novos modelos. Os compradores de carros usados podem usar essas informações para fazer escolhas mais informadas sobre qual carro comprar com base em suas preferências pessoais e necessidades. Governos e reguladores podem usar os dados para entender tendências de eficiência de combustível e emissões ao longo do tempo e formular políticas relacionadas.

Isso é uma citação [1].

## References

- [1] Fazil T. “1985 Automobile Dataset”. In: *Kaggle* 1.1 (2023), pp. 1–1.