

## ESTUDO DOS ACIDENTES OCORRIDOS NA BR-277 EM MESES DE FÉRIAS POR MEIO DA ANÁLISE MULTIVARIADA

Ferreira, M. V. R. P. G.<sup>c</sup>, Balbo, F. A. N.<sup>c</sup>, Gramani, L. M.<sup>a</sup>, Neto, A. C.<sup>b</sup>, Bogo, R. L.<sup>c</sup> e Lobeiro, A. M.<sup>c</sup>

<sup>a</sup>*Departamento de Matemática, Universidade Federal do Paraná, Centro Politécnico, Curitiba, Paraná, Brasil. <http://www.mat.ufpr.br/>*

<sup>b</sup>*Departamento de Estatística, Universidade Federal do Paraná, Centro Politécnico, Curitiba, Paraná, Brasil. <http://www.est.ufpr.br/>*

<sup>c</sup>*Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Centro Politécnico, Curitiba, Paraná, Brasil. <http://www.ppgmne.ufpr.br/>*

**Palavras Chave:** Acidentes em Rodovias, Análise Fatorial, Método das Componentes Principais.

**Resumo.** A otimização e controle do fluxo de tráfego de veículos ao longo de uma pista ou várias pistas, em estradas, rodovias ou vias urbanas, é um interessante e abrangente campo de interação entre matemáticos, engenheiros e estatísticos. Com o aumento significativo do tráfego de veículos em rodovias federais (Brasil), principalmente em meses de férias (dezembro, janeiro, fevereiro e julho), ocorre também um crescimento no número de acidentes. A partir dos dados referentes aos acidentes dos meses de férias no período de 2006 à novembro de 2009, ocorridos na BR-277, Rodovia Federal do Paraná, que vai desde Paranaçu até Foz do Iguaçu, realizou-se um estudo utilizando uma técnica estatística multivariada denominada análise fatorial pelo método das componentes principais, onde foi possível identificar inter-relações entre um grande número de variáveis descritas nos acidentes. Desta forma, foi possível identificar relações entre as causas dos acidentes e condições da pista, para que possa ser feito planejamentos futuros para resolver este problema. Além disso, utilizou-se os escores fatoriais, com os quais foi possível montar um ranking com os pontos mais perigosos da rodovia, dando uma visão geral sobre do perfil dos acidentes na BR-227.

## 1 INTRODUÇÃO

A BR-277 é uma rodovia federal transversal do Brasil 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, por ser conhecida como o “corredor do Mercosul”. Desempenha ainda, um importante papel no escoamento de produtos agropecuários produzidos no sudoeste do Paraná.

Por se localizar em uma região estratégica, cada vez mais o tráfego de veículos tem aumentado e conseqüentemente o número de acidentes também tem crescido<sup>1</sup>. Com o objetivo de encontrar os quilômetros da rodovia com maior número de acidentes e identificar suas possíveis causas, será usado uma técnica estatística multivariada (em um banco de dados referente aos 2.934 acidentes na BR-277 nos meses de janeiro, fevereiro, julho e dezembro entre janeiro de 2007 e novembro de 2009), conhecida como a análise fatorial (Johnson e Wichern, 2002).

A análise fatorial tem por objetivo agrupar variáveis correlacionadas em fatores, resumindo assim a grande quantidade de variáveis envolvidas nos acidentes, em fatores, para depois relacionar os quilômetros das rodovias com as principais causas e tipos de acidentes com o uso dos escores fatoriais. É a técnica multivariada que apresentou melhores resultados, uma vez que a análise de agrupamentos (Mingoti, 2007) é de difícil interpretação devido ao grande número de variáveis e 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.

Na literatura podem ser encontrados trabalhos que usam a análise fatorial para resolver problemas no trânsito, como exemplo temos Bartilotti (2009) que analisa a percepção do motorista no trânsito, Pires e Maia (2009) que faz um estudo sobre as vítimas de acidentes no trânsito, Silva (2004) que estuda o comportamento dos motoristas, entre outros.

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.

## 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

<sup>1</sup>De acordo com o banco de dados da Polícia Rodoviária Federal do Paraná em 2007 foram registrados 2.881 acidentes, em 2008, 3.474 acidentes e de janeiro à novembro de 2009, 3.728 acidentes na BR-277.

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)<sup>2</sup>. 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 e Anderson (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 inter-relações (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.

Em linhas gerais Mingoti (2007) afirma que, o que se espera é que as variáveis originais estejam agrupadas em subconjuntos de novas variáveis mutuamente não correlacionadas. Deste modo, em casos no qual se tem um grande número de variáveis medidas e correlacionadas entre si, seria possível, a partir da análise fatorial, identificar um número menor de novas variáveis alternativas, não correlacionadas e que de algum modo sumarizassem as informações principais das variáveis originais.

A análise fatorial pode ser trabalhada de forma exploratória ou confirmatória. No primeiro caso, busca-se encontrar os fatores subjacentes às variáveis originais amostradas, sem ter noção de quantos fatores serão obtidos e de que maneira as variáveis estarão agrupadas. Já na análise fatorial confirmatória, tem-se um modelo inicial pré-estabelecido, onde se deseja verificar se este modelo proposto é consistente com os dados amostrais disponíveis. Neste trabalho será utilizada apenas a análise fatorial exploratória.

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) que serão descritos abaixo.

#### 1. Teste de Esfericidade de Bartlett

Testa a hipótese da matriz de correlação populacional  $\rho$  ser igual a matriz identidade  $I$ , ou seja, admitir a ausência de correlação entre as variáveis estudadas.

---

<sup>2</sup>Charles 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".

A hipótese do teste é dada por  $H_0 : \rho \neq \mathbf{I}$  e a estatística pode ser calculada por:

$$\chi_\nu^2 = - \left[ (n-1) - \frac{2p+5}{6} \right] \ln |\mathbf{R}| \quad (1)$$

onde  $\chi_\nu^2$  é uma distribuição qui-quadrado com  $\nu = \frac{p(p-1)}{2}$  graus de liberdade,  $n$  o tamanho da amostra,  $p$  o número de variáveis e  $|\mathbf{R}|$  o determinante da matriz de correlação.

Quando rejeita-se a hipótese nula, as variáveis apresentam correlações significativas, indicando a continuidade do procedimento.

## 2. Medida de adequacidade da amostra de Kaiser-Meyer-Olkin (KMO)

Essa medida é representada por um índice (MSA) que avalia a adequação da utilização da análise fatorial, sendo calculada por:

$$MSA = \frac{\sum_{j \neq k} r_{jk}^2}{\sum_{j \neq k} r_{jk}^2 + \sum_{j \neq k} q_{jk}^2} \quad (2)$$

onde:  $r_{jk}$  são os elementos da matriz de correlação original ( $j \neq k$ ) e  $q_{jk}$  são os elementos da matriz de correlação anti-imagem<sup>3</sup> também com  $j \neq k$ . De acordo com [Hair e Anderson \(2002\)](#), valores maiores que 0,5 indicam que a análise fatorial é apropriada.

Mais detalhes sobre ambos os testes podem ser encontrados em [\(Rencher, 2002\)](#).

### 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 específicos. O modelo fatorial é

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots && \vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (3)$$

ou em notação matricial

$$\underbrace{\mathbf{X} - \boldsymbol{\mu}}_{(p \times 1)} = \underbrace{\mathbf{L}}_{(p \times m)} \cdot \underbrace{\mathbf{F}}_{(m \times 1)} + \underbrace{\boldsymbol{\varepsilon}}_{(p \times 1)}, \quad (4)$$

onde  $\ell_{ij}$  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  $\varepsilon_i$  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.

<sup>3</sup>Matriz de correlações parciais entre variáveis. A diagonal contém as medidas de adequação da amostra para cada variável, e os demais valores são correlações parciais entre variáveis

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

$$\begin{aligned} E(\mathbf{F}) &= \underbrace{\mathbf{0}}_{m \times 1}, & \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') &= \underbrace{\mathbf{I}}_{(m \times m)} \\ E(\boldsymbol{\varepsilon}) &= \underbrace{\mathbf{0}}_{(p \times 1)}, & \text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') &= \underbrace{\boldsymbol{\Psi}}_{(p \times p)} \end{aligned} \quad (5)$$

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

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underbrace{\mathbf{0}}_{(p \times m)}. \quad (6)$$

As suposições (5) e (6) juntamente com a relação (4) constituem o modelo fatorial ortogonal. A estrutura de covariância para o modelo fatorial ortogonal, pode ser obtida da relação:

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'], \quad (7)$$

onde de (4), obtém-se

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}, \quad (8)$$

e ainda temos que

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}, \quad (9)$$

logo

$$\begin{aligned} V(X_i) &= \ell_{i1}^2 + \dots + \ell_{im}^2 + \psi_i, \\ \text{Cov}(X_i, X_k) &= \ell_{i1}\ell_{k1} + \dots + \ell_{im}\ell_{km} \end{aligned} \quad (10)$$

e de (9), temos

$$\text{Cov}(X_i, F_j) = \ell_{ij}. \quad (11)$$

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

$$h_i^2 = \ell_{i1}^2 + \dots + \ell_{im}^2 \quad (12)$$

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  $\boldsymbol{\Sigma}$ . Sejam  $(\lambda_i, \mathbf{e}_i)$ ,  $i = 1, \dots, p$ , pares de autovalores-autovetores da matriz  $\boldsymbol{\Sigma}$ , com  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ,  $i = 1, \dots, p$ . Então

$$\boldsymbol{\Sigma} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p'. \quad (13)$$

A equação (13) 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 (13) como

$$\underbrace{\Sigma}_{(p \times p)} = \underbrace{L}_{(p \times p)} \underbrace{L'}_{(p \times p)} + \underbrace{0}_{(p \times p)} = L L' \quad (14)$$

Entretanto na análise fatorial essa representação de  $\Sigma$  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  $\Sigma$ . Negligenciando esta contribuição, obtemos a aproximação

$$\Sigma = \underbrace{L}_{p \times m} \underbrace{L'}_{m \times p} \quad (15)$$

A representação (15) considera que os fatores específicos  $\varepsilon$  definidos no modelo fatorial ortogonal são de menor importância e podem ser ignorados na fatoração de  $\Sigma$ . 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 (15).

Considerando-se os fatores específicos, encontramos a seguinte aproximação

$$\Sigma = L L' + \Psi, \quad (16)$$

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

A representação em (16), 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:

$$\begin{cases} \frac{\lambda_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{para a análise fatorial de } S, \\ \frac{\lambda_j}{p} & \text{para a análise fatorial de } R. \end{cases} \quad (17)$$

O critério dado em (17) 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 e Anderson, 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 Estimación 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:

$$F_{jk} = (L'\Psi^{-1}L)^{-1}L'\Psi^{-1}Z_k, \quad (18)$$

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 2934 acidentes ocorridos no período de janeiro de 2007 à novembro de 2009 no período de férias (janeiro, fevereiro, julho e dezembro) 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

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 92 variáveis qualitativas selecionadas do 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, 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 (2.934 acidentes  $\times$  92 variáveis), desmembrando as variáveis originais, onde cada elemento da matriz indica a ocorrência (1) ou não (0) de uma

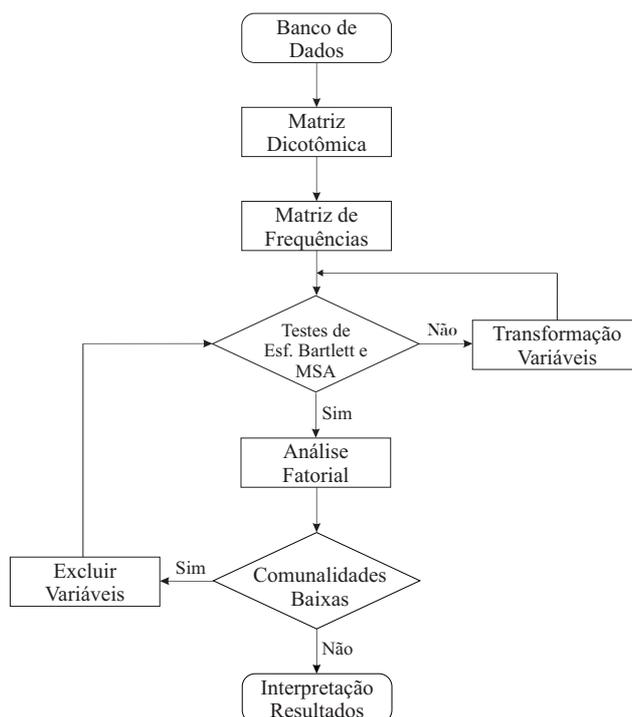


Figura 2: Fluxograma da Aplicação da Análise Fatorial

variável em um determinado acidente. A partir da matriz dicotômica obteve-se a matriz de frequências ( $488 \text{ KM} \times 92$  variáveis) onde cada elemento da matriz é o número de acidentes em um quilômetro com relação a determinada variável. A matriz de correlação da matriz de frequências 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 a um conjunto de variáveis).

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

#### 4 RESULTADOS

A primeira etapa do trabalho foi encontrar os quilômetros da rodovia onde se concentram a maior parte dos acidentes. No histograma (ver Figura 3), podemos observar uma grande concentração de acidentes nos primeiros 100 quilômetros da rodovia (entre Paranaguá - PR e Curitiba - PR) e também do quilômetro 500 até o final da rodovia, ou seja o quilômetro 731 (Entre Nova Laranjeiras - PR e Foz do Iguaçu - PR).

Na primeira análise fatorial, os testes de esfericidade de Bartlett e da Medida de Adequacidade da Amostra (MSA) de Kaiser-Meyer-Olkin foram realizados, onde obteve-se:

- No teste de esfericidade de Bartlett,  $p\text{-value} \cong 0$ , mostrando que a matriz de correlação difere significativamente da matriz identidade;
- Para a Medida de Adequacidade da Amostra (MSA) de Kaiser-Meyer-Olkin,  $MSA = 0,73$ , o que indica a análise fatorial adequada ao conjunto de dados.

Realizando a análise fatorial obteve-se 21 fatores que explicam  $78,2\%$  da variância do conjunto de dados original. O número de fatores foi escolhido pelo critério de Kaiser (Hair e

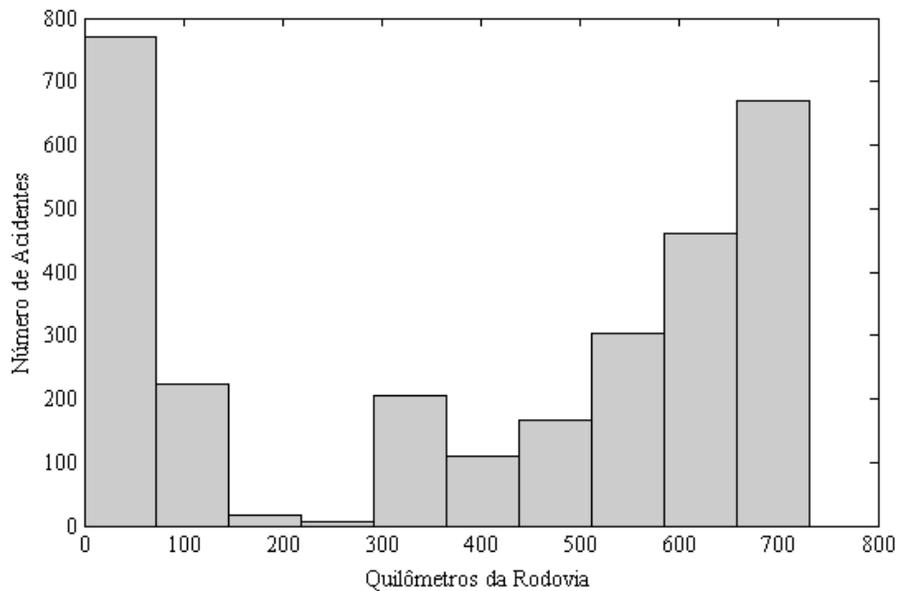


Figura 3: Distribuição dos Acidentes na BR-277 (Jan/2009 à Nov/2009)

Anderson, 2002). Analisando-se as comunalidades, 33 variáveis foram eliminadas com comunalidades abaixo de 0,6.

Aplicando pela segunda vez os testes para a análise fatorial, na nova matriz (488 KM  $\times$  59 variáveis), resulta:

- No teste de esfericidade de Bartlett,  $p$ -value  $\cong 0$ , mostrando grande correlação entre as variáveis;
- Para a Medida de Adequacidade da Amostra (MSA) de Kaiser-Meyer-Olkin,  $MSA = 0,92$ , o que indica a análise fatorial adequada ao conjunto de dados.

Nesta análise fatorial, usando o método das componentes principais e o critério de rotação varimax, obteve-se 10 fatores (critério de Kaiser) que explicam 83,93% da variância total do conjunto de dados. Quatro variáveis foram eliminadas com comunalidades baixas. Os fatores de 1 a 10 são descritos em detalhes abaixo:

**Fator 1:** Sentido da Via\_Crescente ( $X_1$ ), Sentido da Via\_decrescente ( $X_2$ ), Tipo de Acidente\_Colisão Transversal ( $X_3$ ), Tipo de Acidente\_Colisão Traseira ( $X_5$ ), Tipo de Acidente\_Colisão Lateral ( $X_6$ ), Causa do Acidente\_Falta de Atenção ( $X_9$ ), Causa do Acidente\_Não Guardar distância de Segurança ( $X_{11}$ ), Causa do Acidente\_desobediência À Sinalização ( $X_{13}$ ), Condição Pista\_Seca ( $X_{15}$ ), Restrição Visibilidade\_Inexistente ( $X_{18}$ ), Sinalização\_Vertical, Horizontal ( $X_{21}$ ), Fase do dia\_Plena Noite ( $X_{22}$ ), Fase do dia\_Pleno dia ( $X_{24}$ ), Condição Meteorológica\_Céu Claro ( $X_{25}$ ), Condição Meteorológica\_Sol ( $X_{26}$ ), Condição Meteorológica\_Nublado ( $X_{27}$ ), Sinalização Luminosa\_Inexistente ( $X_{29}$ ), Sinalização Luminosa\_Funciona ( $X_{31}$ ), Acostamento\_Sim ( $X_{32}$ ), Acostamento\_Não ( $X_{33}$ ), desnível Acostamento\_Não ( $X_{34}$ ), Acostamento Pavimentado\_Sim ( $X_{35}$ ), Conservação Acostamento\_Bom ( $X_{38}$ ), Canteiro\_Não ( $X_{39}$ ), Obstáculo ao Cruzamento\_Meio-fio ( $X_{40}$ ), Pista\_Simples ( $X_{46}$ ), Pista\_dupla ( $X_{47}$ ), Perfil Pista\_Em Nível ( $X_{49}$ ), Perfil Pista\_Rampa  $< 3\%$  ( $X_{51}$ ), Traçado Pista\_Reta ( $X_{52}$ ), Curva Vertical Pista\_Não ( $X_{54}$ ), Superelevação Pista\_Não ( $X_{55}$ ), Superlargura Pista\_Não ( $X_{56}$ ), Estreitamento Pista\_Não Existe ( $X_{57}$ ) e Conservação Pista\_Bom ( $X_{58}$ ).

**Fator 2:** Tipo de Acidente\_Colisão com Objeto Fixo ( $X_4$ ), Condição Pista\_Molhada ( $X_{12}$ ), Condição Meteorológica\_Chuva ( $X_{16}$ ), Obstáculo Ao Cruzamento\_Muro ( $X_{28}$ ), Traçado Pista\_Curva ( $X_{42}$ ) e Causa do Acidente\_Velocidade Incompatível ( $X_{53}$ ).

**Fator 3:** Canteiro\_Sim( $X_{41}$ ) e Obstáculo Ao Cruzamento\_Sarjeta ( $X_{48}$ ).

**Fator 4:** Desnível Acostamento\_Sim ( $X_{36}$ ).

**Fator 5:** Obstáculo Ao Cruzamento\_Tela ( $X_{44}$ ) e Obstáculo Ao Cruzamento\_Com Anti-ofuscante ( $X_{45}$ ).

**Fator 6:** Sinalização\_Inexistente ( $X_{23}$ ) e Conservação Pista\_Com Erosão ( $X_{59}$ ).

**Fator 7:** Tipo de Acidente\_Atropelamento de Animal ( $X_8$ ) e Causa do Acidente\_Animais Na Pista ( $X_{14}$ ).

**Fator 8:** Restrição Visibilidade\_Poeira/fumaça/neblina ( $X_{19}$ ) e Condição Meteorológica\_Nevoeiro/neblina ( $X_{30}$ ).

**Fator 9:** Causa do Acidente\_defeito Mecânico Em Veículo ( $X_{10}$ ).

**Fator 10:** Restrição Visibilidade\_Veículo Estacionado ( $X_{20}$ ).

A matriz das cargas fatoriais pode ser encontrada na Tabela (1), assim como as communalidades (Com.) e a porcentagem da variância Explicada (Var.). Os fatores são denotados por  $F_i$ ,  $i = 1, 2, \dots, 10$  e as variáveis por  $X_j$ ,  $j = 1, 2, \dots, 59$ . As variáveis com comunalidade abaixo de 0,6 foram eliminadas do modelo.

	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	Com.
$X_1$	<b>0,93</b>	0,04	0,07	-0,03	0,00	0,08	0,04	0,09	0,06	-0,00	0,91
$X_2$	<b>0,77</b>	0,42	0,17	-0,18	0,09	0,12	0,02	0,01	0,01	0,01	0,86
$X_3$	<b>0,88</b>	-0,08	-0,06	-0,05	-0,12	-0,04	0,00	0,06	-0,15	0,04	0,84
$X_4$	0,26	<b>0,66</b>	0,14	0,15	0,18	0,07	-0,05	0,02	0,32	-0,00	0,70
$X_5$	<b>0,85</b>	0,23	-0,06	0,00	0,20	-0,06	0,02	0,07	0,05	-0,03	0,83
$X_6$	<b>0,78</b>	0,05	0,00	0,06	0,00	0,25	-0,01	-0,02	-0,05	-0,05	0,70
$X_7$	0,22	-0,08	0,04	-0,40	-0,03	0,37	-0,07	0,10	-0,11	0,32	0,49*
$X_8$	0,10	-0,04	0,07	-0,03	0,01	0,02	<b>0,96</b>	0,01	-0,00	0,01	0,95
$X_9$	<b>0,96</b>	0,05	0,02	-0,04	-0,04	0,01	0,01	0,02	-0,05	-0,00	0,93
$X_{10}$	0,20	0,12	0,10	-0,08	0,01	-0,01	0,00	0,08	<b>0,82</b>	0,05	0,76
$X_{11}$	<b>0,61</b>	0,33	-0,11	-0,05	0,27	-0,11	-0,01	0,11	-0,05	-0,02	0,60
$X_{12}$	0,16	<b>0,82</b>	0,09	-0,03	0,03	0,13	-0,03	0,02	0,04	0,01	0,74
$X_{13}$	<b>0,77</b>	-0,12	-0,06	0,01	-0,01	0,11	-0,07	0,06	-0,13	0,05	0,66
$X_{14}$	0,09	-0,04	0,04	-0,04	0,01	-0,02	<b>0,95</b>	0,04	0,01	0,03	0,93
$X_{15}$	<b>0,96</b>	0,04	0,11	-0,08	0,05	0,03	0,05	0,04	0,08	0,02	0,97
$X_{16}$	0,52	<b>0,64</b>	0,09	-0,17	0,03	0,19	-0,01	0,10	-0,06	-0,07	0,79
$X_{17}$	0,19	0,16	0,21	0,02	-0,07	0,66	0,02	0,04	0,08	-0,11	0,58*
$X_{18}$	<b>0,94</b>	0,21	0,14	-0,10	0,05	0,11	0,04	0,02	0,04	-0,00	0,99
$X_{19}$	0,19	0,19	-0,04	-0,11	-0,00	-0,05	0,04	<b>0,86</b>	0,07	-0,03	0,84
$X_{20}$	0,01	0,00	0,01	0,02	0,00	-0,03	0,06	-0,04	0,03	<b>0,91</b>	0,84
$X_{21}$	<b>0,94</b>	0,21	0,13	-0,10	0,05	0,01	0,03	0,08	0,03	-0,00	0,98
$X_{22}$	<b>0,53</b>	0,27	0,02	-0,15	0,01	0,42	0,13	-0,06	0,15	-0,02	0,62
$X_{23}$	0,23	0,02	0,07	0,01	0,00	<b>0,77</b>	0,00	-0,03	-0,08	-0,01	0,66
$X_{24}$	<b>0,70</b>	0,21	0,34	-0,15	0,16	0,19	0,13	0,09	-0,01	0,05	0,77
$X_{25}$	<b>0,93</b>	0,21	-0,01	-0,04	-0,00	0,06	-0,03	0,01	0,06	-0,02	0,93
$X_{26}$	<b>0,91</b>	0,00	0,19	-0,13	-0,00	-0,00	0,05	0,06	0,03	0,04	0,90
$X_{27}$	<b>0,83</b>	0,02	-0,02	-0,00	0,04	0,08	-0,00	-0,04	0,16	0,02	0,73
$X_{28}$	0,47	<b>0,62</b>	0,16	-0,17	-0,02	0,20	0,00	0,02	-0,08	-0,06	0,72

$X_{29}$	<b>0,76</b>	0,28	0,03	0,00	0,19	0,17	0,06	-0,05	0,05	-0,03	0,74
$X_{30}$	0,12	0,03	0,03	0,00	0,05	-0,00	0,02	<b>0,92</b>	0,00	-0,01	0,87
$X_{31}$	<b>0,85</b>	0,28	0,18	-0,24	0,05	0,16	0,07	0,08	0,00	0,01	0,96
$X_{32}$	<b>0,74</b>	-0,02	-0,13	0,39	0,01	-0,04	-0,06	-0,01	0,15	-0,03	0,75
$X_{33}$	<b>0,67</b>	0,28	0,27	-0,54	0,12	0,01	0,13	0,07	0,11	-0,01	0,96
$X_{34}$	<b>0,87</b>	0,09	-0,05	0,36	-0,04	0,17	-0,06	0,02	-0,03	0,02	0,95
$X_{35}$	<b>0,92</b>	0,26	0,15	-0,00	0,07	0,14	0,03	0,05	0,05	0,01	0,98
$X_{36}$	0,52	-0,10	-0,08	<b>-0,57</b>	-0,08	-0,13	0,06	0,07	0,01	-0,08	0,66
$X_{37}$	0,66	0,28	0,30	-0,53	0,13	-0,00	0,11	0,06	0,11	-0,02	0,94
$X_{38}$	<b>0,67</b>	0,29	0,27	-0,52	0,12	-0,04	0,12	0,08	0,12	-0,02	0,95
$X_{39}$	<b>0,81</b>	0,10	-0,44	-0,18	-0,09	0,12	0,01	0,10	0,09	-0,03	0,96
$X_{40}$	<b>0,64</b>	0,25	0,67	0,02	0,17	0,05	0,04	-0,00	-0,02	0,03	0,97
$X_{41}$	0,22	-0,11	<b>0,82</b>	-0,21	-0,09	0,04	0,09	0,06	0,14	0,02	0,82
$X_{42}$	0,13	<b>0,86</b>	0,00	0,01	0,01	0,04	-0,02	0,12	0,11	-0,00	0,78
$X_{43}$	0,33	0,11	0,64	-0,08	-0,11	0,01	0,13	-0,08	-0,03	-0,08	0,58*
$X_{44}$	0,12	0,02	0,04	-0,04	<b>0,94</b>	0,01	0,02	-0,01	0,00	0,00	0,90
$X_{45}$	0,10	-0,00	0,01	0,01	<b>0,92</b>	-0,01	0,01	0,06	0,01	-0,00	0,87
$X_{46}$	<b>0,76</b>	0,03	0,23	0,30	-0,09	0,03	-0,07	-0,06	-0,24	0,09	0,82
$X_{47}$	<b>0,78</b>	-0,14	-0,47	-0,22	-0,15	0,00	0,03	0,08	-0,02	-0,03	0,94
$X_{48}$	0,50	0,45	<b>0,64</b>	0,00	0,21	0,16	0,03	0,00	0,04	0,03	0,94
$X_{49}$	<b>0,87</b>	-0,00	0,28	0,04	0,15	0,24	0,09	-0,03	-0,01	-0,02	0,93
$X_{50}$	0,21	0,48	-0,11	-0,43	-0,14	-0,10	-0,01	0,22	0,09	0,02	0,57*
$X_{51}$	<b>0,62</b>	0,42	-0,21	-0,16	-0,15	-0,17	-0,08	0,13	0,13	0,07	0,74
$X_{52}$	<b>0,79</b>	-0,01	0,19	-0,30	0,18	0,23	0,12	0,07	0,21	-0,03	0,91
$X_{53}$	0,34	<b>0,79</b>	0,00	-0,05	-0,10	-0,08	-0,07	0,04	-0,10	0,04	0,78
$X_{54}$	<b>0,92</b>	0,23	0,19	-0,08	0,07	0,11	0,05	0,03	0,04	0,00	0,97
$X_{55}$	<b>0,94</b>	0,23	0,13	-0,10	0,05	0,11	0,04	0,05	0,04	0,00	0,99
$X_{56}$	<b>0,94</b>	0,22	0,13	-0,13	0,05	0,11	0,04	0,06	0,02	0,00	0,99
$X_{57}$	<b>0,93</b>	0,23	0,13	-0,13	0,04	0,10	0,04	0,06	0,04	0,00	0,99
$X_{58}$	<b>0,94</b>	0,22	0,14	-0,12	0,05	-0,02	0,04	0,07	0,05	-0,00	0,99
$X_{59}$	0,35	0,13	-0,05	0,05	0,04	<b>0,79</b>	-0,03	-0,07	0,01	0,02	0,79
Var.(%)	51,08	7,25	5,52	4,74	3,67	3,13	2,63	2,39	1,79	1,73	83,93

Tabela 1: Cargas Fatoriais, Comunalidades(Com.) e Variância Explicada (Var.).

No gráfico (4) relacionando os fatores 1 e 2 observa-se o agrupamento das variáveis descrito na tabela acima, que são os fatores mais significativos.

De acordo com os fatores acima, foram calculados os escores fatoriais, onde pode-se identificar os quilômetros da rodovia com maior número de acidentes relacionados com cada fator, que está organizado na tabela (2).

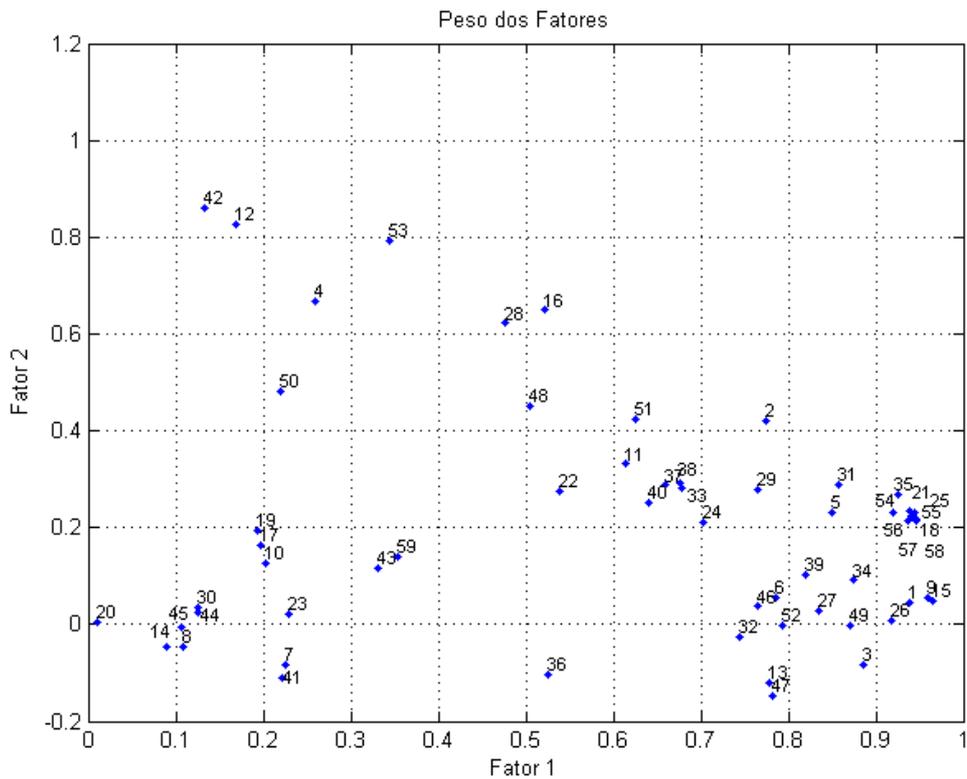


Figura 4: Peso dos fatores: Fator1  $\times$  Fator 2

Fatores	1 <sup>a</sup> Posição	2 <sup>a</sup> Posição	3 <sup>a</sup> Posição
1	584	728	657
2	37	38	33
3	721	723	725
4	584	728	60
5	80	82	79
6	7	4	1
7	702	583	710
8	42	312	649
9	464	74	60
10	727	347	701

Tabela 2: *Ranking* dos quilômetros de acordo com cada fator

## 5 CONCLUSÕES

O objetivo principal da análise fatorial aplicado à matriz de frequências obtida por meio do banco de dados é conseguir fatores que possam agrupar variáveis e também determinar os escores fatoriais, para encontrar os pontos mais críticos com relação ao número de acidentes da BR-277. Interpretando conjuntamente os fatores e os escores fatoriais obtidos na segunda análise fatorial pode-se concluir:

- O primeiro fator, que é o mais significativo, reúne algumas variáveis importantes, tais como três tipos de colisão (transversal, traseira e lateral), ligado a falta de atenção dos motoristas, desobediência à sinalização e também o fato de não guardarem a distância de segurança. Pode-se ainda perceber que as condições meteorológicas estão quase todas presentes nesse fator, exceto chuva e os acidentes ocorrem em pista seca. Quanto a estrutura da pista observa-se as variáveis pista dupla em nível ou inclinada, com traçado reto, e sem superlevação, superlargura, estreitamento, canteiro ou acostamento. Analisando os escores fatoriais desse fator e selecionando apenas os três primeiros (mais expressivos), consegue-se identificar os quilômetros KM 584 (próximo a Cascavel - PR), KM 728 (Foz do Iguaçu - PR) e KM 657 (próximo a Medianeira - PR) que estão intimamente ligados com as características desse fator.
- O segundo fator é bem interessante, pois observa-se a junção das variáveis velocidade incompatível com pista curva, pista molhada e com chuva, além de colisão com objeto fixo e também obstáculo ao cruzamento com muro. Pelos escores fatoriais, consegue-se identificar os três quilômetros onde ocorrem acidentes com essa característica: KM 37, KM 38 e KM 33 (Entre Paranaguá - PR e Curitiba - PR), o que indica que maioria dos acidentes com esse perfil ocorrem nesses quilômetros da BR-277.
- O terceiro fator agrupa as variáveis relacionadas a canteiro e obstáculo ao cruzamento com sarjeta (valas geralmente localizadas ao longo das vias pavimentadas, para onde escoam as águas da chuva). Acidentes com essas características se concentram principalmente nos quilômetros KM 721, KM 723 e KM 725 (próximo a Foz do Iguaçu - PR).
- No quarto fator temos uma única variável referente ao acostamento em desnível. Os acidentes que ocorrem na rodovia com essa característica estão concentrados principalmente nos quilômetros KM 584 (Cascavel - PR), KM 728 (Foz do Iguaçu - PR) e KM 60 (próximo a Curitiba - PR).
- No quinto fator temos duas variáveis que se destacam: obstáculo ao cruzamento com tela e com anti-ofuscante. Acidentes envolvendo as com esses tipos de obstáculos ao cruzamento ocorrem principalmente nos quilômetros KM 80, KM 82 e KM 79 (próximo a Curitiba - PR).
- O sexto fator é composto por duas variáveis que se destacam: pista com erosão e sinalização inexistente. Pelos escores fatoriais podemos observar que, bem no início da rodovia, nos quilômetros KM 7, KM 4 e KM 1 (Paranaguá - PR) ocorrem acidentes devido a essa característica da pista.
- O sétimo fator reúne as variáveis atropelamento de animais e animais na pista. Nesse fator os escores fatoriais apontam os quilômetros KM 702, KM 710 (entre Medianeira - PR e Foz do Iguaçu - PR) e KM 583 (Cascavel - PR) como sendo os de maior ocorrência deste tipo de acidente.
- O oitavo fator envolve acidentes ligados às condições meteorológicas (nevoeiro e neblina) e restrição de visibilidade (fumaça, poeira e neblina). Acidentes com esse perfil ocorrem principalmente nos quilômetros KM 42 (entre Paranaguá - PR e Curitiba), KM 312 (entre Irati - PR e Guarapuava - PR) e KM 649 (próximo a Medianeira - PR).

- O nono fator envolve acidentes devido a falta de defeitos mecânicos em veículos. Acidentes com esse perfil ocorrem principalmente nos quilômetros KM 74, KM 60 (entre Paranaguá - PR e Curitiba) e KM 464 (próximo a Nova Laranjeiras - PR).
- Finalmente o décimo fator, menos representativo, está ligado à acidentes devido a restrição de visibilidade causada por veículos estacionados. Mesmo assim podemos identificar os quilômetros KM 727 (Guaraniaçu - PR), KM 727, KM 701 (próximo a Foz do Iguaçu - PR) e KM 347 (próximo a Guarapuava - PR) ligados a esse fator.

Os 10 fatores acima dão uma visão mais clara dos acidentes que ocorreram na BR-277. Como pode ser observado, no histograma da Figura (3) ocorre uma concentração muito grande de acidentes no início e no final da rodovia. Os fatores e escores fatoriais indicaram em sua grande maioria a mesma tendência.

Esses resultados podem ser utilizados para que sejam feitos planejamentos futuros para a rodovia, na tentativa de evitar o grande número de acidentes que vem ocorrendo, principalmente em meses de férias em que muitos viajam, e também os custos desses acidentes, principalmente na área da saúde.

## REFERÊNCIAS

- Bartilotti C.B. *Fatores da Senso-Percepção Relacionados à Atividade do Condutor no Sistema Trânsito: Construção e Validação de um Instrumento de Medida*. Tesis de Doutorado, UFSC, 2009.
- Hair J.F. e Anderson R.E. *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. Factorial analysis in selecting the zeolite catalyst for 2,6-dimethylnaphthalene synthesis. *Coke and Chemistry*, 50:200–206, 2007.
- Pires T. e Maia A. Relação entre saúde física e transtorno de estress 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*. Tesis de Mestrado, FFCLRP, 2004.
- Spearman C. General intelligence objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904.