

## O uso dos resíduos padronizados ajustados como medida auxiliar na Análise de Correspondência Simples aplicada a dados de acidentes em rodovias federais no Brasil

Damião Flávio dos Santos<sup>†</sup>

Universidade de Brasília - UNB

**Resumo:** *Os acidentes ocorridos em rodovias federais no Brasil são considerados uma importante causa de impactos sociais e econômicos no país. Esses acidentes todos os anos matam milhares de pessoas e/ou deixam em situação de vulnerabilidade as famílias das vítimas. Com o uso dos dados disponibilizados pela Polícia Rodoviária Federal (PRF), este trabalho objetiva identificar padrões de associação entre a classificação do acidente e o dia da semana de sua ocorrência, por meio da técnica bivariada análise de correspondência simples (ANACOR) com o auxílio dos resíduos padronizados ajustados. A aplicação do teste de hipótese qui-quadrado possibilitou aferir que a associação das duas variáveis não se dá de maneira aleatória, o que permitiu a análise dos resíduos padronizados ajustados de forma conjunta com ANACOR, sendo possível verificar que os acidentes com vítimas fatais e sem vítimas têm associação estatisticamente significativa ao nível de 5% de significância com os dias do final de semana (sábado e domingo).*

**Palavras-chave:** Análise de dados categóricos; Análise bivariada; ANACOR.

**Abstract:** *The accidents that occurred on federal highways in Brazil are considered an important cause of social and economic impacts in the country. These accidents every year kill thousands of people and/or leave the families of the victims in a situation of vulnerability. With the use of the data made available by the Federal Highway Police (PRF), this work aims to identify patterns of association between the classification of the accident and the day of the week of its occurrence, through the bivariate technique simple correspondence analysis (ANACOR) with the aid of adjusted standardized waste. The application of the chi-square hypothesis test made it possible to assess that the association of the two variables does not occur randomly, which allowed the analysis of the standardized residues adjusted jointly with ANACOR, being possible to verify that accidents with fatal victims and without victims have a statistically significant association at the level of 5% significance with the days of the weekend (Saturday and Sunday).*

**Keywords:** Categorical data analysis; Bivariate analysis; ANACOR.

## Introdução

Mundialmente os acidentes de trânsito representam uma das principais causas de morte, sendo a principal entre jovens na faixa etária de 15 a 29 anos, um fato que vai além do pesado fardo que as lesões e mortes ocorridas no trânsito representam para as economias nacionais e para as famílias (WHO, 2016).

No Brasil, os números elevados de acidentes ocorridos no trânsito em todo o território e em rodovias federais devem ser tratados como um problema relevante a ser inserido em políticas públicas do país.

---

<sup>†</sup>Autor correspondente: [d.flaviostate@gmail.com](mailto:d.flaviostate@gmail.com).

Por ano, milhares de pessoas têm suas vidas perdidas em eventos como esses, sejam eles originados por falha humana, falha na via, falha mecânica ou outros motivos. De acordo com o Departamento de Análise em Saúde e Vigilância das Doenças Não Transmissíveis (DASNT, 2021), em 2020, houve 31.088 mortes por acidentes de transporte, e, de acordo com os dados da Polícia Rodoviária Federal (PRF), no mesmo ano, registrou-se 5.291 óbitos devido a acidentes em rodovias brasileiras - ou seja, 17,0% dos óbitos totais.

Desde 2007, a Polícia Rodoviária Federal (PRF) adaptou uma política de dados abertos e fornece as informações dos acidentes em rodovias federais. Esses dados são bastante relevantes, pois a comunidade, de maneira geral poderá consumi-los de forma a identificar quais são as principais causas e/ou padrões desses acidentes, principalmente em relação àqueles que geram vítimas fatais, uma pauta de bastante interesse, tanto no que se refere à integridade humana quanto ao fator econômico do país, dado que, por ano, são milhares de vidas perdidas. Segundo estimativas da Confederação Nacional do Transporte (CNT, 2021), só em 2020, os acidentes ocorridos nas rodovias federais geraram um prejuízo em torno de R\$ 10,2 bilhões, sendo que cerca de R\$ 4,0 bilhões foram custos relacionados aos acidentes com vítimas fatais.

Ao considerar a importância dessas mortes, é necessário explorar alguns fatores que elevam a probabilidade de ocorrência desses eventos. De acordo com a série histórica dos dados da PRF, os acidentes de trânsito nas rodovias federais ocorrem com maior frequência aos dias de finais de semana, sendo um fator que ocasiona o aumento na probabilidade da ocorrência de mortes, conforme visto em Júnior et al. (2019).

Neste trabalho, será apresentado uma análise de correspondência simples, conhecida também por ANACOR, e a importância do auxílio dos resíduos padronizados ajustados advindos de uma tabela de dupla entrada entre a classificação do acidente e o dia da semana da ocorrência dos acidentes em rodovias federais no primeiro semestre de 2021. Com isso, será possível identificar as associações das categorias dessas variáveis. Ou seja, na prática permitirá identificar padrões de letalidade dos acidentes durante os dias da semana de ocorrência.

De acordo com Cunha (1997), a análise de correspondência estuda a associação entre variáveis categóricas de forma bivariada por meio da ANACOR ou multivariada (análise de correspondência múltipla), por meio de um conjunto de mapas (gráficos) perceptuais/intuitivos, nos quais nestes mapas verifica-se a existência de proximidade, chamada de similaridade ou dissimilaridade entre os objetos propostos, isto é, entre as categorias das variáveis em um específico estudo. No entanto, de acordo com Naito (2007), a interpretação gráfica dos resultados da análise de correspondência é um tema controverso, que ainda não reuniu consenso por parte da comunidade científica especializada e que, apesar da facilidade de representação gráfica da ANACOR, aconselha-se alguma precaução na sua interpretação.

Uma forma de avaliar as associações entre as categorias das variáveis é o uso dos resíduos padronizados ajustados, após rejeição da hipótese de nulidade do teste qui-quadrado. De acordo com Pereira (2001), a análise de resíduos padronizados ajustados complementa a avaliação do teste qui-quadrado, indicando onde há probabilidade de excesso ou falta de ocorrências. Para tanto, esses valores devem ser comparados com os pontos críticos de uma curva normal padronizada. Ao nível de confiança de 95%, a existência de valores positivos superiores a 1,96 indica mais ocorrências observadas que esperadas, e valores negativos inferiores a -1,96 sinalizam o contrário.

## Materiais e Métodos

### Teste de hipóteses $\chi^2$ (qui-quadrado)

O teste de hipóteses  $\chi^2$  de Pearson é bastante utilizado na literatura da área da saúde e em estudos sociais. De acordo com Bussab e Morettin (2002), um dos principais objetivos do teste é verificar se existe associação entre duas variáveis categóricas, a partir de uma tabela de contingência. O teste compara as frequências observadas ( $O_{ij}$ ) na amostra com as frequências esperadas ( $E_{ij}$ ) sob a hipótese de independência ( $H_0$ ).

Suponha que temos duas variáveis categóricas X e Y, classificadas em  $j$  categorias para X e  $j$  categorias para Y. Então uma tabela de dupla entrada ou tabela de contingência será composta pelos seguintes elementos:

- $n_{ij}$  = número de elementos pertencentes à  $i$ -ésima categoria de X e  $j$ -ésima categoria de Y;
- $n_{i.}$  = número de elementos da  $i$ -ésima categoria de X (Total marginal de X);
- $n_{.j}$  = número de elementos da  $j$ -ésima categoria de Y (Total marginal de Y);
- $n$  = número total de elementos.

Sob a hipótese de que as variáveis X e Y não sejam associadas ou independentes, temos que os valores esperados são dados por:

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{n}, \quad (1)$$

ou seja, o total marginal da coluna (em que está a célula), multiplicado pelo total marginal da linha (em que está a célula), dividido pelo tamanho da amostra.

Dessa forma, o teste  $\chi^2$  é dado por:

$$\chi_{calc}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

Os valores de  $\chi_{calc}^2$  seguem, aproximadamente, uma distribuição  $\chi^2$  com  $\nu = (I - 1) \cdot (J - 1)$  graus de liberdade. A conclusão do teste será dada pelo “valor-p”, que é calculado e confrontado com o nível de significância, o que significa que caso o “valor-p” seja inferior a  $\alpha = 0,05$ , rejeitamos a hipótese  $H_0$ , ou seja, conclui-se que há relação ou associação entre as variáveis estudadas, ou que a associação não se dá de forma aleatória, ao nível de significância de 5%. As hipóteses, portanto, são definidas da seguinte forma:

$$\begin{cases} H_0 : \text{As duas variáveis categóricas se associam de forma aleatória,} \\ H_1 : \text{A associação entre as duas variáveis categóricas não se dá de forma aleatória.} \end{cases}$$

Para a aplicação do teste, utilizou-se o *software* livre R (R CORE TEAM, 2021), por meio da função por meio da função `chisq.test()` do pacote `stats` (R CORE TEAM, 2021).

De acordo com Barnett e Lewis (1994), pode-se definir os resíduos padronizados em uma tabela de contingência dividindo-se em cada célula o valor do resíduo calculado pela raiz quadrada da respectiva frequência absoluta esperada. Dessa forma, para  $ij$  ( $i = 1, \dots, I$  e  $j = 1, \dots, J$ ), tem-se que:

$$e_{pad_{ij}} = \frac{(O_{ij} - E_{ij})}{\sqrt{E_{ij}}} \quad (3)$$

Ao considerar a diferença entre o valor observado e o valor esperado como resíduo bruto, Sharpe (2015) destaca que as células com o maior valor esperado, também produzem os maiores resíduos brutos e para superar essa redundância, o resíduo padronizado é calculado e que a partir da análise residual é possível identificar as células específicas que fazem a maior contribuição para o resultado do teste qui-quadrado.

A partir da equação 3 calcula-se os resíduos padronizados ajustados proposto por Haberman (1973), em que para cada célula para  $ij (i = 1, \dots, I \text{ e } j = 1, \dots, J)$  são calculados da seguinte forma:

$$e_{pad \text{ ajustado}_{ij}} = \frac{e_{pad_{ij}}}{\sqrt{\left(1 - \frac{n_{i.}}{n}\right) \times \left(1 - \frac{n_{.j}}{n}\right)}} \quad (4)$$

Com a equação 4, podemos definir uma matriz de resíduos padronizados ajustados. De acordo com Fávero e Belfiore (2017), caso determinada célula apresente um resíduo padronizado ajustado com valor superior a 1,96, poderemos caracterizar ao nível de significância de 5% a associação entre as duas categorias. Dessa forma, tão importante quanto avaliar a existência de associação estatisticamente significativa entre duas variáveis categóricas é estudar a relação de dependência entre cada par de categorias, o que, inclusive, facilita na leitura do mapa perceptual de uma análise de correspondência simples.

### Análise de Correspondência Simples (ANACOR)

De acordo com Greenacre (2006), a história da análise de correspondência pode ser rastreada até Hirsfeld (1935), que deu uma formulação algébrica da correlação entre linhas e colunas de uma tabela de contingência. No entanto, a análise de correspondência se popularizou na década de 1960 e 1970 por meio do estatístico francês Jean-Paul Benzécri, primeiramente na França e se estendendo pela Europa. Suas origens, porém, estão em trabalhos do início do século XX feitos por Pearson e Fisher, dois dos mais importantes nomes da estatística.

De acordo com Fávero e Belfiore (2017), a análise de correspondência simples, também conhecida como ANACOR, é uma técnica de análise bivariada por meio da qual é estudada a associação entre duas variáveis categóricas e entre suas categorias. Assim como o teste qui-quadrado, a aplicação da análise de correspondência parte de uma tabela de contingência, na qual, nas linhas, são apresentadas as categorias de uma variável; nas colunas, as categorias de outra variável; e, nas células as frequências observadas de objetos ou indivíduos.

De acordo com Mingoti (2005), a matriz de correspondência  $\mathbf{P}_{p \times q}$  é constituída das proporções  $p_{ij} = \frac{n_{ij}}{n}$ , ou seja, o número de elementos que pertencem à categoria  $i$  da variável X e à categoria  $j$  da variável Y, dividido pelo total de indivíduos e, sendo assim, a soma dos elementos de  $\mathbf{P}_{p \times q}$  passa a ser, portanto, igual a 1. Assim, temos a seguinte matriz:

$$\mathbf{P}_{p \times q} = \begin{bmatrix} \frac{n_{11}}{n} & \frac{n_{12}}{n} & \dots & \frac{n_{1J}}{n} \\ \frac{n_{21}}{n} & \frac{n_{22}}{n} & \dots & \frac{n_{2J}}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{J1}}{n} & \frac{n_{J2}}{n} & \dots & \frac{n_{IJ}}{n} \end{bmatrix}$$

A partir da matriz  $\mathbf{P}$ , calcula-se a matriz de perfil das linhas e das colunas, em que necessita-se obter os vetores de frequências relativas marginais  $\mathbf{r}$  e  $\mathbf{c}$ , vetores linhas e colunas respectivamente, também chamada de massa, para cada célula da tabela de contingência, podendo ser construído a partir dos elementos dos vetores  $\mathbf{r}$  e  $\mathbf{c}$  das seguintes formas:

$$r_i = \sum_{j=1}^J p_{ij} = \sum_{j=1}^J \frac{n_{ij}}{n}, \quad c_j = \sum_{i=1}^I p_{ij} = \sum_{i=1}^I \frac{n_{ij}}{n}.$$

De acordo com Fávero e Belfiore (2017), pode-se definir o conceito de massa como uma medida de influência ou preponderância de determinada categoria em relação às demais, com base em sua frequência observada, ou seja, representa o quanto do total de cada categoria está presente no estudo (vertical e horizontal).

Com a criação dos vetores de perfil de linhas e colunas, segue para a construção das matrizes de perfil de linhas e de colunas, onde as mesmas são compostas da matriz diagonal dos elementos de  $\mathbf{r}$  e de  $\mathbf{c}$ , respectivamente e assim, temos as matrizes  $\mathbf{D}_r$  e  $\mathbf{D}_c$ .

Agora, precisa-se fazer a decomposição em coordenadas principais, conforme visto em Minototi (2005). Para isso, considere a matriz  $\tilde{\mathbf{P}}$ , dada por:

$$\tilde{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}'.$$

De acordo com Infantosi (2014), pré-multiplicando e pós-multiplicando a matriz  $\tilde{\mathbf{P}}$  por  $\mathbf{D}_c^{-1/2}$  e  $\mathbf{D}_r^{-1/2}$  obtêm-se um nova matriz conforme equação (5), que é a padronização da matriz de perfis. A ideia por trás da construção desta matriz está na necessidade da disposição gráfica de vetores multivariados em utilizar a origem das próprias variáveis, ou determinar uma nova origem por meio do método de mínimos quadrados, no qual a soma dos quadrados das distâncias dos pontos até os eixos é minimizada. O método computacional mais utilizado para tal minimização é o algoritmo de Decomposição por Valores Singulares (DVS), em que a matriz de perfis é fatorada em três matrizes.

$$\mathbf{S} = \mathbf{D}_c^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_r^{-1/2}, \quad (5)$$

que também pode ser escrito da seguinte maneira:

$$s_{ij} = (p_{ij} - r_i c_j) / \sqrt{r_i c_j}.$$

A partir da matriz  $\mathbf{S}$ , será feita a DVS, na qual, de acordo com Johnson e Wichern (2007), pode ser escrito como

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}',$$

em que  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_k\}$ , com  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$ ,  $\mathbf{U}$  é uma matriz ortogonal de ordem  $n \times k$ , isto é,  $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ . O conjunto de valores  $\{\lambda_i\}$  são chamados de valores singulares de  $\mathbf{S}$ . Se  $\mathbf{U}$  e  $\mathbf{V}$  são escritos em termos de seus vetores coluna,  $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k]$  e  $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_k]$ , então  $\{\mathbf{u}_i\}$  são os vetores singulares à esquerda de  $\mathbf{S}$  e  $\{\mathbf{v}_i\}$  são os vetores singulares à direita de  $\mathbf{S}$ . A matriz  $\mathbf{S}$  pode então ser escrita como

$$\mathbf{S} = \sum_{i=1}^k \lambda_i \mathbf{u}_i \mathbf{v}_i'.$$

Com as matrizes  $\mathbf{U}$ ,  $\mathbf{V}$  e  $\mathbf{\Lambda}$ , pode-se construir as coordenadas principais das linhas e colunas. As coordenadas principais das linhas são definidas como:

$$\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U}\mathbf{\Lambda}, \quad \mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{S}\mathbf{V}.$$

As coordenadas principais das colunas são:

$$\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V}\mathbf{\Lambda}, \quad \mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{S}'\mathbf{U}.$$

Por meio das matrizes  $\mathbf{F}$  e  $\mathbf{G}$ , constroem-se os gráficos de dispersão das coordenadas principais das linhas e colunas, também chamados de mapas perceptuais ou gráficos de correspondência, para que se possa analisar a associação entre as categorias de suas variáveis. A partir desse gráfico, é possível identificar padrões de associação entre as categorias das variáveis, ao observar a proximidade entre as coordenadas das categorias.

Para a aplicação da análise, utilizou-se o *software* livre R (R CORE TEAM, 2021), por meio da função `ca()` do pacote `FactoMineR` (HUSSON et al., 2020).

## Banco de Dados

Os dados utilizados nesse estudo são referentes aos acidentes rodoviários no Brasil no primeiro semestre de 2021, que estão disponíveis no banco de dados oficial da PRF, na seção de dados abertos. No total, foram registradas 36.703 ocorrências de acidentes e 2.638 pessoas mortas, entre as datas de 01/01/2021 e 31/07/2021.

Para este estudo, considerou-se as variáveis **classificação do acidente** (sem vítimas, com vítimas feridas e com vítimas fatais) e **dia da semana** (domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira e sábado).

## Resultados e Discussão

A Tabela 1 fornece uma análise preliminar da distribuição percentual por linha, ou seja, a distribuição percentual da classificação do acidente para cada dia da semana. De forma apenas descritiva, observamos que a distribuição percentual do sábado e domingo são muito próximas, mas diferentes com relação aos outros dias da semana. Isso pode ser indícios de diferença entre as distribuições percentuais durante a semana e o final de semana.

Tabela 1: Tabela de contingência entre o dia da semana e a classificação do acidente, com frequência absoluta observada e distribuição percentual por linha

Dia da Semana	Classificação do Acidente			Total
	Sem Vítimas N (%)	Com Vítimas Feridas N (%)	Com Vítimas Fatais N (%)	
Domingo	1.231 (20,6%)	4.260 (71,1%)	499 (8,3%)	5.990
Segunda-feira	800 (16,4%)	3.728 (76,6%)	336 (6,9%)	4.864
Terça-feira	716 (16,1%)	3.440 (77,3%)	292 (6,6%)	4.448
Quarta-feira	781 (17,1%)	3.520 (76,9%)	274 (6,0%)	4.575
Quinta-feira	807 (17,3%)	3.539 (75,9%)	317 (6,8%)	4.663
Sexta-feira	1.028 (17,5%)	4.464 (75,9%)	393 (6,7%)	5.885
Sábado	1.255 (20,0%)	4.496 (71,6%)	527 (8,4%)	6.278
Total	6.618 (18,0%)	27.447 (74,8%)	2.638 (7,2%)	36.703

No entanto, é necessária a realização do teste de hipóteses qui-quadrado para verificar se as categorias das duas variáveis associam-se de forma aleatória ou se existe uma relação de dependência entre as mesmas. Ou seja, testar as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As variáveis Dia da Semana e Classificação do Acidente se associam de forma aleatória,} \\ H_1 : \text{As variáveis Dia da Semana e Classificação do Acidente não se associam de forma aleatória.} \end{cases}$$

Tabela 2: Teste qui-quadrado de Pearson entre as variáveis Dia da Semana e Classificação do Acidente

Qui-quadrado	g.l.	Valor-p
123,81	12	$1,070724 \times 10^{-20}$

Conforme podemos verificar por meio da Tabela 2, o valor-p da estatística  $\chi_{calc}^2$  é consideravelmente menor do que 0,05, ou seja, rejeitamos a hipótese  $H_0$  e verificamos que a associação das variáveis Dia da Semana e Classificação do Acidente não se dá de forma aleatória. No entanto, de acordo com Fávero e Belfiore (2017), embora o resultado do teste qui-quadrado tenha mostrado a existência de um padrão de dependência entre o dia da semana do acidente e a classificação do acidente, é a análise dos resíduos padronizados ajustados que revelará os padrões

característicos de cada categoria do dia da semana segundo o excesso ou a falta de ocorrências de sua combinação com cada categoria da classificação do acidente.

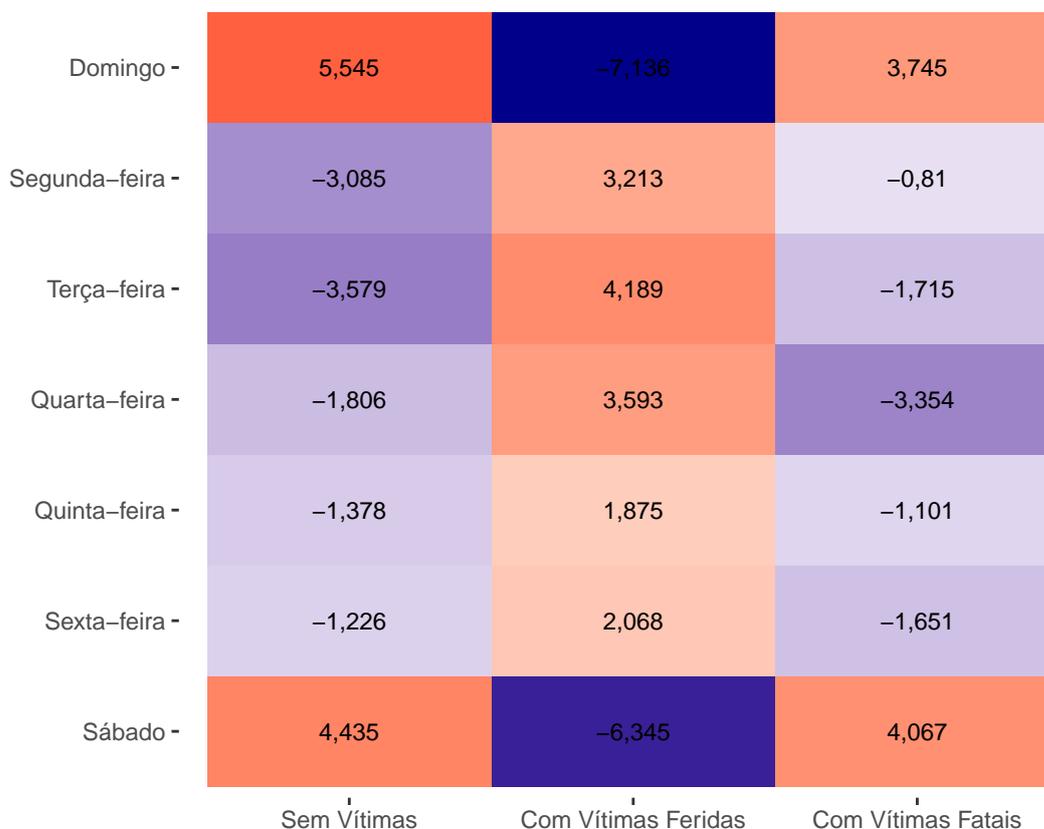


Figura 1: Mapa de calor dos resíduos padronizados ajustados da associação entre o dia da semana do acidente e sua classificação

Conforme descrito na subseção , os resíduos padronizados ajustados com valores positivos superiores a 1,96, apresentados na Figura 1, correspondem ao excesso de ocorrências em cada célula, ao nível de significância de 5%. Dessa forma, pode-se afirmar, ao nível de 5% de significância, que os acidentes **sem vítimas** e **com vítimas fatais** se associam com os dias do final de semana (sábado e domingo), já os acidentes **com vítimas feridas** se associam com os dias da semana (exceto a quinta-feira com 5% de significância).

É importante destacar que o sábado e domingo, se comparados aos outros dias, são os que mais registram acidentes nas rodovias federais no Brasil, principalmente acidentes sem vítimas. Destaca-se também que a partir de 2015 os registros de ocorrência desses acidentes passaram a ser realizados diretamente pelos usuários, por meio da declaração eletrônica de acidente de trânsito (e-DAT) na internet, substituindo o boletim que vinha sendo elaborado pessoalmente pelos policiais rodoviários federais. Esse fato pode estar levando a subnotificação dos registros de acidente sem vítimas, principalmente durante a semana.

Com base no mapa perceptual da Figura 2, podemos verificar que há uma segmentação clara entre os eventos ocorridos durante a semana (de segunda-feira a sexta-feira) e final de semana (sábado e domingo). No entanto, destacamos alguns equívocos de muitos pesquisadores ao analisar esse tipo de gráfico isoladamente.

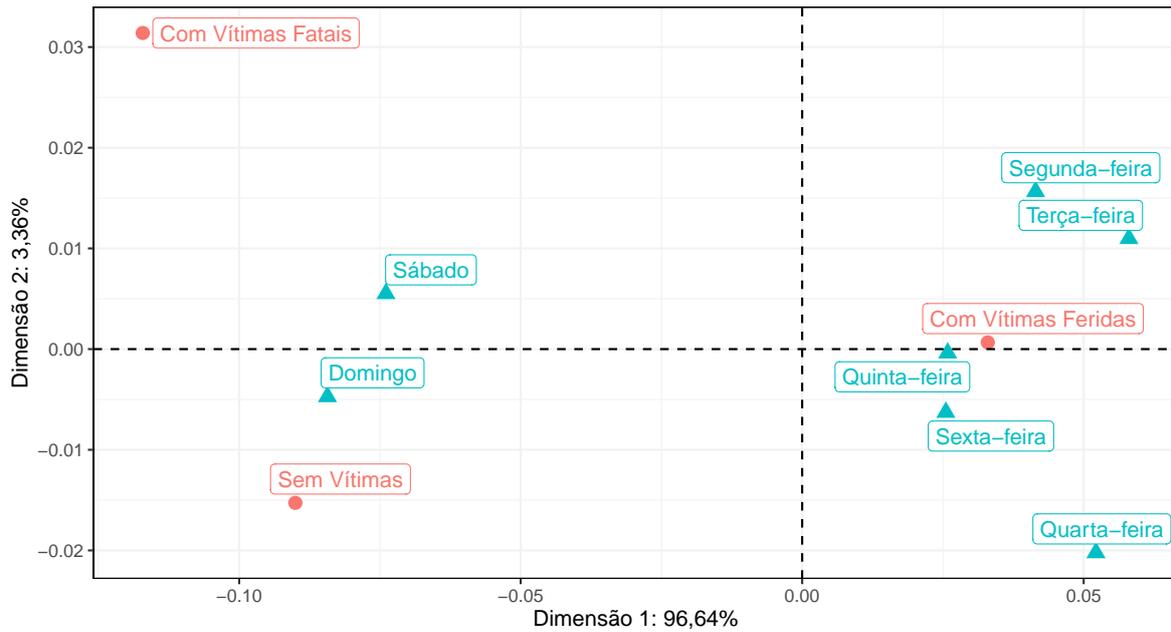


Figura 2: Mapa perceptual para a classificação do acidente segundo o dia da semana da ocorrência do acidente

Há uma má interpretação do mapa perceptual com relação às proximidades entre os pontos, pois nem sempre os mais próximos significam necessariamente associação maior entre as categorias - por exemplo, a ocorrência de acidentes com vítimas feridas está mais próxima da ocorrência da quinta-feira, no entanto, vimos, pela análise de resíduos padronizados, que a quinta-feira não está associada ao nível de 5% de significância a essa classificação de acidente.

Além disso, ao se observar o mapa perceptual isoladamente, poderia concluir-se visualmente que os acidentes sem vítimas se associam fortemente com o domingo e acidentes com vítimas fatais, com o sábado, mas não só isso deve ser considerado. Conforme apresentado na análise de resíduos, as duas classificações de acidentes se associam com os dois dias do final de semana e não de forma excludente. Esses resultados são condizentes com as descobertas realizadas por Júnior et al. (2019) e Miranda et al. (2021).

Conforme visto em Júnior et al. (2019), por meio de um modelo de regressão logística binária, com dados dos acidentes em rodovias federais no ano de 2016, as chances de um acidente de trânsito ser letal são, em média, 35% maiores aos domingos e 18% aos sábados, em comparação à segunda-feira. Porém não verificou-se diferença significativa entre a letalidade dos acidentes que ocorreram nos demais dias da semana quando comparados à segunda-feira. De acordo com Miranda et al. (2021), pela modelagem via regressão logística multinomial, identificou-se que os fins de semana elevaram as chances de falecimento de pedestres em 22,4% em relação aos dias úteis.

Além desses resultados, ressalta-se que há outros fatores não explorados neste trabalho que elevam a probabilidade de ocorrência de acidentes com vítimas fatais, como região do acidente, tipo de pista, tipo de veículo, fase do dia, tipo de acidente, clima e outros, conforme descrito em Miranda et al. (2021), Júnior et al. (2019) e Roquim et al. (2019).

## Considerações Finais

Observou-se, por meio do teste qui-quadrado, que as variáveis categóricas Classificação do Acidente e Dia da Semana que ocorreu o acidente não se dá de forma aleatória e, a partir daí, foi possível utilizar a análise de correspondência simples, com o intuito de investigar as associações entre as categorias das variáveis.

Inicialmente, ao utilizar os resíduos padronizados ajustados, verificou-se que a ocorrência dos acidentes com vítimas fatais e sem vítimas tem uma associação estatisticamente significativa, ao nível de 5% de significância com os dias do final de semana (sábado e domingo), e os acidentes com vítimas feridas têm associação com os dias da semana. No entanto, apesar de muito próximo, como o resíduo padronizado ajustado (1,875) da associação entre a quinta-feira e acidentes com vítimas feridas foi menor do que 1,96, a quinta-feira é o único dia da semana que não tem associação estatisticamente significativa, ao nível de 5%.

Ademais, essa conclusão não seria possível apenas observando o mapa perceptual da análise de correspondência simples, pois, como apresentado, a coordenada da quinta-feira no mapa perceptual ficou muito próximo da categoria “acidentes com vítimas feridas”, o que levaria a uma análise equivocada apenas observando essa proximidade. Apesar disso, não faz da ANACOR uma técnica que pode ser substituída apenas pelos resíduos padronizados ajustados, pois técnicas distintas que devem ser utilizadas de forma conjunta. Com o mapa perceptual da análise de correspondência, foi possível identificar além da associação entre as categorias das variáveis, uma segmentação entre a classificação dos acidentes ocorridos aos finais de semana e durante a semana, corroborando com a análise de resíduos padronizados ajustados.

Por fim, de acordo com a literatura apresentada, há outros fatores identificados por meio de modelos de regressão que elevam a probabilidade de óbitos em acidentes nas rodovias federais. Ressalta-se que, para a estimação desses modelos, são realizados alguns testes estatísticos para a identificação de variáveis significativas para o modelo e, dessa forma, a análise de correspondência simples, a análise de resíduos padronizados ou até mesmo uma análise de correspondência múltipla, também poderá ser utilizada para fins de exploração dos dados em trabalhos futuros, de modo a corroborar com análises preditivas dos modelos de regressão.

## Referências

BARNETT, V.; LEWIS, T. *Outliers in statistical data*. 3. ed. Chichester: John Wiley & Sons, 1994.

BUSSAB, W. O.; MORETTIN, P. A., *Estatística básica*. 5ª edição. São Paulo: Saraiva, 2002.

CNT. Confederação Nacional do Transporte. *Painel CNT de Consultas Dinâmicas dos Acidentes Rodoviários - 2020*. Disponível em: <https://www.cnt.org.br/painel-acidente>. Acesso em: 06 set. 2021. 6 p.

CUNHA, J. *Análise Multidimensional de Dados Categóricos: A aplicação das Análises de Correspondência Simples e Múltipla em Marketing e sua Integração com Técnicas de Análise de Dados Quantitativos*. Rio Grande do Sul: UFRGS, 1997.

DASNT, Departamento de Análise em Saúde e Vigilância das Doenças Não Transmissíveis. Disponível em: <https://shortest.link/dantpspaineis-de-monitoramentomortalidadecid10>. Acesso em: 18 set. 2021.

FÁVERO, L. P.; BELFIORE, P. *Manual de Análise de dados: estatística e modelagem*

*multivariada com Excel®*, *SPSS®* e *Stata®*. Rio de Janeiro: Elsevier, 2017.

GREENACRE, M.; BLASIUS, J. *Multiple Correspondence Analysis and Related Methods*. New York: Chapman & Hall/CRC, 2006.

HIRSFELD, H. O. *A connection between correlation and contingency*. Proceedings of the Cambridge Philosophical Society, p. 520, 524, 1935.

HUSSON, F.; JOSSE, J.; LE, S.; MAZET, J. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. R package version 2.4. 2020. Disponível em:  
<https://CRAN.R-project.org/package=FactoMineR>

HABERMAN, S. J. *The analysis of residuals in cross-classified tables*. Biometrics, v. 29, n. 1, p. 205 - 220, 1973.

INFANTOSI, A. F. C. ; GAMA DIAS COSTA, F. C.; ALMEIDA, R. M. V. R. *Análise de correspondência: bases teóricas na interpretação de dados categóricos em ciências da saúde*. Cadernos de Saúde Pública, n. 30(3), p.475 - 486, 2014.

JOHNSON, R. A. ; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 2007.

JUNIOR, G. T. B; BERTHO, A. C. S.; VEIGA, A. de C. *A letalidade dos acidentes de trânsito nas rodovias federais brasileiras*. Revista Brasileira de Estudos de População, 36, p. 1 - 22, 2019.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. Belo Horizonte, UFMG, 2005.

MIRANDA, R.; SILVA, W.P.; DUTT-ROSS, S. *Identificação de fatores determinantes da severidade das lesões sofridas por pedestres nas rodovias federais brasileiras entre 2017 e 2019: Análise via regressão logística multinomial*. Scientia Plena, v.17, n. 4, 2021. Disponível em:  
<https://www.scientiaplenu.org.br/sp/article/view/5897/2382>

NAITO, S. D. N. P. *Análise de Correspondências Generalizada*. Dissertação (Mestrado em Bioestatística). Universidade de Lisboa, 2007.

PEREIRA, J. C. R. *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais*. 3. ed. São Paulo: Editora da Universidade de São Paulo, 2001.

PRF, Polícia Rodoviária Federal. Disponível em:  
<https://arquivos.prf.gov.br/arquivos/index.php/s/n1T3lymvIdD0zzb>. Acesso em: 06 set. 2021.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL <https://www.R-project.org/>.

R CORE TEAM and contributors worldwide. *stats: The R Stats Package*. R package version 4.2.0. 2021. Disponível em:  
<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>

ROQUIM, F. V.; NAKAMURA, L. R.; RAMIRES, T. G.; LIMA, R. R. *Regressão Logística: o que leva um acidente rodoviário a ser uma tragédia?*. Sigmae, Alfenas, v.8, n.2, p. 19-28. 2019. Disponível em:

<https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/902/605>

SHARPE, D. (2015). *Your Chi-Square Test is Statistically Significant: Now What?*. Practical Assessment, Research & Evaluation, 20(8). Disponível em:

<http://pareonline.net/getvn.asp?v=20&n=8>

WORLD HEALTH ORGANIZATION (WHO). *Global status report on road safety 2015*. Geneva: WHO, 2015. Disponível em:

<https://shortest.link/whointviolenceinjuryprevention>.