

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO
CAMPUS SÃO PAULO**

TECNÓLOGO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

GUSTAVO BEZERRA DE ANDRADE – SP313475X

JEFFERSON BISPO DE LIMA DE JESUS – SP3139328

SEUNG JIN JUCÁ CASIERRA – SP3134229

VITOR AUGUSTO DOS SANTOS – SP3132048

**ANÁLISE ESTATÍSTICA DA RELAÇÃO ENTRE VEÍCULOS ELÉTRICOS E PADRÕES
DE ACIDENTES DE TRÂNSITO EM NOVA YORK**

SÃO PAULO

2025

Sumário	
INTRODUÇÃO	3
METODOLOGIA	3
Dados e Amostra	3
Variáveis	4
Testes Estatísticos Aplicados	4
ESTATÍSTICA DESCRITIVA	5
Proporção Geral de Feridos Graves	5
Número de Feridos por Acidente	5
Probabilidades Condicionais	7
INFERÊNCIA ESTATÍSTICA E PODER ESTATÍSTICO	8
Testes de Normalidade e Homogeneidade	8
Testes Paramétricos	8
Testes Não Paramétricos	8
Testes de Proporções	9
Ajustes de Distribuição	9
Poder Estatístico e Tamanho Amostral	9
CONCLUSÃO	10
REFERÊNCIAS	10

INTRODUÇÃO

O presente relatório estatístico visa avaliar a existência de alguma diferença significativa entre os *veículos elétricos (EV)* e *veículos não elétricos (frota geral)* quanto à gravidade dos acidentes – sendo mensurada por meio de: proporção de acidentes com feridos graves; e número médio de feridos por acidente.

Esta análise utilizará exclusivamente os dados presentes nos *datasets Motor Vehicle Collisions - Crashes* e *Electric Vehicle Population Data*, que totalizam 2.223.272 acidentes, dos quais 724 envolvem veículos elétricos.

Diante desse contexto, a presente análise concentra-se em testar se a proporção de acidentes com feridos graves envolvendo veículos elétricos difere daquela observada na frota geral. Para isso, formula-se a hipótese nula (H_0), que estabelece igualdade entre as proporções ($p_{EV}=p_{geral}$), e a hipótese alternativa (H_1), que assume que tais proporções não são iguais ($p_{EV}\neq p_{geral}$). As inferências estatísticas serão conduzidas considerando um nível de significância de 5% ($\alpha=0,05$).

METODOLOGIA

As análises foram estruturadas em cinco blocos principais, sendo estes os seguintes – como enumerados abaixo:

1. Estatística Descritiva
2. Testes de Normalidade e Homogeneidade
3. Probabilidade e Teorema de Bayes
4. Inferência Estatística (testes paramétricos e não-paramétricos)
5. Poder Estatístico e Tamanho de Amostra

Dados e Amostra

Assim, foram separados os dados da amostra. Abaixo, seguem o total de acidentes, a quantidade de acidentes – dos quais os veículos elétricos estão envolvidos – e a quantidade, também em porcentagem, de veículos não elétricos.

- Total de acidentes: **2.223.272**
- EVs: **724 (0,03%)**

- Não-EVs: **2.222.548 (99,97%)**

Sabe-se que, a desproporção observada entre a quantidade de acidentes entre cada tipo de veículo é um fator crítico que compromete de forma direta o *poder estatístico* dos testes que visam analisar/comparar as proporções.

Variáveis

Abaixo, seguem os tipos e as principais variáveis utilizadas durante a análise (Tabela 1).

Tabela 1: Variáveis.

Tipo	Variável	Descrição
Exposição	is_EV	1 = elétrico, 0 = não elétrico
Desfecho	ferido_grave	1 = ferido grave, 0 = sem ferido grave
Contagem	n_persons_injured	nº de feridos por acidente

Fonte: Própria dos autores, 2025.

Testes Estatísticos Aplicados

Com base no que fora passado pela professora, durante o curso das aulas, foram feitos os seguintes testes – a fim de garantir que as hipóteses levantadas pudessem ser analisadas corretamente.

- Normalidade: **Shapiro-Wilk**
- Homogeneidade: **Levene, Bartlett, Fligner-Killeen, F-test**
- Proporções: **Teste z, Qui-quadrado de independência**
- Médias: **Teste t tradicional, Teste t de Welch**
- Não paramétrico: **Mann-Whitney U**
- Ajustes de distribuição: **Poisson, Binomial Negativa**

- Poder estatístico: **cálculo via efeito observado (3,1%)**

ESTATÍSTICA DESCRITIVA

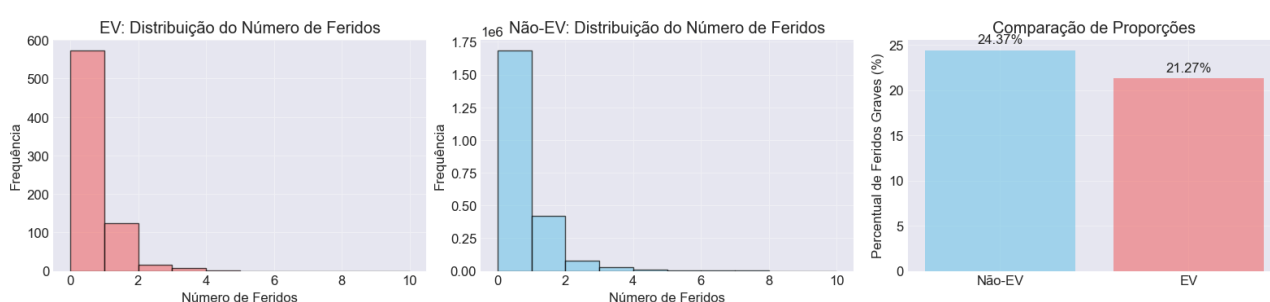
Proporção Geral de Feridos Graves

A análise descritiva – da variável de desfecho *ferido_grave* – mostra que, ao observar em todo o conjunto de dados, a proporção geral de feridos graves equivale a 24,37%; em volume absoluto de casos, esse valor equivale a cerca de 541.934 (quinhentos e quarenta e um mil, novecentos e trinta e quatro) – conforme detalhado na Imagem 1.

É possível observar na Imagem 1 a comparação das proporções de feridos graves entre os grupos definidos. Assim sendo, é visto que na frota geral (a saber, veículos não elétricos) os feridos classificados como graves é de 24,37%, enquanto nos acidentes que envolvem, especificamente, os veículos elétricos, a proporção é de 21,27%.

Essa diferença de 3,10 pontos percentuais (3,1%) sugere que a proporção de feridos graves é numericamente inferior nos acidentes com veículos elétricos. Este valor será o ponto central da etapa de *inferência estatística*, onde será determinada se essa diferença é significativa estatisticamente para rejeitar a hipótese nula ($H_0: p_{EV} = p_{geral}$) ou se é resultado apenas da variabilidade amostral.

Imagem 1: gráfico de proporção geral de feridos graves.



Fonte: Própria dos autores, 2025.

Número de Feridos por Acidente

Seguindo com a análise descritiva, observa-se na Tabela 2 e na Imagem 2 (com os boxplots das distribuições) que a Mediana e a Moda iguais a zero em ambos os grupos indicam que a maioria dos acidentes não apresenta feridos — algo visível pela forte

concentração de valores na base dos gráficos. Ainda assim, é possível notar que a média de feridos por acidente é menor nos veículos elétricos ($\bar{x}_{EV} = 0,2624$) em comparação à frota geral ($\bar{x}_{\text{Não-EV}} = 0,3286$), sugerindo que acidentes envolvendo veículos elétricos tendem a resultar em uma menor quantidade de feridos. Esse resultado será explorado na etapa de comparação entre as médias.

Outro ponto é a evidente *não normalidade* das distribuições, confirmada pelos elevados valores de *Assimetria* e *Curtose*, especialmente no grupo de veículos não elétricos (*Curtose* = 46,6660; *Assimetria* = 4,0848). Esse comportamento, típico de variáveis de contagem com predominância de zeros, inviabiliza o uso de testes paramétricos tradicionais, como o teste t, reforçando a necessidade de empregar métodos não paramétricos – como, por exemplo, Mann-Whitney U – ou modelos específicos para dados de contagem, como Poisson ou Binomial Negativa.

Ademais, observa-se que a variabilidade é maior na frota geral ($s = 0,7151$) do que no grupo de veículos elétricos ($s = 0,5835$), indicando maior dispersão no número de feridos por acidente entre veículos não elétricos. Essa diferença de variâncias precisa ser considerada na etapa inferencial e, caso se opte por um teste paramétrico, o teste t de Welch mostra-se mais adequado por não assumir homogeneidade de variâncias. Em síntese, os resultados descritivos sugerem menor média e menor variabilidade de feridos em EVs, ao passo que reforçam a não normalidade como o principal desafio metodológico da análise.

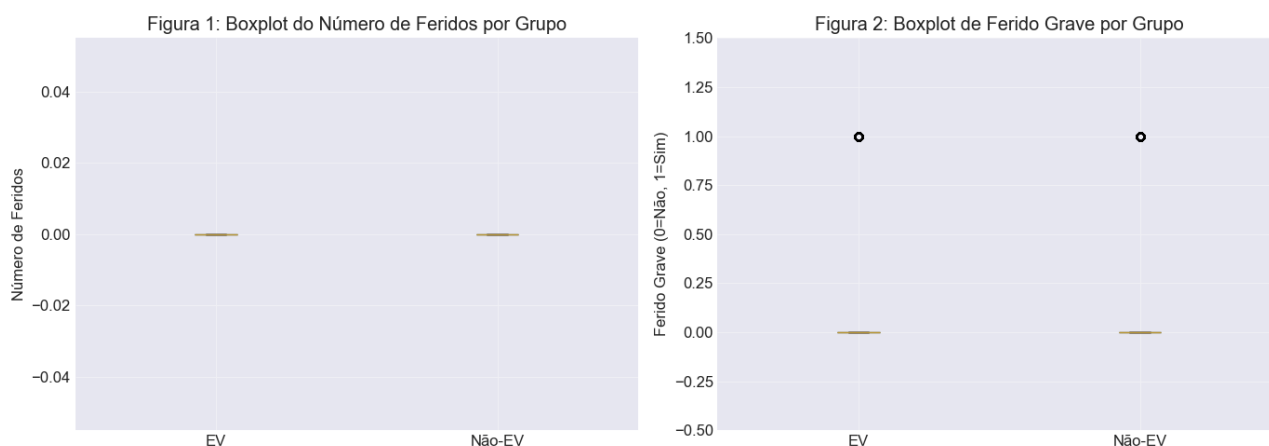
Tabela 2: Dados visualizados no boxplot.

Métrica	EV (n=724)	Não-EV (n=2.222.548)
Média	0,2624	0,3286
Mediana	0	0
Moda	0	0
Variância	0,3404	0,5114
Desvio-padrão	0,5835	0,7151

Assimetria	2,7672	4,0848
Curtose	9,3564	46,6660

Fonte: Própria dos autores, 2025.

Imagem 2: Figuras dos boxplots do número de feridos por grupo.



Fonte: Própria dos autores, 2025.

PROBABILIDADE

Probabilidades Condicionais

A análise das Probabilidades Condicionais permite detalhar como a ocorrência de feridos graves varia de acordo com o tipo de veículo envolvido no acidente. De forma geral, a probabilidade de um acidente qualquer resultar em um ferido grave, $P(\text{Ferido Grave})$, é de 0,2437 (24,37%). Já a probabilidade a priori de um veículo envolvido ser elétrico, $P(\text{EV})$, é extremamente baixa — apenas 0,0003 (0,03%) — refletindo a baixa participação dos EVs dentro do total de 2.223.272 acidentes analisados.

Ao observar as probabilidades condicionais, nota-se que a proporção de acidentes com feridos graves é menor quando o veículo é elétrico. A probabilidade de haver um ferido grave dado que o veículo é EV, $P(\text{Ferido Grave}|\text{EV})$, é de 0,2127 (21,27%), enquanto para veículos não elétricos, $P(\text{Ferido Grave}|\text{Não-EV})$, o valor é de 0,2437 (24,37%). A diferença de -3,10 pontos percentuais indica que, proporcionalmente, acidentes envolvendo EVs apresentam uma ocorrência ligeiramente menor de feridos graves em comparação com a frota geral.

Aplicando-se o Teorema de Bayes para estimar a probabilidade de um veículo ser elétrico dado que houve um ferido grave, $P(EV | \text{Ferido Grave})$, obtém-se:

$$P(EV | FG) = [P(FG | EV) \times P(EV)] / P(FG) = (0.2127 \times 0.0003) / 0.2437 \approx 2.62 \times 10^{-4} \approx 0.0003$$

O resultado, 0,0003 (0,03%), é idêntico ao valor da probabilidade a priori. Isso indica que saber que houve um ferido grave não altera a chance de o veículo envolvido ser elétrico. A razão é simples: a participação dos EVs na base de dados é extremamente pequena (0,03%), de modo que o evento “ferido grave” não contém informação suficiente para modificar a probabilidade inicial de a ocorrência envolver um EV.

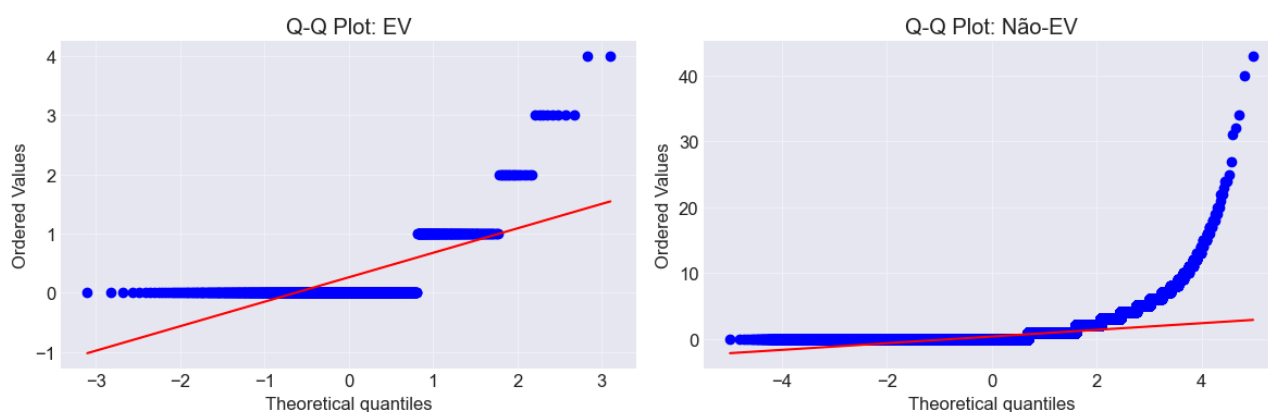
INFERÊNCIA ESTATÍSTICA E PODER ESTATÍSTICO

Nesta etapa, são apresentadas as análises inferenciais realizadas com o objetivo de avaliar se as diferenças observadas entre veículos elétricos (EV) e veículos não elétricos (frota geral) — quanto à gravidade dos acidentes — podem ser atribuídas ao acaso ou representam evidências estatisticamente significativas.

Testes de Normalidade e Homogeneidade

Inicialmente, avaliou-se a distribuição da variável de contagem *n_persons_injured*. Os resultados do teste de Shapiro–Wilk indicaram violação da normalidade tanto para EVs ($W = 0,5000$; $p \approx 0$) quanto para não-EVs ($W = 0,5041$; $p \approx 0$), confirmando que a variável não segue distribuição normal.

Imagem 3: Q-Q Plot.



Fonte: Própria dos dos autores, 2025.

Em seguida, foram aplicados diferentes testes de homogeneidade de variâncias (Levene, Bartlett, Fligner-Killeen e F-test), todos apontando para a heterogeneidade entre os grupos. Dessa forma, tanto a não normalidade quanto a desigualdade das variâncias orientaram a escolha por testes paramétricos robustos e testes não paramétricos.

Testes Paramétricos

Apesar da violação das premissas de normalidade, realizou-se inicialmente o teste t tradicional para fins comparativos. Os resultados indicaram diferença significativa entre as médias ($t = -2,4878$; $p = 0,0129$).

Considerando a heterogeneidade das variâncias, aplicou-se o teste t de Welch, mais apropriado nesse cenário, o qual reforçou a significância estatística da diferença ($t = -3,0487$; $p = 0,00238$).

Conclui-se, portanto, que veículos elétricos apresentam, em média, menor número de feridos por acidente.

Testes Não Paramétricos

Devido às características da distribuição dos dados, utilizou-se o teste de Mann–Whitney U. O resultado ($U = 775.376.369$; $p = 0,0237$) indica que as distribuições dos números de feridos diferem significativamente entre EVs e não-EVs, corroborando os achados dos testes paramétricos robustos.

Testes de Proporções

Para investigar a diferença na proporção de acidentes com feridos graves (*ferido_grave* = 1), aplicou-se o teste z para duas proporções. O resultado ($z = -1,9434$; $p = 0,05197$) indica que, ao nível de significância de 5%, não se rejeita a hipótese nula de igualdade entre as proporções.

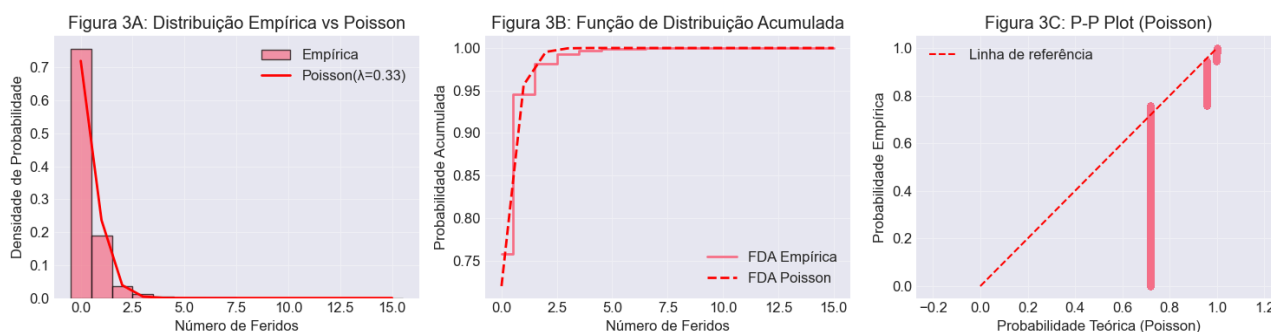
O teste qui-quadrado de independência confirmou essa conclusão ($\chi^2 = 3,610$; $p = 0,0574$), também não detectando associação estatisticamente significativa entre o tipo de veículo e a ocorrência de feridos graves.

Apesar disso, observa-se que a proporção de feridos graves em EVs (21,27%) é numericamente menor do que na frota geral (24,37%), correspondendo a uma diferença de 3,10 pontos percentuais.

Ajustes de Distribuição

Para compreender melhor o comportamento da variável de contagem, ajustaram-se modelos de Poisson e Binomial Negativa. O modelo de Poisson apresentou ajuste inadequado ($p \approx 0$), evidenciando sobredispersão (variância/média = 1,556). Em contraste, a distribuição Binomial Negativa apresentou parâmetros ajustados de forma consistente ($n = 0,5904$; $p = 0,6425$), mostrando-se mais apropriada aos dados.

Imagem 4: Figuras de distribuição.



Fonte: Própria dos autores, 2025.

Poder Estatístico e Tamanho Amostral

O poder estatístico do teste de comparação de proporções — considerando o efeito observado de 3,10% — foi calculado em 0,4931 (49,31%), valor inferior ao mínimo recomendado de 80% para inferências robustas.

Com base nisso, estimou-se o tamanho amostral necessário de veículos elétricos para detectar a diferença observada com poder adequado: aproximadamente 2.874 EVs seriam necessários. Entretanto, a base de dados conta com apenas 724 EVs, quantidade insuficiente para garantir alta capacidade de detecção do efeito.

Essa limitação explica por que a diferença entre as proporções, embora numericamente relevante, não atingiu significância estatística. A reduzida participação de EVs na amostra compromete diretamente o poder dos testes, impedindo a rejeição da hipótese nula.

CONSIDERAÇÕES FINAIS

Os resultados obtidos ao longo da análise demonstram que acidentes envolvendo veículos elétricos (EV) apresentam, em média, menor número de feridos por ocorrência

quando comparados à frota geral. Essa diferença foi confirmada de forma consistente pelos testes paramétricos robustos (t de Welch) e pelo teste não paramétrico de Mann–Whitney U, mesmo diante da clara violação das premissas de normalidade e homogeneidade das variâncias. Além disso, o ajuste dos dados a modelos específicos para variáveis de contagem evidenciou que a distribuição Binomial Negativa descreve adequadamente o comportamento do número de feridos, ao passo que o modelo de Poisson mostrou-se inadequado devido à sobredispersão presente.

No que se refere à proporção de acidentes com feridos graves, embora os veículos elétricos apresentem um percentual numericamente inferior (21,27%) quando comparados aos demais veículos (24,37%) — diferença de 3,10 pontos percentuais — os testes de hipóteses aplicados (teste z para duas proporções e qui-quadrado de independência) não forneceram evidências suficientes, ao nível de significância de 5%, para rejeitar a hipótese nula de igualdade entre as proporções. A ausência de significância estatística é explicada principalmente pelo baixo poder estatístico do teste ($\approx 49\%$), diretamente influenciado pela baixa representatividade de veículos elétricos na base de dados: apenas 724 ocorrências, correspondendo a 0,03% do total analisado.

Diante desse cenário, destaca-se que a amostra atual de EVs não oferece capacidade suficiente para detectar, com segurança, a diferença observada entre as proporções. As estimativas mostram que seriam necessários cerca de 2.874 registros envolvendo veículos elétricos para alcançar o poder estatístico adequado (80%) e permitir conclusões mais robustas. Assim, apesar de os EVs apresentarem melhor desempenho em relação ao número médio de feridos, não é possível afirmar, com evidência estatística suficiente ao nível de 5%, que a proporção de acidentes com feridos graves difere significativamente entre veículos elétricos e a frota geral.

Recomenda-se que estudos futuros contemplem bases de dados com maior representatividade de EVs e incorporem modelos estatísticos mais robustos — como regressões de contagem e análises multivariadas — capazes de controlar potenciais variáveis de confusão, tais como ano, tipo de colisão e características ambientais. Somente com um conjunto de dados mais equilibrado e métodos mais avançados será possível produzir inferências conclusivas sobre a relação entre tipo de veículo e gravidade dos acidentes.

REFERÊNCIAS

MOTOR VEHICLE COLLISIONS - CRASHES. Dataset. *Data.gov*. Disponível em: <https://catalog.data.gov/dataset/motor-vehicle-collisions-crashes>. Acesso em: 2 dez. 2025.

ELECTRIC VEHICLE POPULATION DATA. Dataset. *Data.gov*. Disponível em: <https://catalog.data.gov/dataset/electric-vehicle-population-data>. Acesso em: 2 dez. 2025.