

DADOS ESTATÍSTICOS - CdD-ML

23 DE ABRIL DE 2025

OBJETIVO DO DOCUMENTO

Este documento tem como objetivo apresentar uma análise estatística descritiva dos gastos realizados pelos deputados federais brasileiros no período de 2022 a 2025, utilizando dados públicos disponibilizados pela API de Dados Abertos da Câmara dos Deputados. A análise inclui medidas como média, mediana, moda, desvio padrão e a identificação de valores extremos (outliers), com foco em entender padrões de comportamento, inconsistências e possíveis desvios nos dados.



AQUISIÇÃO DOS DADOS

Os dados foram obtidos diretamente da API de Dados Abertos da Câmara dos Deputados, acessível publicamente via:

<https://dadosabertos.camara.leg.br/api/v2/deputados>

- Os dados foram consultados utilizando um script em Python com a biblioteca `requests`.
- As requisições foram feitas por deputado e por ano (2022 a 2025).
- As informações foram processadas e salvas em arquivos CSV comprimidos para cada ano.



DADOS COLETADOS

Análise Geral	
 Ano	 Registros Coletados
2022	32 registros
2023	24.530 registros
2024	49.253 registros
2025	36.171 registros
TOTAL GERAL	109.986 registros

As colunas analisadas incluem


- `valorDocumento`;
- `nomeDeputado`;
- `tipoDespesa`;
- `fornecedor`;
- `dataDocumento`;

ANÁLISE ANO A ANO

Ano 2022	
 Estatística	 Valor (R\$)
Média	7.644,60
Mediana	3.285,00
Moda	400,00
Desvio Padrão	11.685,89
Máximo	50.000,00
Mínimo	24,41


Observações importantes:

- **IQR (Q3 - Q1):** R\$ 5.462,50;
- **Outliers superiores:** valores > R\$ 14.968,75
- **Inferência:** Alta média comparada à mediana indica influência de grandes despesas.

Ano 2023	
 Estatística	Ⓜ Valor (R\$)
Média	1.729,52
Mediana	320,99
Moda	100,00
Desvio Padrão	4.415,31
Máximo	164.900,00
Mínimo	-3.569,91


Observações importantes:

- **IQR:** R\$ 1.128,20
- **Outliers superiores:** valores > R\$ 3.012,30
- **Outliers inferiores:** valores < R\$ -1.500,50
- **Inferência:** Despesas muito altas e negativas foram identificadas.

Ano 2024	
 Estatística	Ⓜ Valor (R\$)
Média	1.581,67
Mediana	300,27
Moda	100,00
Desvio Padrão	4.121,47
Máximo	209.948,00
Mínimo	-3.592,01

Observações importantes:


- **IQR:** R\$ 955,39
- **Outliers superiores:** valores > R\$ 2.578,48
- **Outliers inferiores:** valores < R\$ -1.253,52
- **Inferência:** O valor máximo ultrapassa 200 mil, o que é um caso extremo.

Ano 2025	
 Estatística	Ⓜ Valor (R\$)
Média	1.298,06
Mediana	284,82
Moda	200,00
Desvio Padrão	3.138,24
Máximo	70.000,00
Mínimo	-3.912,52

Observações importantes:

- **IQR:** R\$ 842,20
- **Outliers superiores:** valores > R\$ 2.263,30
- **Outliers inferiores:** valores < R\$ -1.103,70
- **Inferência:** Apesar da média ser menor que anos anteriores, ainda há altos desvios e valores negativos.




ANÁLISE GERAL (2022 - 2025)

Todos os anos	
 Estatística	Ⓜ Valor (R\$)
Média	1.523,14
Mediana	300,00
Moda	100,00
Desvio Padrão	3.907,73
Máximo	209.948,00
Mínimo	-3.912,52

Observações importantes:

- **IQR:** R\$ 947,58
- **Outliers superiores:** valores > R\$ 2.548,47
- **Outliers inferiores:** valores < R\$ -1.440,46
- **Inferência:** A diferença entre média e mediana reforça que **há forte assimetria positiva.**

RESUMO GERAL DOS OUTLIERS

Resumo			
 Ano	 Valores Negativos?	 Limite Superior (Outlier)	Ⓜ Gasto Máximo
2022	Não	~R\$ 14.968,75	R\$ 50.000,00
2023	Sim	~R\$ 3.012,30	R\$ 164.900,00
2024	Sim	~R\$ 2.578,48	R\$ 209.948,00
2025	Sim	~R\$ 2.263,30	R\$ 70.000,00
Geral	Sim	~R\$ 2.548,47	R\$ 209.948,00

POSSÍVEIS CAUSAS PARA OS VALORES EXTREMOS

Valores muito altos:

- Despesas acumuladas.
- Verbas indenizatórias específicas.
- (SUPosição) Lançamentos errados com casas decimais incorretas (ex: R\$ 2.000,00 sendo inserido como R\$ 200.000,00).
- Reembolso de despesas múltiplas em uma só operação.

Valores negativos:

- Estornos de valores pagos anteriormente.
- Correções administrativas ou glosas.
- Lançamentos duplicados com posterior anulação.

CONCLUSÃO

A nossa análise evidencia a existência de valores extremos e gastos inconsistentes, que devem ser cuidadosamente tratados em qualquer modelo de Machine Learning.

Por isso, temos de focar em:

- Normalizar ou remover os outliers;
- Segmentar por tipo de despesa;
- Investigar mais os gastos anômalos.