

UMA ANÁLISE MULTIVARIADA DE DADOS SOB A ÓTICA DE ACIDENTES OCORRIDOS NA BR 304-RN

Mileno Alexandre Barbosa Epifânio

Existe uma correlação positiva entre o aumento da motorização da população e o número de acidentes de trânsito. Como resultado, milhares de pessoas morrem anualmente no Brasil; e, outras tantas ficam com sequelas. Além do sofrimento pessoal, os acidentes geram um custo aos cofres públicos. Considerando o aumento no número de veículos e de acidentes registrados, essa pesquisa busca analisar, por meio de uma análise multivariada, através da seleção de variáveis com a Regressão Múltipla e a Clusterização. Os dados relativos aos acidentes entre 2017 e 2021 ocorridos na BR 304-RN por meio de uma pesquisa com características: aplicada, descritiva, qualitativa e método de modelagem e simulação. Para isso, a motivação que norteia essa pesquisa é a necessidade de auxiliar na diminuição de mortes ocorridas em rodovias, contribuindo para os acessos básicos da população serem realizados de uma forma mais segura. Como resultado, foi encontrado que a base de dados pode ser classificada em 7 clusters. Analisando somente o Cluster 1(CL1), que representou a maior taxa de mortalidade visualizada (12% de incidência de óbitos), o dia mais letal foi nas quartas-feiras, tendo destaque a cidade de Santa Maria, durante a fase do amanhecer e noite, com causas com maiores incidências de mortes associadas a entrada inapropriada de pedestres e uso de entorpecentes. A partir dos resultados encontrados, os gestores podem se concentrar em implementar medidas para diminuição desses acidentes, como aumento do policiamento, inserção de faixas de pedestres, aumento de radares e abertura de postos como SAMU e Bombeiros.

Palavras-chave: Engenharia de Produção. Pesquisa Operacional. Regressão múltipla. Clusterização. Acidentes em rodovias.

1. Introdução

Com o avanço da tecnologia e mobilidade urbana, houve também um aumento nas formas de locomoção, sendo exemplo a popularização dos automóveis, que já passam dos 1.5 bilhões de carros e caminhões em todo o planeta, com aproximadamente 8 milhões nas categorias puramente elétricos e híbridos plugáveis (ESTADÃO, 2023; OLIVEIRA, 2023). Por outro lado, somente no Brasil, existem cerca de 100 milhões no somente no Brasil (PAIXÃO, 2022).

Como consequência desse aumento, a cada 24 segundos, é registrado uma morte no trânsito, com média de 1 milhão e 350 mil mortes anuais (ONU, 2021). Somente no Brasil, ocorrem em média 3 mortes por hora no trânsito, sendo a oitava maior causa de morte no nosso país (BIERNATH, 2021), com aumento de 12% nos acidentes em vias federais entre os anos de 2017 e 2021, segundo os dados abertos da PRF, a Polícia Rodoviária Federal do Brasil (RPF, 2022).

Na granularidade do Estado do Rio Grande do Norte (RN), a frota de automóveis visualizou um aumento de mais de 100% entre os anos de 2008 a 2018 (SILVESTRE, 2019). Em comparação dentre os dados abertos relativos aos anos de 2017 a 2021, foi lavrado um aumento de 26% nos acidentes, variando entre 3% e 5% de óbitos.

Somente na BR 304, que é o objeto deste estudo e é a única ligação entre as duas maiores cidades do estado, Natal e Mossoró, ocorrem 25% dos acidentes registrados, com 4% de acidentes resultante em óbito nos anos analisados em média no mesmo período (BRASIL, 2022). No ano de 2022, a BR 304 foi a que deteve o maior percentual de mortes das BR's que cruzam o estado nos primeiros cinco meses do ano, registrando 27% do total de óbitos e é alvo de debates acerca da sua duplicação no futuro (JÁCOME, 2022).

Como consequência dessa realidade, o Senado Nacional, divulgando uma pesquisa do Instituto de Pesquisa Econômica Aplicada (IPEA), cita que as cerca de 45 mil mortes anuais foram responsáveis por R\$ 50 bilhões a menos aos cofres públicos anualmente (BRASIL, 2022).

Nesta temática, existem algumas produções já realizadas na área, que são consideradas para elaborar o critério competitivo deste artigo. Nicácio e Lacerda (2021), Amorim (2019) utilizam uma abordagem semelhante, não havendo a abordagem de possíveis soluções. Enquanto isso, Barduchi (2018) praticou uma análise com base com os mesmos objetivos, mas somente utilizando as variáveis de rede pavimentada, população, frota, investimentos e quantidade de frota de fiscalização, somente entre os anos de 2010 a 2014, e não abordou tópicos para a solução dos problemas visualizados. Por fim, Nunes et al. (2022) realizou a clusterização com base em variáveis, mas sem entrar em uma possível solução para os problemas abordados.

Com isso, define-se que o critério de avanço nas discussões sobre o tema que esta produção acadêmica trará é a análise focada em um único estado, na granularidade da unidade das rodovias, criando consigo a abordagem de possíveis soluções.

A partir das estatísticas, pode-se inferir que a grande quantidade de acidentes é onerosa para a população, pois acarreta perdas de vidas e outras pessoas podem ter sequelas para o resto da vida, sendo necessárias medidas para sua resolução (OMS, 2022). Por tratar-se de um problema complexo, é necessário um estudo acerca de quais variáveis podem ser mais impactantes para os resultados visualizados no referido local. Por isso, o seguinte trabalho tem como objetivo analisar, por meio de uma regressão linear múltipla e clusterização, quais variáveis mais impactam a taxa de mortalidade de ocorrência na BR-304.

Portanto, a pergunta que norteia esta pesquisa é *“Quais variáveis são mais relevantes para a criação de soluções que impactem positivamente na prevenção acidentes com vítimas fatais em acidentes ocorridos no trecho norterio-grandense da BR 304?”*.

O objetivo deste artigo é analisar quais variáveis impactam mais em óbitos ocorridos na rodovia supracitada, por meio da análise de regressão múltipla focada na seleção das variáveis para o processo de clusterização. Do ponto de vista prático, espera-se auxiliar os órgãos públicos nas decisões acerca de melhorias a serem realizadas no trecho analisado com uma análise baseada em dados históricos, diminuindo consigo tanto a quantidade de mortes, quanto os recursos financeiros que são, anualmente, disponibilizados nestas operações de maneira paliativa. Para isso, é necessária uma pesquisa aplicada, descritiva e quantitativo por meio de modelagem e simulação. Portanto, esta pesquisa foi realizada em 4 etapas e utilizados dados abertos acerca dos acidentes ocorridos no referido local e disponibilizado pela PRF.

Esta produção acadêmica está abordada em 6 seções: (1) esta seção tem caráter introdutório; (2), o arcabolo teórico, sobre análise multivariada de dados, regressão múltipla, optimal scalling e seleção forward e clusterização; (3) os procedimentos metodológicos para a execução deste trabalho; (4) resultados e discussões; (5) as considerações finais acerca do tema e, por fim, (6), as referências bibliográficas.

2. Referencial Teórico

Para o pleno entendimento deste trabalho, é necessário abordar alguns tópicos teórico que embasam o processo de criação do mesmo. Com isso, se faz necessário abordar alguns tópicos durante o referencial, que são os de análise multivariada de dados, regressão múltipla, *Optimal Scalling* e Seleção *Forward* e Clusterização, respectivamente 2.1, 2.2, 2.3 e 2.4.

2.1 Análise multivariada de dados

Loesch e Hoeltgebaum (2012, p. 4) descrevem uma processo de Análise Multivariada de Dados (AMD) como um conjunto de técnicas que utilizam da estatística para realizar uma “análise de dados envolvendo mais de duas variáveis em uma amostra de observações. Emprega, com frequência, a análise exploratória de dados para identificar relações sistemáticas entre variáveis”. Enquanto isso, Hair et al. (2009, p. 23) descrevem uma análise multivariada como “todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação”.

Loesch e Hoeltgebaum (2012) afirmam que a existência de várias técnicas dentro de AMD se dá baseado em dois fatores, que são: a existência, ou não, de variáveis que são independentes; Tipo de variável, podendo ser quantitativa (numéricas) ou qualitativas (categóricas).

2.2 Regressão Múltipla

O modelo de regressão múltipla é uma técnica estatística que se assemelha ao visualizado na regressão simples, que segundo Favero (2015), busca entender o comportamento de uma variável (y), chamada de dependente, em relação as variáveis explicativas (x). No caso da regressão múltipla, essa variação da variável independente é visualizada tendo em base mais de uma variável x no modelo estudado (HAIR et al., 2009).

Como resultado deste processo, são obtidos os pesos que denotam a contribuição associada a cada variável independente na previsão, facilitando a interpretação de influência (HAIR et al., 2009). Matematicamente, Chein (2019) o cálculo de regressão múltipla pode ser visualizado como:

$$Y_c = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + e_i$$

Onde,

Y_c = valor previsto;

b_0 = intercepto estimado;

$b_1 + b_2 + \dots + b_n$ = coeficientes angulares estimados;

$x_1 + x_2 + \dots + x_n$ = variáveis;

e_i = erro aleatório associado.

No caso desta produção acadêmica, essa regressão será crucial para entender quais variáveis mais impactam para que o acidente resulte em morte.

2.3 Optimal Scalling e Seleção Foward

Esse método é utilizado dentro do *software* SPSS, da empresa americana IBM, visando atribuir, por meio de um modelo, a quantificação numérica às variáveis categóricas presentes na base de dados utilizada por meio de um critério de otimização do procedimento utilizado (MEULMAN; HEISER, 2005).

Esse processo de atribuição é realizado tendo como base no critério de otimização do procedimento que será utilizado, tendo propriedades próprias, onde essa metodologia é bastante recomendada para usuários que detém de bases com poucas observações, muitas variáveis ou muitos valores por variável (MEULMAN; HEISER, 1999).

Enquanto isso, o método de seleção Forward é um dos mais referenciados da literatura e tem como objetivo a entrada de uma variável no modelo como critério baseado na significância estatística (p) da mesma no conjunto, caso o valor seja abaixo do aceito, a variável contínua no modelo, e caso seja abaixo, a variável é retirada do modelo e é necessário reprocessamento sem a mesma (IBM, 2012; MASIERO; ANZANELLO; GUSBERTI, 2012).

2.4 Clusterização

Análise de agrupamentos como um processo que visa a utilização de um conjunto voltado às técnicas exploratórias que buscam verificar a existência de comportamentos semelhantes entre os objetos estudados (indivíduos, empresas, municípios, países, consumidores) em relação a determinadas variáveis, visando a criação de grupos (clusters), em busca de uma maior taxa de homogeneidade interna dentro desse cluster (FAVERO, 2015).

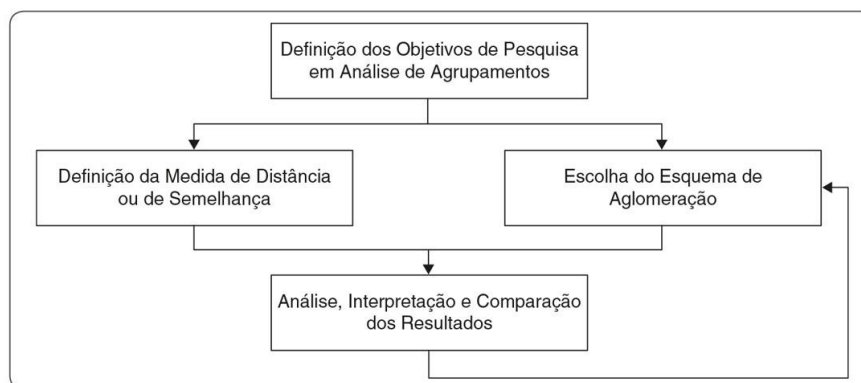
Segundo Hair et al. (2009), para que uma análise de agrupamentos seja viável, são necessários três passos considerados essenciais, que são:

- Medida de alguma forma de similaridade ou associação entre as entidades envolvidas na análise, visando determinar quantos grupos realmente existem dentro da amostra selecionada;
- Processo de agrupamento, onde as entidades são particionadas em agrupamentos/clusters;
- Estabelecer o perfil das pessoas ou variáveis para determinar sua composição interna. Durante esse processo, é crucial entender que o resultado será formado por pequenos agrupamentos, compostos por unidades heterogêneas entre si, que conseguem representar o comportamento conjunto das observações a partir de determinadas variáveis determinadas no início da análise (FAVERO, 2015).

Às técnicas voltadas a análise de agrupamento são consideradas como exploratórias, também chamadas de interdependência, porque as aplicações realizadas não representam um caráter

preditivo, já que novas variáveis podem entrar na análise e resultar em uma realocação de agrupamentos já existentes, alterando todo o resultado até então (FAVERO, 2015). Para o processo de lógica de criação de uma análise de agrupamentos, Favero & Belfiore (2022) descreve como um processo cíclico, que é representado na Figura 1.

Figura 1: Criação de uma análise de agrupamentos



Fonte: Favero (2015).

Por fim, Loesch e Hoeltgebaum (2012) descrevem que o resultado do processo de clusterização pode ser sintetizado por meio de um dendrograma, que é uma representação de árvore classificatória para diversos níveis de agrupamento. Essa representação possibilita a visualização desde grupos mais gerais até subgrupos mais específicos, como se fossem as folhas da árvore. Porém, na construção de processos de clusterização, também é necessário um tratamento de base, voltada à análise e possíveis retiradas de *outliers* (FAVERO, 2015).

Considerando o contexto, na seção seguinte serão apresentados os procedimentos metodológicos que norteiam essa pesquisa.

3. Procedimentos metodológicos

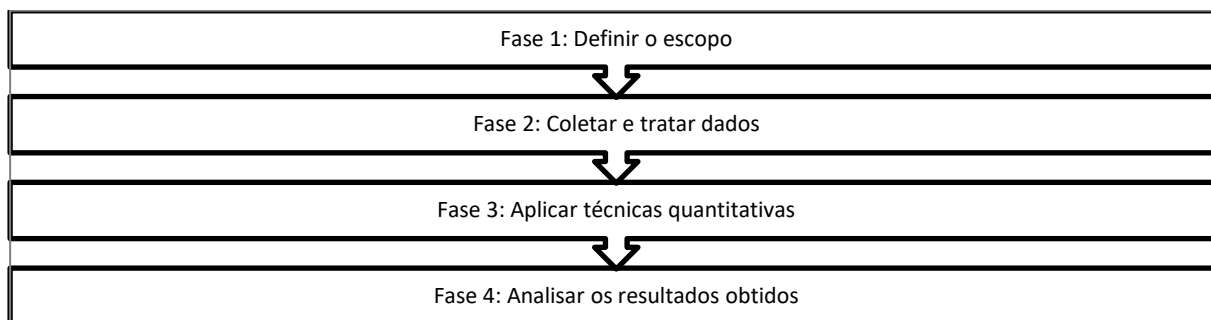
Segundo Miguel (2010), a metodologia de pesquisa que está inserida na área da Engenharia de Produção é estruturada em quatro pilares fundamentais: natureza, objetivos, abordagens e métodos. Em relação a natureza, a pesquisa se caracteriza como aplicada, pois busca uma aplicação real dos conhecimentos na solução de um problema prático (TURRIONI; MELLO, 2012).

Em relação a natureza, a pesquisa se caracteriza como aplicada, que busca uma validação real dos conhecimentos na solução de um problema prático (TURRIONI; MELLO, 2012). Quanto aos objetivos, o estudo se classifica como descritivo, que é quando o pesquisador descreve situações, contextos e fenômenos sem realizar nenhum processo de intervenção nos mesmos (SAMPIERI; COLLADO; LUCIO, 2013). A abordagem se dá no estilo quantitativo, que é um

tipo de pesquisa voltado a coleta de informação e tratamento das mesmas por meio de processos estatísticos (LOUZADA; NUNES, 2019).

Com base nos processos apresentados, as etapas desta pesquisa podem ser abordadas, conforme mostra a Figura 2.

Figura 2: Fases para elaboração do trabalho



Fonte: Autoria própria (2023).

Para solucionar cada fase descrita na Figura 2, serão necessários a realização de sub-tarefas alocadas dentro das fases, descritas nos sub-tópicos 3.1 a 3.4.

3.1. Definir o escopo

A primeira fase realizada foi a pesquisa e definição do tema, pensando no impacto que o projeto possa trazer a sociedade. Com isso, foi visualizado nas bases de dados abertas que os índices de acidentes tiveram uma progressão de aumento ao longo do tempo, segundo a PRF.

Além disso, houve a publicação de notícias informando acerca do aumento dos acidentes em rodovias do Rio Grande do Norte, em especial a 304. Portanto, visualizou-se a importância do tema para a comunidade, além dos procedimentos, que são análise de regressão múltipla e clusterização, e bases de dados necessários.

3.2. Coletar e tratar dados

A partir da coleta de dados, é possível conduzir pesquisas voltadas ao método de modelagem e simulação, onde o pesquisador, por meio de um computador, realiza pesquisas e manipula variáveis e níveis, mas não na realidade (MIGUEL, 2010). Esse método é utilizado quando se é necessário experimentar um sistema real por meio de um modelo, podendo determinar como as variáveis e níveis se comportam e como as mesmas irão responder às mudanças analisadas (TURRIONI; MELLO, 2012).

O banco de dados contém 5667 registros e está disponível para consulta no domínio governamental¹, que realizam a distribuição de dados públicos do nosso país. Foram

¹ <https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-acidentes>

selecionados dados referentes aos anos de 2017 a 2021 em *hall*, distribuídos em 14 variáveis para este estudo, distribuídos na Tabela 1. Essa etapa encerrou a preparação da base por meio da aplicação da técnica de *Optimal Scaling* no *software* SPSS, gerando a base final para realizar as análises necessárias. Por fim, foi realizado por meio da linguagem *Python*, necessitando a criação do resultado final obtido.

Tabela 1: Variáveis presentes no estudo

Variáveis	Descrição da variável
1 - Dia da semana	Dia da semana do ocorrido.
2 - Município	Município em que o acidente ocorreu.
3 - Causa do acidente	Causa do acidente ocorrido. Ex: mal súbito, ingestão de álcool e etc.
4 - Tipo do acidente	Identificação do tipo do acidente. Ex: Colisão, incêndio e etc.
5 - Fase do dia	Fase do dia no momento do acidente. Ex: anoitecer, pleno dia e etc.
6 - Condição meteorológica	Condições registrada na hora do acidente. Ex: Céu claro, chuva e etc.
7 - Tipo da pista	Tipo da pista no local do acidente. Ex: Simples, duplicada e etc.
8 - Traçado da via	Característica da via. Ex: Reta, curva e etc.
9 - Tipo do veículo	Classificação do veículo. Ex: motocicleta, automóvel, caminhão e etc.
10 - Ilesos	Se houve pessoas ilesas.
11- Feridos leves	Se houve pessoas com ferimentos leves.
12 - Feridos graves	Se houve pessoas com ferimentos graves.
13 - Mortes	Se houve pessoas mortas.

Fonte: Autoria própria (2023).

3.3. Aplicar técnicas quantitativas

Essa etapa consistiu no processo das técnicas quantitativas definidas na fase de coleta e tratamento de dados. Durante a aplicação do método de clusterização, foi utilizado o método de *seleção forward* e retirar os valores de $p > 5\%$ de relevância. Após isso, foi realizada o processo de clusterização com as variáveis consideradas influentes no resultado, visando descobrir quantos clusters podem ser formados com os dados encontradas, sendo possível seguir para o tópico de análise.

3.4. Analisar os dados obtidos

Por fim, foi realizada a análise dos dados obtidos no processo de criação deste trabalho, assim como geração de *insights* relativos a situação visualizada e problema real encontrado, que estão disponíveis no tópico 4 deste trabalho.

Durante a realização da análise, por meio da linguagem *Python*, foi realizado a análise dos resultados obtidos, que são entendidos como as descrições dos clusters. Por último, foi realizada

a análise do trecho, visando encontrar possíveis formas de adaptação e criação de atitudes que podem ser realizadas, visando a mitigação das estatísticas encontradas.

Após a aplicação destes tópicos, foi possível realizar a análise e visualizar os resultados da pesquisa, disponíveis no tópico 4.

4. Resultados e discussões

Após a aplicação dos passos descritos, foi possível chegar a um resultado favorável ao objetivo deste trabalho, que é identificar quais variáveis são mais decisivas em acidentes que resultam em usuários mortos em acidentes ocorridos na BR 304 do RN. Após a transformação utilizando a técnica de *Optimal Scaling*, foi encontrado o resultado para a regressão linear, presente na Figura 3.

Figura 3: Regressão Múltipla com todas as variáveis do estudo

OLS Regression Results						

Dep. Variable:	TRA13_1	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	4.499e+31			
Date:	Sat, 11 Feb 2023	Prob (F-statistic):	0.00			
Time:	00:55:10	log-likelihood:	1.7846e+05			
No. Observations:	5585	AIC:	-3.569e+05			
Df Residuals:	5572	BIC:	-3.568e+05			
Df Model:	12					
Covariance Type:	nonrobust					

	coef	std err	t	P> t	[0.025	0.975

Intercept	-4.993e-07	4.3e-17	-1.16e+10	0.000	-4.99e-07	-4.99e-07
TRA1_1	1.801e-15	4.4e-17	40.941	0.000	1.71e-15	1.89e-15
TRA2_1	1.608e-15	5.48e-17	29.320	0.000	1.5e-15	1.72e-15
TRA3_1	0.150e-16	4.94e-17	18.547	0.000	8.10e-16	1.01e-15
TRA4_1	-1.569e-15	5.04e-17	-31.122	0.000	-1.67e-15	-1.47e-15
TRA5_1	-7.331e-16	4.64e-17	-15.804	0.000	-8.24e-16	-6.42e-16
TRA6_1	2.234e-16	4.45e-17	5.024	0.000	1.36e-16	3.11e-16
TRA7_1	2.412e-16	5.21e-17	4.633	0.000	1.39e-16	3.43e-16
TRA8_1	-2.579e-17	4.49e-17	-0.575	0.565	-1.14e-16	6.22e-17
TRA9_1	-6.956e-16	5.47e-17	-12.709	0.000	-8.03e-16	-5.88e-16
TRA10_1	-2.3849	1.11e-16	-2.14e+16	0.000	-2.385	-2.385
TRA11_1	-2.3488	1.06e-16	-2.22e+16	0.000	-2.349	-2.349
TRA12_1	-1.7578	8.43e-17	-2.08e+16	0.000	-1.757	-1.757

Omnibus:	23.089	Durbin-Watson:	0.437			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22.472			
Skew:	-0.135	Prob(JB):	1.32e-05			
Kurtosis:	2.847	Cond. No.	5.95			

Fonte: Autoria própria (2023).

Como o valor relativo à variável 8 retornou uma significância de P acima de 5%, foi aplicado a técnica de *seleção forward*, retirando a variável e realizando novamente o processamento de dados, atingindo o resultado disponíveis na Figura 4.

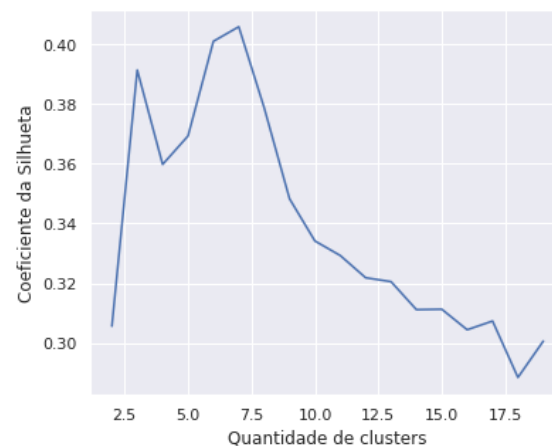
Figura 4: Regressão Múltipla sem a variável 8

OLS Regression Results						
=====						
Dep. Variable:	TRA13_1		R-squared:	1.000		
Model:	OLS		Adj. R-squared:	1.000		
Method:	Least Squares		F-statistic:	3.484e+3		
Date:	Sat, 11 Feb 2023		Prob (F-statistic):	0.0		
Time:	00:55:10		Log-Likelihood:	1.7750e+0		
No. Observations:	5585		AIC:	-3.550e+0		
Df Residuals:	5573		BIC:	-3.549e+0		
Df Model:	11					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975
Intercept	-4.993e-07	5.11e-17	-9.78e+09	0.000	-4.99e-07	-4.99e-07
TRA1_1	-1.089e-15	5.22e-17	-20.871	0.000	-1.19e-15	-9.86e-16
TRA2_1	5.408e-16	6.5e-17	8.324	0.000	4.13e-16	6.68e-16
TRA3_1	4.912e-16	5.86e-17	8.386	0.000	3.76e-16	6.06e-16
TRA4_1	3.817e-16	5.08e-17	6.383	0.000	2.64e-16	4.99e-16
TRA5_1	2.378e-16	5.49e-17	4.330	0.000	1.3e-16	3.45e-16
TRA6_1	-1.661e-15	5.27e-17	-31.502	0.000	-1.76e-15	-1.56e-15
TRA7_1	-3.246e-16	6.11e-17	-5.309	0.000	-4.44e-16	-2.05e-16
TRA9_1	1.563e-15	6.49e-17	24.081	0.000	1.44e-15	1.69e-15
TRA10_1	-2.3849	1.32e-16	-1.81e+16	0.000	-2.385	-2.385
TRA11_1	-2.3488	1.26e-16	-1.87e+16	0.000	-2.349	-2.349
TRA12_1	-1.7570	1e-16	-1.76e+16	0.000	-1.757	-1.757
=====						
Omnibus:	107.793		Durbin-Watson:	1.12		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	220.45		
Skew:	0.061		Prob(JB):	1.34e-4		
Kurtosis:	3.966		Cond. No.	5.8		

Fonte: Autoria própria (2023).

Como nenhuma variável retornou um valor de p acima do valor de 5%, não foi necessário realizar nenhuma retirada. Com base nos dados da Figura 04, foi possível averiguar e validar a quantidade ideal de agrupamentos, que são 7 clusters, que é o ponto mais alto do coeficiente da Silhueta visualizado, como demonstrados na Figura 5.

Figura 5: Análise da quantidade de clusters



Fonte: Autoria própria (2023).

Com base no apresentado, é possível analisar qual cluster recebeu o maior índice de mortalidade, onde é possível visualizar na Tabela 2.

Tabela 2: Índice de % de mortalidade em casa cluster

CLUSTER	MORTES	TOTAL DE CASOS	% DE MORTALIDADE
1 (CL1)	78	660	12%
2 (CL2)	26	1101	2%
3 (CL3)	66	1433	5%
4 (CL4)	23	557	4%
5 (CL5)	29	979	3%
6 (CL6)	13	298	4%

7 (CL7)	15	556	3%
TOTAL	250	5584	4%

Fonte: Autoria própria (2023).

Com base na porcentagem de mortes registrada, os *insights* e análises serão focados em CL1 pela maior porcentagem de mortalidade em comparação com os outros clusters. Portanto, as informações que são tabuladas levando em consideração o critério de óbito por acidentes nos indicam que:

- Dentre os tipos de acidentes, colisão frontal, Saída de leito carroçável e atropelamento de pedestre registraram 24%, 18% e 12%, nesta ordem;
- Foi registrado 58% e 19% dos acidentes com óbito ocorreram nas fases do Amanhecer e Plena Noite;
- O dia da semana mais letal foi nas quartas feiras e sábados, com 22% dos ocorridos em cada dia, seguido por domingo, com 18%;
- A condição metereológica que registrou a maior quantidade de acidentes com vítimas fatais é a de sol claro, seguida pelo Nublado, com 81% e 14%, respectivamente;
- As cidades de Açu (23%), Angicos (15%), Macaíba (13%) e Santa Maria (13%) foram as que registraram o maior índice de acidentes com óbitos quando é visualizado na granularidade de cidade;
- A pista de tipo simples foi a que mais registrou ocorrências, com 97% dos casos;
- Quando visualizamos na ótica das causas de acidentes, os motivos de “Falta de Atenção à Condução”, “Desobediência às normas de trânsito pelo condutor” e “Ultrapassagem Indevida” registraram maior índices de mortes, com 22%, 19% e 10%, respectivamente.

Entretanto, pela importância do tema e por lidar com acidentes em status de vítimas fatais, não se pode deixar de citar variáveis que se destacaram de maneira distinta em outros clusters desta análise. Entre os clusters 4 e 7, o turno de pleno dia foi o mais letal, variando entre os valores de 61% a 87%. Colocando uma lupa sobre mortalidade em um dia específico da semana, o cluster 6 obteve o status recorde, também na quarta feira (67%). Quando a ótica é voltada a cidade mais letal, Mossoró se destaca entre os clusters 2 a 7, sendo o recorde visualizado no cluster 5, com 52% dos obitos ocorrendo nessa cidade. Por fim, sob a visão da condição metereológica, o status de Nublado aparece na maioria em segundo lugar;

Portanto, com base nos dados apresentados, é crucial a abordagem de processos que podem melhorar a situação relativa as ocorrências de acidentes com vítimas fatais na BR 304-RN. Para isso, existem algumas ações que podem ser tomadas e algumas já foram, inclusive, citadas na

literatura semelhante, como a inserção de passarelas nas cidades, garantindo que o pedestre não necessite passar pelo trecho na mesma altura do veículo (MORAIS, 2017). Com isso, também como colaboração deste trabalho, os autores indicam formas para que este mesmo propósito seja atendido nos tópicos a seguir.

- Aumento de policiamento militar entre as cidades, principalmente no turno do amanhecer, em uma forma de parceria contínua entre as áreas urbanas, visando diminuir a taxa de acidentes provenientes a uso de álcool, entorpecentes e velocidade acima do permitido, principalmente durante o fim de semana e dias com festividades nas cidades que estão próximas a BR 304;
- Adição de radares de velocidade na modalidade fixas e móvel, nos perímetros urbanos das cidades que mais representaram % de mortes;
- Processos de sinalização de proibidas ultrapassagens, com adição de outdoors e outros tipos de sinalização de trechos perigosos no trajeto, assim como novos modelos de sinalização de trechos perigosos;
- Campanhas digitais incentivando o uso de aplicativos de deslocamento entre cidades que já tem essa operação, além de incentivos para que usuários sigam as regras de trânsito;
- Manutenção das *guard rail* presentes no trajeto, assim como revitalização dos já existentes, visando que caso ocorra algum acidente, o usuário não saia da pista e tenha um processo de primeiros socorros mais eficientes;
- Abertura e incentivo de novos locais de atendimento para o SAMU (Serviço de Atendimento Móvel de Urgência) ou Pronto Socorro em uma cidade estratégica do trecho, como Lages ou São Paulo do Potengi, que mesmo não estando dentro da BR 304-RN, pode ser um local de rápida movimentação e chegada a locais de atendimento.

Com base nesses *insights*, espera-se que seja registrado uma diminuição de pelo menos 25% das mortes registradas em Clusters com a mesma característica que do analisado. Isso representa aproximadamente 20 situações envolvendo mortes na respectiva BR.

5. Considerações Finais

Este trabalho acadêmico teve como objetivo realizar uma análise multivariada de dados nos dados históricos dos acidentes visualizados na BR 304, localizada no Rio Grande do Norte. A rodovia 304-RN liga as cidades de Natal a Mossoró, as duas principais cidades do referido

estado. Foi realizada uma pesquisa de cunho aplicada, descritivo, quantitativo e modelagem e simulação, com uso de dados abertos da PRF entre os anos de 2017 e 2021.

Foi realizado uma análise multivariada de dados para encontrar as variáveis mais significativas, assim como clusterização do resultado encontrado, resultando em 7 clusters formados por meio das características dos acidentes que resultaram em óbitos. Após essa etapa, foi escolhido o CL1, por ter 12% de incidência de óbitos.

Dentre as variáveis inicialmente alocadas para a realização deste trabalho, somente o “Traçado da via” não obteve um valor de p menos que 5%. Como contribuições práticas estão os períodos de Amanhecer é crucial, com dia mais letal na quarta-feira, com a causa mais fatal estando relacionada a falta de atenção, além de Açu sendo considerada a cidade mais letal. Como formas de insights de minimização de problemas, foram citadas atitudes que podem ser tomadas pelo poder público visando minimizar o problema, como patrulhamento direcionado a determinados dias, criação de faixas de pedestres, instalação de radares fixos e móveis, manutenção e criação de mais *guard rail*, campanhas digitais de incentivo e etc.

Por fim, pode-se concluir que os objetivos propostos foram atendidos de maneira satisfatória, sendo possível realizar as aplicações e na forma de indicação de melhorias ao poder público. Como sugestão para trabalhos futuros, é possível criar uma aplicação teste das soluções em determinados dias, aumentando a segurança da BR-304 e visando a validação dos direcionais, assim como a criação de um modelo multicritério que indique a fila de execução das soluções aqui abordadas.

6. Referências

- AMORIM, B. de S. P.; **Uso de Aprendizado de Máquina para Classificação de Risco de Acidentes em Rodovias**. 2016. 106 f. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Computação, Coordenação de Pós-Graduação em Ciência da Computação, Ufmg, Campina Grande, 2016.
- BIERNATH, A. **Três morrem por hora no trânsito no Brasil**: veja quem corre mais risco. veja quem corre mais risco. 2022. Disponível em: <https://www.bbc.com/portuguese/brasil-61181312#:~:text=%22Essas%20mortes%20costumam%20seguir%20um,Observat%C3%B3rio%20Nacional%20de%20Seguran%C3%A7a%20Vi%C3%A1ria..> Acesso em: 07 mar. 2023.
- BARDUCHI, C. S.; **Segurança das Rodovias Federais Brasileiras: uma abordagem pela análise envoltória de dados**. 2018. 105 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia de Produção, Departamento de Engenharia de Produção, UNESP, Bauru, 2018.
- CHEIN, F.; **O modelo de regressão linear múltipla**. Em: **Introdução aos modelos de regressão linear**. 1. ed. Brasília, DF: Escola Nacional de Administração Pública, 2019. v. 1p. 33–58.

ESTADÃO. **Vendas de veículos elétricos e híbridos no mundo devem chegar a 8,3 milhões em 2022.** 2023. Disponível em: <https://mobilidade.estadao.com.br/inovacao/vendas-de-veiculos-eletricos-e-hibridos-no-mundo-devem-chegar-a-83-milhoes-em-2022/>. Acesso em: 03 mar. 2023.

FAVERO, L. P.; Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®. 1. ed. Rio de Janeiro: Elsevier Editora Ltda, 2015. v. 1.

FAVERO, L. P.; BELFIORE, P. Manual de Análise de Dados: Técnicas Multivariadas Exploratórias com SPSS e STATA. 1. ed. Rio de Janeiro: LTC, 2022. v. 1.

HAIR, J. F. et al. Análise multivariada de dados. 5. ed. Porto Alegre: Bookman editora, 2009.

IBM. **Métodos de seleção de variáveis de regressão logística:** uma abordagem pela análise envoltória de dados. Uma Abordagem pela Análise Envoltória de Dados. 2022. Disponível em: <https://www.ibm.com/docs/pt-br/spss-statistics/saas?topic=regression-logistic-variable-selection-methods>. Acesso em: 07 mar. 2023.

JÁCOME, I. BR-304 é a rodovia federal com maior número de mortes em acidentes no RN.

LOESCH, C.; HOELTGEBAUM, M. Métodos estatísticos multivariados. 1. ed. São Paulo: Editora Saraiva, 2012. v. 1.

LOUZADA, G.; NUNES, K. DA S. Metodologia Científica. Porto Alegre: SAGAH, 2019.

MEULMAN, J. J.; HEISER, W. J. SPSS Categories 10.0. 1. ed. Chicago: SPSS Inc, 1999. v. 1.

MEULMAN, J. J.; HEISER, W. J. SPSS Categories® 14.0. 1. ed. Chicago: SPSS Inc, 2005. v. 1.

MIGUEL, P. A. C.; Metodologia de pesquisa em engenharia de produção e gestão de operações. 3. ed. Rio de Janeiro: Elsevier Ltd, 2010.

OLIVEIRA, R.; **O mundo já tem mais de 1,44 bilhão de veículos.** 2023. Disponível em: <https://www.noticiasautomotivas.com.br/o-mundo-ja-tem-mais-de-1-bilhao-de-veiculos/#:~:text=J%C3%A1%20s%C3%A3o%20em%20torno%20de,outros%20ve%C3%ADculos%20de%20uso%20industrial..> Acesso em: 03 mar. 2023.

OMS. **Road traffic injuries.** 2022. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>. Acesso em: 13 mar. 2023.

ONU. **Acidentes de trânsito são a maior causa de morte de pessoas de 5 a 29 anos.** 2021. Disponível em: <https://news.un.org/pt/story/2021/11/1771092>. Acesso em: 07 mar. 2023.

PAIXÃO, A.; **Quantos carros tem no Brasil?**. 2022. Disponível em: <https://autoesporte.globo.com/curiosidades/noticia/2022/05/quantos-carros-tem-no-brasil.ghml>. Acesso em: 07 mar. 2023.

PRF. **Acidentes**. 2023. Disponível em: <https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-acidentes>. Acesso em: 07 mar. 2023.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, M. DEL P. B. Metodologia de Pesquisa. 5. ed. Porto Alegre: Penso, 2013.

SILVESTRE, F. L.; ACIDENTES DE TRÂNSITO EM RODOVIAS FEDERAIS BRASILEIRAS: DETERMINAÇÃO DE PONTOS CRÍTICOS NO ESTADO DO RIO GRANDE DO NORTE. TCC—Angicos: UFERSA, 2019.

MASIERO, M. S.; ANZANELLO, M.; GUSBERTI, T. D. H. Seleção de variável preditivas em preditivas em processos logísticos de distribuição. ENEGEP, out. 2012.

TURRIONI, J. B.; MELLO, C. H. P. Metodologia de pesquisa em engenharia de produção. 1. ed. Itajubá: UNIFEI, 2012.