

Estudo dos Acidentes em Rodovias por Meio da Análise Multivariada

Fábio André Negri Balbo
PPGMNE - UFPR*
andrebalbo@gmail.com

Liliana Madalena Gramani
Departamento de Matemática, UFPR*
l.gramani@gmail.br

Anselmo Chaves Neto
Departamento de Estatística, UFPR*
anselmo@ufpr.br

Tatiane Cazarin da Silva
PPGMNE - UFPR*
tatianecazarin@gmail.com

Nivaldo A. Minervi
PPGEP - UFPR*
nivaldominervi@yahoo.com.br

Sandro Rodrigues
PPGMNE - UFPR*
sandrordrigues23@gmail.com

*Centro Politécnico, Jardim das Américas, CEP 81531-990 - Curitiba - Paraná.

RESUMO

Este trabalho faz um estudo dos acidentes ocorridos na BR-277 no período de janeiro de 2007 a novembro de 2009, buscando relacionar os diferentes tipos de acidentes com as condições de pista, tipos de acostamento, condições meteorológicas, sentido da rodovia, entre outras. O método utilizado para determinar essas relações foi a Análise Fatorial. Os dados utilizados nesse trabalho são compostos por 35 planilhas, cedidas pela Polícia Rodoviária Federal do estado do Paraná, sendo que cada planilha contém os dados (qualitativos e quantitativos) dos 10.096 acidentes ocorridos entre janeiro de 2007 e novembro de 2009. Como resultados pôde-se verificar, em todas as análises, que o quilômetro 584 (Cascavel - PR) é o mais problemático com relação ao número de acidentes, e que a falta de atenção do motorista pode ser considerada a principal responsável por esses acidentes.

PALAVRAS-CHAVE. Acidentes em Rodovias. Análise Fatorial. Análise Multivariada. EST - Estatística.

ABSTRACT

In the work reported here is presented a study of accidents in the BR-277 from January 2007 to November 2009, seeking to relate the different types of accidents with the track conditions, sidewalk types, weather, direction of the highway, among others. The methodology used to determine these relations was the factorial analysis. The data used in this study consists of 35 sheets, provided by the Federal Highway Police in the state of Paraná, and each worksheet contains the data (quality) of the 10,096 accidents that occurred between January 2007 and November 2009. The results could be verified in all tests, the 584 km (Cascavel - PR) is more problematic with regard to the number of accidents, and the inattention drivers can be considered the main responsible for these injuries.

KEY-WORDS. Accidents on Highways. Factorial Analysis. Multivariate Analysis. EST - Statistics.

1 Introdução

A BR-277 é uma rodovia federal da malha rodoviária do Brasil localizada no estado do Paraná. Foi inaugurada em março de 1969, e tem aproximadamente 730 quilômetros de extensão, com início no Porto de Paranaguá e término na Ponte da Amizade, em Foz do Iguaçu. É considerada a principal via rodoviária do Paraná e uma das mais importantes do Brasil, e é conhecida como o “corredor do Mercosul”. Desempenha ainda, um importante papel no escoamento de produtos agropecuários produzidos no estado do Paraná rumo à exportação.

Como a BR-277 se localiza em uma região estratégica, cada vez mais o tráfego de veículos tem aumentado e conseqüentemente o número de acidentes registrados também têm crescido¹. Com o objetivo de encontrar os quilômetros desta rodovia com maior número de acidentes e identificar suas possíveis causas, será usado uma técnica estatística multivariada (em um banco de dados referente aos 10.096 acidentes na BR-277 no período de janeiro de 2007 à novembro de 2009), conhecida como Análise Fatorial (Johnson e Wichern, 2002). Foram realizadas 3 análises em períodos diferentes, afim de verificar se os fatores obtidos e os piores pontos da rodovia com relação ao número de acidentes permanecem os mesmos.

Na literatura podem ser encontrados trabalhos que usam a Análise Fatorial para estudar problemas de trânsito, como exemplo tem-se Bartilotti (2009) que analisa a percepção do motorista no trânsito, Pires e Maia (2009) que fazem um estudo sobre as vítimas de acidentes de trânsito, Silva (2004) que estuda o comportamento dos motoristas, entre outros.

A utilização da Análise Fatorial neste trabalho tem por objetivo agrupar variáveis correlacionadas em fatores, resumindo assim a grande quantidade de variáveis envolvidas nos acidentes, para depois relacionar os quilômetros da rodovia com as principais causas e tipos de acidentes, com o uso dos escores fatoriais. A Análise Fatorial foi a técnica multivariada que mais adequada, uma vez que a análise de agrupamento (Mingoti, 2007) é de difícil interpretação devido ao grande número de variáveis. Já a análise de correlação canônica (Johnson e Wichern, 2002) necessita de grupos de variáveis pré-definidos, o que supõe o conhecimento das variáveis que são correlacionadas.

A estrutura deste trabalho está dividido em mais quatro seções envolvendo a descrição matemática da análise fatorial, a apresentação dos materiais e métodos usados no problema, a discussão dos resultados obtidos e a conclusão. Em detalhes:

Na Seção 2 será feito o desenvolvimento matemático da Análise Fatorial, onde explica-se os testes que verificam se a análise fatorial é adequada ao conjunto de dados, o modelo fatorial ortogonal, o modelo fatorial pelo método das componentes principais, critério de rotação varimax e por fim a obtenção dos escores fatoriais pelos mínimos quadrados.

Na Seção 3 será descrito todo o processo de análise de dados, tais como, a composição da amostra utilizada, os tipos de variáveis envolvidas, a maneira como as técnicas e testes multivariados foram aplicados.

Na Seção 4 será discutido os resultados obtidos com a análise fatorial e a rotação dos fatores pelo critério varimax, tais como os fatores observados, comunalidades e escores fatoriais que irão ajudar na interpretação e conclusão dos resultados na Seção 5.

¹De acordo com o banco de dados da Polícia Rodoviária Federal no Paraná em 2007 foram registrados 2.887 acidentes, em 2008, 3.470 acidentes e de janeiro à novembro de 2009, 3.739 acidentes na BR-277.

2 Desenvolvimento do Tema

Ao aplicar a análise fatorial sobre um conjunto de dados, surge a discussão se os dados são apropriados, para a utilização dessa técnica. Dois testes são apresentados brevemente no início dessa seção: o teste de esfericidade de Bartlett e o MSA - medida de adequabilidade da amostra de Kaiser-Meyer-Olkin que analisam se a estrutura de correlação da matriz de dados é adequada a análise fatorial. Nesta seção também será apresentado a descrição matemática da análise fatorial usando o modelo fatorial ortogonal e estendido para o método das componentes principais e ainda será abordado o critério de rotação varimax e a obtenção dos escores fatoriais utilizando-se os mínimos quadrados.

2.1 Análise Fatorial

A análise fatorial foi introduzida por Spearman (1904)². Inicialmente usada nas áreas de psicologia e ciências sociais, na tentativa de identificar os fatores relacionados com a inteligência humana e ligá-los, de algum modo à etnia. Devido a subjetividade e falta de unicidade de suas soluções, tem sido alvo de críticas ao longo dos anos (Hills, 1977). Porém, é uma das técnicas multivariadas mais conhecidas e tem sido muito utilizada em áreas como química (Ozerenko, 2007), educação (Marques, 2010), geologia (Queiroz, 2007), entre outras.

Segundo Hair (2002), a análise fatorial é um conjunto de métodos estatísticos multivariados cujo propósito principal é definir a estrutura subjacente em uma matriz de dados. Em termos gerais, a análise fatorial aborda o problema de analisar a estrutura das interrelações (correlações) entre um grande número de variáveis, definindo um conjunto de dimensões latentes comuns, chamadas de fatores. O objetivo principal da análise fatorial é resumir os dados, obtendo dimensões latentes, que quando interpretadas e compreendidas, descrevem os dados em um número muito menor de conceitos do que as variáveis originais individuais.

A seguir será desenvolvido o embasamento matemático da análise fatorial.

2.1.1 Testes para a Análise Fatorial

Para aplicar a análise fatorial a um conjunto de dados, uma das primeiras preocupações é saber se o conjunto de dados é adequado para aplicar o método multivariado. Existem dois testes capazes de determinar se o conjunto de dados é apropriado: teste de esfericidade de Bartlett e a medida de adequabilidade da amostra de Kaiser-Meyer-Olkin (KMO). Ambos os testes analisam a estrutura de correlação da matriz de dados.

2.1.2 Modelo Fatorial Ortogonal

De acordo com Johnson e Wichern (2002), o vetor aleatório \mathbf{X} , com p variáveis observáveis, tem vetor de média $\boldsymbol{\mu}$, e matriz de covariância $\boldsymbol{\Sigma}$. O modelo fatorial \mathbf{X} é linearmente dependente sobre as poucas variáveis aleatórias não observáveis F_1, F_2, \dots, F_m com ($m \leq p$) chamados fatores comuns e p fontes de variações ativas $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas de erros ou fatores

²Charles Edward Spearman (10 de setembro de 1863 - 7 de setembro de 1945) foi um psicólogo inglês conhecido pelo seu trabalho na área da estatística, como um pioneiro da análise fatorial e pelo coeficiente de correlação de postos de Spearman. Ele também fez bons trabalhos de modelos da inteligência humana, incluindo a descoberta de que escores em testes cognitivos incompatíveis exibiam um fator geral único, batizado de fator "g".

específicos. O modelo fatorial é

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (1)$$

ou em notação matricial

$$\underbrace{\mathbf{X} - \boldsymbol{\mu}}_{(p \times 1)} = \underbrace{\mathbf{L}}_{(p \times m)} \cdot \underbrace{\mathbf{F}}_{(m \times 1)} + \underbrace{\boldsymbol{\varepsilon}}_{(p \times 1)}, \quad (2)$$

onde ℓ_{ij} é denominado como a carga da i -ésima variável no j -ésimo fator, portanto a matriz \mathbf{L} é chamada de matriz das cargas fatoriais. O i -ésimo fator específico ε_i é associado somente com a variável resposta X_i . Os p desvios $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são representados pelas $p + m$ variáveis aleatórias $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, as quais não são observáveis.

Com tantas quantidades não observáveis e para tornar útil o modelo fatorial, algumas pressuposições sobre \mathbf{F} e $\boldsymbol{\varepsilon}$ devem ser feitas, tais como

$$\begin{aligned} E(\mathbf{F}) &= \underbrace{\mathbf{0}}_{m \times 1}, & \text{Cov}(\mathbf{F}) &= E(\mathbf{F}\mathbf{F}') = \underbrace{\mathbf{I}}_{(m \times m)} \\ E(\boldsymbol{\varepsilon}) &= \underbrace{\mathbf{0}}_{(p \times 1)}, & \text{Cov}(\boldsymbol{\varepsilon}) &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \underbrace{\boldsymbol{\Psi}}_{(p \times p)} \end{aligned} \quad (3)$$

Onde \mathbf{I} é a matriz identidade e $\boldsymbol{\Psi}$ é uma matriz diagonal contendo as variâncias específicas. E ainda \mathbf{F} e $\boldsymbol{\varepsilon}$ são independentes, então

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underbrace{\mathbf{0}}_{(p \times m)}. \quad (4)$$

As suposições (3) e (4) juntamente com a relação (2) constituem o modelo fatorial ortogonal. A estrutura de covariância para o modelo fatorial ortogonal, pode ser obtida da relação:

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'], \quad (5)$$

onde de (2), obtém-se

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}, \quad (6)$$

e ainda temos que

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}, \quad (7)$$

logo

$$\begin{aligned} V(X_i) &= \ell_{i1}^2 + \dots + \ell_{im}^2 + \psi_i, \\ \text{Cov}(X_i, X_k) &= \ell_{i1}\ell_{k1} + \dots + \ell_{im}\ell_{km} \end{aligned} \quad (8)$$

e de (7), temos

$$\text{Cov}(X_i, F_j) = \ell_{ij}. \quad (9)$$

A parte da variância que a i -ésima variável contribui para os m fatores comuns é chamada de comunalidade, que é dada por

$$h_i^2 = \ell_{i1}^2 + \dots + \ell_{im}^2 \quad (10)$$

onde $\psi_i = 1 - h_i^2$ denotam as variâncias específicas.

Mais detalhes sobre o modelo fatorial ortogonal podem ser encontrados em Johnson e Wichern (2002).

2.1.3 Método das Componentes Principais

Usando a decomposição espectral podemos fatorar a matriz de covariância Σ . Sejam (λ_i, e_i) , $i = 1, \dots, p$, pares de autovalores-autovetores da matriz Σ , com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, $i = 1, \dots, p$. Então

$$\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_p e_p e_p' \quad (11)$$

A equação (11) ajusta a estrutura de covariância prescrita para a análise do modelo fatorial, tendo como fator qualquer variável ($m = p$) e variância específica $\psi_i = 0$ para todo i . A matriz de cargas tem a j -ésimo coluna dada por $\sqrt{\lambda_j} e_j$. Desta maneira, podemos reescrever a equação (11) como

$$\underbrace{\Sigma}_{(p \times p)} = \underbrace{L}_{(p \times p)} \underbrace{L'}_{(p \times p)} + \underbrace{0}_{(p \times p)} = L L' \quad (12)$$

Entretanto na análise fatorial essa representação de Σ é exata, isto é, o número de fatores comuns é igual ao número de variáveis. Podemos optar por modelos que expliquem a estrutura de covariância usando poucos fatores comuns. Para isso, devemos eliminar os últimos $p - m$ autovalores, que pouco contribuem para Σ . Negligenciando esta contribuição, obtemos a aproximação

$$\Sigma = \underbrace{L}_{p \times m} \underbrace{L'}_{m \times p} \quad (13)$$

A representação (13) considera que os fatores específicos ε definidos no modelo fatorial ortogonal são de menor importância e podem ser ignorados na fatoração de Σ . Se os fatores específicos forem incluídos no modelo, a contribuição deles pode ser estimada tomando-se os elementos da diagonal de $\Sigma - L L'$, onde $L L'$ é dado em (13).

Considerando-se os fatores específicos, encontramos a seguinte aproximação

$$\Sigma = L L' + \Psi, \quad (14)$$

onde $\psi_i = \sigma_{ii} - \sum_{j=1}^m \ell_{ij}^2$ para $i = 1, 2, \dots, p$.

A representação em (14), quando aplicada a matriz de covariância amostral S ou a matriz de correlação amostral R é conhecida como a solução por componentes principais do modelo fatorial.

Em geral a proporção da variância total amostral no j -ésimo fator é dada por:

$$\begin{cases} \frac{\lambda_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{para a análise fatorial de } S, \\ \frac{\lambda_j}{p} & \text{para a análise fatorial de } R. \end{cases} \quad (15)$$

O critério dado em (15) geralmente é usado para determinar o número de fatores comuns apropriado. Quanto maior o número de fatores, maior é a proporção da variância amostral explicada pelo modelo.

Quando se fatora a matriz R na análise fatorial, o número de fatores comuns a ser extraído, geralmente é dado pelo número de autovalores maiores que 1 segundo o critério de Kaiser (Hair, 2002).

2.1.4 Critério de Rotação Varimax

O Método foi proposto por Kaiser (1958) e tem por objetivo encontrar uma matriz $\underbrace{T}_{m \times m}$ responsável pela maximização da variação dos quadrados das cargas fatoriais originais das colunas da matriz $\underbrace{L}_{p \times m}$. Mais detalhes podem ser encontrados em Mingoti (2007).

2.1.5 Estimação dos Escores (Mínimos Quadrados)

Após a identificação e análise dos fatores rotacionados relacionados com as variáveis X_i , $i = 1, 2, \dots, p$ é necessário calcular os escores para cada elemento amostral, de modo a utilizar esses valores em outras análises. Neste trabalho os escores fatoriais serão necessários para a determinação dos quilômetros da rodovia relacionados com cada fator.

Os escores fatoriais são determinados usando-se os mínimos quadrados e é dado por:

$$F_{jk} = (L' \Psi^{-1} L) L' \Psi^{-1} Z_k, \tag{16}$$

onde $Z_k = (Z_{1k} \ Z_{2k} \ \dots \ Z_{pk})$ é o vetor de observações do k -ésimo elemento amostral padronizado.

3 Materiais e Métodos

A partir do banco de dados cedido pela Polícia Rodoviária Federal do Paraná, contendo as informações referentes aos 10.096 acidentes ocorridos no período de janeiro de 2007 à novembro de 2009 na BR-277 (Figura 1), será analisado os pontos onde ocorrem maior número de acidentes, encontrando as possíveis variáveis que estejam relacionadas com esses acidentes.



Figura 1: BR-277: início em Paranaguá (KM 0) e término em Foz do Iguaçu (KM 731) - PR

O Banco de dados contém informações sobre os tipos de acidentes (capotamento, colisão transversal, colisão traseira, entre outras), condições meteorológicas (chuva, céu claro, nublado, entre outras), tipo de pista (simples, dupla e múltipla), traçado da pista (curva, reta), causas do acidente (falta de atenção, sono, ingestão de bebidas alcoólicas, não guardar distância de segurança, entre outras), inclinação da pista (muito inclinado, pouco inclinado), conservação da pista (bom, ruim, regular), fase do dia em que ocorreu o acidente (manhã, tarde, anoitecer, noite), sentido da via (crescente ou decrescente), entre outras. Ao todo são 80 variáveis qualitativas selecionadas no banco de dados para fazer a análise fatorial. Todas essas variáveis serão analisadas de modo a agrupá-las, para uma melhor visualização do perfil dos acidentes ocorridos na BR-277.

O trabalho foi realizado de acordo com o fluxograma abaixo:

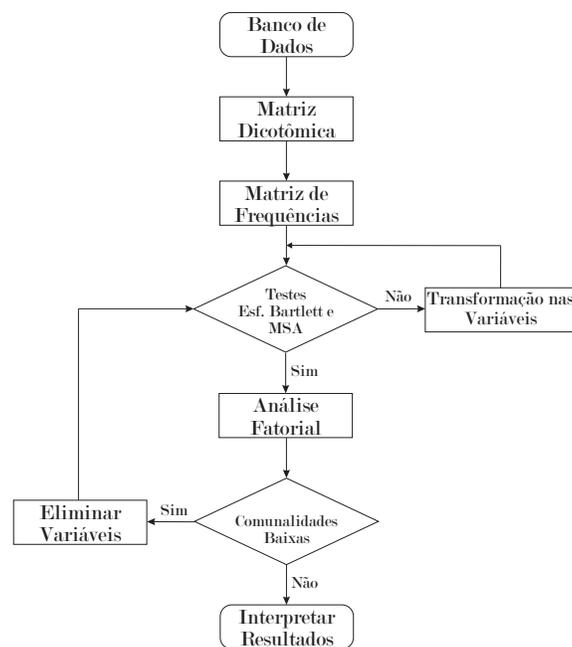


Figura 2: Fluxograma da Aplicação da Análise Fatorial

no qual observamos inicialmente o banco de dados de onde foi selecionado um conjunto de variáveis qualitativas, onde, em seguida obteve-se uma matriz dicotômica, desmembrando as variáveis originais, onde cada elemento da matriz indica a ocorrência (1) ou não (0) de uma variável em um determinado acidente. A partir da matriz dicotômica obteve-se a matriz de frequências (quilômetro da rodovia \times variáveis) onde cada elemento da matriz é o número de acidentes em um quilômetro e em determinada variável. A matriz de correlação da matriz de frequências de tamanho n é utilizada na análise fatorial para obter os fatores e escores fatoriais. O cálculo dos escores fatoriais é utilizado para obter um *ranking* dos quilômetros da BR-277 de acordo com cada fator (onde cada fator está associado com um conjunto de variáveis).

Devido a constantes mudanças na rodovia, tais como duplicação, modificação de trevos, pavimentação de acostamentos, entre outros, a Análise Fatorial foi realizada três vezes, em períodos distintos, para verificar se houve modificação dos quilômetros mais perigosos, de acordo com cada período. Esses períodos são:

AF1. de janeiro de 2007 a novembro de 2009;

AF2. de janeiro de 2009 a novembro de 2009;

AF3. meses de férias (janeiro, fevereiro, julho e dezembro) de janeiro de 2007 a novembro de 2009.

Para cada um destes períodos foi obtida uma matriz de frequências diferente, sendo que o número de variáveis de uma análise para a outra difere, pois não houve a ocorrência de determinadas variáveis em alguns períodos. Além disso, algumas variáveis foram eliminadas, por serem combinações lineares de outras, impedindo a realização do Teste de Esfericidade de Bartlett que utiliza o determinante da matriz de correlação da matriz de frequências.

Na **AF1** foi utilizada uma matriz de 637 linhas \times 93 variáveis. Neste caso a análise fatorial foi aplicada três vezes sendo que as duas primeiras tinham o objetivo de eliminar variáveis com comunalidades baixas (abaixo de 0,7). Na Análise Fatorial interpretada (última) foram consideradas comunalidades acima de 0,5. Para determinação do número de fatores foi utilizado o critério de Kaiser (número de fatores igual ao número de autovalores maiores que a unidade). O critério de rotação varimax foi utilizado apenas na última análise, para facilitar a interpretação dos fatores, sendo que nesta última Análise Fatorial, variáveis com comunalidades abaixo de 0,5 não foram incluídas no modelo. Outros critérios foram utilizados, porém não produziram fatores mais simples que o critério varimax. Quanto aos escores, eles foram obtidos utilizando-se os mínimos quadrados ponderados.

Os mesmos procedimentos foram utilizados na **AF2** e **AF3**, com 574 linhas \times 80 variáveis e 488 linhas \times 92 variáveis, respectivamente. Nestes dois casos, a primeira Análise Fatorial foi feita para eliminar variáveis com comunalidades baixas e a segunda foi interpretada.

Todos os cálculos efetuados nesse trabalho foram realizados por meio de um programa implementado em Fortran.

4 Resultados

Os resultados apresentados a seguir são da **AF1**. Os resultados das outras análises não serão apresentados uma vez que são análogos aos da **AF1**.

Ao conjunto de dados inicial (637 linhas \times 93 variáveis) aplicou-se três vezes a análise fatorial. É importante destacar que as duas primeiras análises foram realizadas com o objetivo de eliminar variáveis, devido a grande quantidade de variáveis presentes na matriz. Apenas a última Análise Fatorial foi interpretada. Os resultados de cada Análise Fatorial podem ser encontrados na tabela 2, lembrando que a porcentagem da variância explicada na primeira Análise Fatorial foi calculada para as 93 variáveis utilizadas, na segunda, 73 variáveis, e na terceira, 66 variáveis.

Tabela 2: Resultados de cada Análise Fatorial

Análise Fatorial	<i>p</i> - value Teste de Esfericidade de Bartlett	MSA (KMO)	Número de Fatores	Variância Explicada (%)
1	≈ 0	0,81	17	80,94
2	≈ 0	0,91	11	84,09
3	≈ 0	0,91	9	86,86

Pode-se observar que os dois testes (Teste de Esfericidade de Bartlett e MSA de Kaiser-Meyer-Olkin) indicam a Análise Fatorial adequada ao conjunto de dados em cada Análise Fatorial feita. O valor de p no Teste de Esfericidade de Bartlett indica que a hipótese nula deve ser rejeitada, ou seja, a matriz de correlações pode ser considerada diferente da matriz identidade. O critério usado para eliminar variáveis foi a comunalidade abaixo de 0,7 nas duas primeiras análises. Na terceira Análise Fatorial apenas uma variável (Causa Do Acidente_Defeito Mecânico Em Veículo) teve comunalidade abaixo de 0,5, portanto não foi incluída no modelo.

Os fatores obtidos foram:

Fator 1: Sentido da Via_Crescente, Sentido da Via_decrescente, Tipo de Acidente_Colisão Transversal, Tipo de Acidente_Colisão Traseira, Tipo de Acidente_Colisão Lateral, Causa do Acidente_Falta de Atenção, Causa do Acidente_Não Guardar distância de Segurança, Causa do Acidente_desobediência à Sinalização, Condição Pista_Seca, Condição Pista_Molhada, Restrição Visibilidade_Inexistente, Sinalização_Vertical, Horizontal Sinalização_Vertical, Fase do Dia_Plena Noite, Fase do Dia_Pleno Dia, Fase do Dia_Anoitecer, Condição Meteorológica_Céu Claro, Condição Meteorológica_Sol, Condição Meteorológica_Nublado, Sinalização Luminosa_Inexistente, Sinalização Luminosa_Funciona, Acostamento_Sim, Acostamento_Não, Desnível Acostamento_Não, Conservação Acostamento_Bom, Canteiro_Não, Canteiro_Sim, Obstáculo ao Cruzamento_Meio-fio, Pista_Simples, Perfil Pista_Em Nível, Perfil Pista_Rampa < 3%, Traçado Pista_Reta, Curva Vertical Pista_Não, Superelevação Pista_Não, Superlargura Pista_Não, Estreitamento Pista_Não, Existe Estreitamento Pista_Provisão e Conservação Pista_Bom.

Fator 2: Tipo de Acidente_Colisão com Objeto Fixo, Causa do Acidente_Velocidade Incompatível, Condição Meteorológica_Chuva, Obstáculo ao Cruzamento_Muro e Traçado Pista_Curva.

Fator 3: Tipo de Acidente_Saída de Pista, Obstáculo ao Cruzamento_Sarjeta, Obstáculo ao Cruzamento_Não Existe, Obstáculo ao Cruzamento_Canal e Pista_Dupla.

Fator 4: Condição Pista_Com Buraco, Sinalização_Inexistente, Conservação Acostamento_Ruim e Conservação Pista_Com Erosão.

Fator 5: Obstáculo Ao Cruzamento_Tela e Obstáculo Ao Cruzamento_Com Anti-ofuscante.

Fator 6: Desnível Acostamento_Sim, Acostamento Pavimentado_Sim, Perfil Pista_Rampa \geq 3%.

Fator 7: Sinalização_Vertical, Horizontal, Manual.

Fator 8: Restrição Visibilidade_Poeira/fumaça/neblina e Condição Meteorológica_Nevoeiro/neblina.

Fator 9: Causa do Acidente_Animalis na Pista e Tipo de Acidente_Atopelamento de Animal.

A descrição dos nove fatores segue abaixo:

- O fator 1 (falta de atenção do motorista), é o mais significativo, pois explica 44,98% da variância dos dados. Todas as variáveis deste fator têm comunalidades acima de 0,69. Este fator reúne variáveis, tais como os três tipos de colisão (transversal, traseira e lateral), ligado a falta de atenção dos motoristas e também o fato de não guardarem a distância de segurança e desobediência à sinalização. Pode-se ainda perceber que as condições meteorológicas estão quase todas presentes nesse fator, exceto a chuva, o mesmo ocorre com as condições de pista, entre outras. Quanto ao tipo de pista, observa-se pista reta, simples e com inclinação menor que 3% sem superelevação, superlargura ou estreitamento. Analisando os escores fatoriais desse fator e selecionando apenas os três primeiros (mais expressivos), consegue-se identificar os quilômetros km 584 (Cascavel - PR), km 726 e km 728 (Foz do Iguaçu - PR) que estão intimamente ligados com as características desse fator.

- O fator 2 (imprudência do motorista em dias de chuva) explica 8,89% da variância dos dados. Todas as variáveis deste fator têm comunalidades acima de 0,79. Neste fator observa-se a junção das variáveis velocidade incompatível com pista inclinada, curva, obstáculo ao cruzamento com muro e colisão com objeto fixo. Pelos escores fatoriais, consegue-se identificar os três quilômetros onde ocorrem acidentes com essa característica: km 37, km 36 e KM 84 (entre Paranaguá - PR e Curitiba - PR), o que indica que a maioria dos acidentes com esse perfil estão ocorrendo nesses quilômetros da BR-277.
- O fator 3 (desobediência do motorista) agrupa as variáveis relacionadas ao tipo de acidente saída de pista aliado à pista dupla, sem obstáculos ao cruzamento, com canal e sarjeta. O nome desse fator se deve ao fato da variável pista dupla, que permite que os motoristas atinjam velocidades maiores, e também a variável saída de pista. Este fator explica 7,38% da variância dos dados e as comunalidades das variáveis são maiores ou iguais a 0,72. Acidentes com essa característica se concentram principalmente nos quilômetros km 723, km 721 e km 726 (Foz do Iguaçu - PR).
- No fator 4 (pista mal conservada) tem-se as variáveis: pista com buraco e erosão, sinalização inexistente e conservação de acostamento ruim. A variância explicada deste fator é 6,52% e as variáveis possuem comunalidades maiores ou iguais a 0,74. Os acidentes que ocorrem na rodovia com essa característica estão concentrados principalmente nos quilômetros km 3, km 2 e km 1 (Paranaguá - PR).
- No fator 5 (pista com antiofuscante) tem-se as variáveis: obstáculo ao cruzamento com tela e com antiofuscante. Este fator explica 4,27% da variância e as comunalidades são altas, maiores ou iguais a 0,87. Acidentes envolvendo esses tipos de obstáculos ao cruzamento ocorrem principalmente nos quilômetros KM 80, km 81 e km 82 (Curitiba - PR).
- O fator 6 (pista inclinada com desnível para acostamento) agrupa as variáveis: acostamento em desnível e pista com inclinação maior que ou igual a 3%. Este fator explica 5,49% da variância dos dados e as comunalidades das variáveis estão acima de 0,7. Nesse fator os escores fatoriais apontam os quilômetros km 593, km 583 e km 588 (próximo a Cascavel - PR) como sendo os de maior peso para acidentes na rodovia com essas características.
- O fator 7 (pista sinalizada) é composto pelas variáveis causa do acidente (ingestão de álcool) com correlação baixa, estreitamento de pista provisão e sinalização vertical, horizontal e manual. Esse fator explica 2,92% da variância dos dados e as comunalidades das variáveis estão acima de 0,68. Pelos escores fatoriais pode-se observar que, nos quilômetros km 66, km 60 e km 65 (entre Paranaguá - PR e Curitiba - PR) ocorrem acidentes com pista sinalizada.
- O fator 8 (restrição de visibilidade) envolve acidentes ligados à condições meteorológicas (nevoeiro e neblina) e restrição de visibilidade (fumaça, poeira e neblina), explicando 3,16% da variância. As variáveis deste fator apresentam comunalidades altas: acima de 0,9. Acidentes com esse perfil ocorrem principalmente nos quilômetros km 42, km 43 (entre Paranaguá - PR e Curitiba - PR) e o km 488 (próximo a Nova Laranjeiras - PR).

- O fator 9 (animais na pista), está ligado à acidentes com atropelamento de animais e animais na pista, explicando 3,26% da variância dos dados. Assim como no fator restrição de visibilidade as variáveis apresentam comunalidades acima de 0,9. Pode-se identificar os quilômetros km 702 (entre Medianeira - PR e Foz do Iguaçu - PR), km 3 (Paranaguá - PR) e km 76 (próximo a Curitiba - PR) ligados a esse fator.

Calculando-se os escores fatoriais finais ponderados, por meio dos nove fatores obtidos, e criando uma escala entre 0 e 10, encontra-se os quilômetros mais problemáticos de acordo com o número de acidentes registrados.

Tabela 3: Escores fatoriais finais ponderados em ordem decrescente

Quilômetro	Local	Escore
584	Cascavel	10,0
726	Foz do Iguaçu	3,6
728	Foz do Iguaçu	3,4
1	Paranaguá	2,9
593	Cascavel	2,9
4	Paranaguá	2,8
583	Cascavel	2,7
657	Matelândia	2,6
725	Foz do Iguaçu	2,6
⋮	⋮	⋮
180	próx. Palmeira	0,0
443	próx. Laranjeiras do Sul	0,0
386	próx. Guarapuava	0,0

5 Conclusão

A partir das três análises realizadas anteriormente, pôde-se verificar que todas as análises realizadas apresentaram resultados semelhantes, ou seja, a maior parte dos fatores foram coincidentes e os piores quilômetros com relação ao número de acidentes também. O fator com maior variabilidade explicada, em todas as análises foi a falta de atenção do motorista. Isso mostra que independente do período analisado, os motoristas contribuem muito nos acidentes. Levando em consideração o fator imprudência do motorista, a contribuição dos motoristas nos acidentes aumenta ainda mais. Outros fatores tais como animais na pista, restrições de visibilidade e pista mal conservada influenciam, porém de maneira menos significativa.

Um outro detalhe que pode ser observado é que na Análise Fatorial entre 2007 e 2009, os fatores são mais definidos do que na análise aplicada somente em 2009. Quanto à Análise Fatorial aplicada em meses de férias pode-se observar a ocorrência de dois novos fatores que não apareceram nas análises anteriores: defeito mecânico em veículo e restrição de visibilidade devido a veículos estacionados. Isso se deve ao fato de muitas famílias saírem de viagem sem verificar as condições do automóvel.

Todas as análises definem o mesmo perfil, quanto aos fatores, com pequenas diferenças entre uma e outra. Entretanto, cabe ressaltar que a falta de atenção do motorista é o maior problema quanto aos acidentes de trânsito na BR-277 no período de janeiro de 2007 a novembro de 2009. Outras condições influenciam, porém de maneira bem menos representativas.

Analisando os escores fatoriais finais ponderados, obtidos em cada análise observa-se que o quilômetro 584 que se encontra no município de Cascavel é o mais problemático com relação

aos acidentes, com 336 acidentes entre 2007 e novembro de 2009. Uma possível causa é a grande quantidade de veículos que circula nesse quilômetro, o que se deve ao trevo cataratas localizado nesse quilômetro que faz a junção da BR-277 com a BR-369, BR-467 e a entrada de Cascavel.

Outros quilômetros, tais como o km 726 e km 728 em Foz do Iguaçu, km 74 em Curitiba e km 1 em Paranaguá também aparecem como quilômetros com muitos acidentes, porém com escores menores do que o observado no km 584. Mesmo assim nesses quilômetros foram registrados 473 acidentes entre 2007 e novembro de 2009. Esses quilômetros estão localizados em perímetros urbanos, nesses locais o fluxo de veículos geralmente é intenso, causando uma quantidade maior de acidentes.

Referências

- Bartilotti, C. B.**, Fatores da Senso-Percepção Relacionados à Atividade do Condutor no Sistema Trânsito: Construção e Validação de um Instrumento de Medida. Tese de Doutorado, UFSC, 2009.
- Hair, J. F.**, et al, *Applied Multivariate Statistical Analysis*. New York, 2002.
- Hills, M.**, Book Review. *Applied Statistics*. 4, 339–340, 1977.
- Johnson, R. A. e Wichern, D. W.**, *Applied Multivariate Statistical Analysis*. New York, 2002.
- Kaiser, H. F.**, The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200, 1958.
- Marques, A. F.**, Aplicação da análise multivariada na infraestrutura e no desempenho das escolas públicas do Ensino Fundamental e Médio pertencentes ao Núcleo Regional de Educação de Paranavaí. *Acta Scientiarum*, 32, 75–81, 2010.
- Mingoti, S. A.**, *Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada*. Belo Horizonte, 2007.
- Ozerenko, A. A.**, et al, Factorial analysis in selecting the zeolite catalyst for 2,6-dimethylnaphthalene synthesis. *Coke and Chemistry*, 50, 200–206, 2007.
- Pires, T., Maia, A.**, Relação entre saúde física e transtorno de estresse pós-traumático em vítimas de acidentes rodoviários. *Série Ciência Social Ambiental*, 1, 14–26, 2009.
- Queiroz, J. C. B.**, Geochemical characterization of heavy metal contaminated area using multivariate factorial kriging. *Environmental Geology*, 55, 95–105, 2007.
- Rencher, A. C.**, *Methods of Multivariate Analysis*. New York, 2002.
- Silva, L. C. P.**, Análise Psicométrica do Inventário Multidimensional do Estilo de Dirigir. Dissertação de Mestrado, FFCLRP, 2004.
- Spearman, C.**, General Intelligence Objectively Determined and Measured. *American Journal of Psychology*, 15, 201–293, 1904.