

UNIVERSIDADE DE SÃO PAULO  
Escola de Engenharia de São Carlos  
Departamento de Engenharia de Transportes  
Programa de Pós-Graduação em Engenharia de Transportes

**DIEGO CAMARGO**

**ANÁLISE DO COMPORTAMENTO DE CONDUTORES DE  
TRANSPORTE PÚBLICO E A RELAÇÃO COM ACIDENTES DE  
TRÂNSITO – ESTUDO DE CASO NA CIDADE DE RIBEIRÃO PRETO**

São Carlos

2016



**DIEGO CAMARGO**

**ANÁLISE DO COMPORTAMENTO DE CONDUTORES DE  
TRANSPORTE PÚBLICO E A RELAÇÃO COM ACIDENTES DE  
TRÂNSITO – ESTUDO DE CASO NA CIDADE DE RIBEIRÃO PRETO**

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte integrante dos requisitos para obtenção do título de Mestre em Engenharia de Transportes.

Área de Concentração: Infraestrutura de Transportes.

Orientador: Prof. Dr. Paulo Cesar Lima Segantine

SÃO CARLOS

Junho 2016

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTA TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Cl72a. Camargo, Diego  
ANÁLISE DO COMPORTAMENTO DE CONDUTORES DE  
TRANSPORTE PÚBLICO E A RELAÇÃO COM ACIDENTES DE  
TRÂNSITO □ ESTUDO DE CASO NA CIDADE DE RIBEIRÃO PRETO /  
Diego Camargo; orientador Paulo Cesar Lima Segantine,  
São Carlos, 2016.

Dissertação (Mestrado) - Programa de Pós-Graduação  
em Engenharia de Transportes e Área de Concentração em  
Infra-Estrutura de Transporte -- Escola de Engenharia  
de São Carlos da Universidade de São Paulo, 2016.

1. Transporte Público Urbano. 2. Comportamento dos  
condutores de transporte público. 3. Acidentes de  
trânsito. I. Título.

## FOLHA DE JULGAMENTO

Candidato: Engenheiro **DIEGO CAMARGO**.

Título da dissertação: "Análise do comportamento de condutores de transporte público e a relação com acidentes de trânsito – estudo de caso na cidade de Ribeirão Preto".

Data da defesa: 30/09/2016.

### Comissão Julgadora:

Prof. Associado **Paulo Cesar Lima Segantine**  
**(Orientador)**  
(Escola de Engenharia de São Carlos/EESC)

Profa. Dra. **Anabela dos Santos Aleixo Simões**  
(Universidade de Lisboa/Portugal)

Profa. Dra. **Bárbara Stolte Bezerra**  
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP - Bauru)

### Resultado:

Aprovado

Aprovado

Aprovado

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:  
Profa. Dra. **Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:  
Prof. Associado **Luis Fernando Costa Alberto**



*Dedico este trabalho ao meu pai.*



## AGRADECIMENTOS

---

Agradeço a toda a minha família pela compreensão e apoio, em especial meus primos Miguel e Fabiane.

Ao professor Segantine, pela orientação.

À professora Anabela Simões, por toda a ajuda prestada a este trabalho.

Aos professores do Departamento de Transportes, em especial Ana Larocca e Cira que ajudaram neste trabalho.

Aos meus amigos do mestrado, que transformaram essa experiência ainda mais valiosa.

À *CAPES* (Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio Financeiro.



---

Camargo, D. **Análise do comportamento de condutores de transporte público e a relação com acidentes de trânsito – estudo de caso na cidade de Ribeirão Preto.** 97p. Dissertação de Mestrado – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2016.

Este trabalho apresenta um estudo a respeito dos acidentes de trânsito envolvendo o transporte público urbano por ônibus e objetiva, principalmente, a relação dos acidentes *versus* comportamento dos condutores. Os dados utilizados, a partir do estudo de caso realizado na cidade de Ribeirão Preto-SP, têm duas origens: dados da operação do sistema (quantidade de horas operadas, quilometragem e frota) e dados gerados pelo monitoramento por câmeras. Este último tem como principais variáveis o comportamento dos condutores durante a condução dos veículos. Através de índices de exposição, utilizando as variáveis da operação do sistema, juntamente com os acidentes por linha, foi possível identificar quais as linhas com piores indicadores, ou seja, quais linhas merecem maior atenção na criação de intervenções ou campanhas para redução do número de acidentes. Foram tratados aproximadamente 72 mil dados e a partir dos dados extraídos e processados estatisticamente para obtenção das variáveis mais significativas com relação aos acidentes. A variável com maior peso foi a utilização de telefones celulares durante a condução e que tem alta utilização nos horários de pico, da ordem de 53% dos eventos ocorrem em horários de maior movimento de passageiros e de tráfego intenso. O tempo de utilização do celular durante a condução do ônibus é majoritariamente maior que 5 minutos, ou seja, 33% dos eventos mostram que os condutores utilizam o telefone celular por mais de 5 minutos. Criou-se uma taxonomia do comportamento dos condutores, baseando-se, principalmente, no banco de dados e tem como função instituir uma base teórica dos comportamentos, ajudando a descrevê-los e entendê-los. É dessa maneira que a segunda variável foi discutida. O avanço do sinal amarelo, com nível de significância alta, não representa em sua totalidade um comportamento decidido (decisões conscientes do condutor), mas algumas vezes comportamento involuntário (falhas e lapsos). Essa distinção de comportamento decidido ou involuntário é complexa, mas sabemos que decisões conscientes são mais frequentes. Este trabalho identificou quais as linhas que necessitam de intervenções e quais os problemas com o comportamento dos condutores, direcionando o operador do sistema de transporte às campanhas necessárias para redução dos

acidentes, ou mesmo possibilitando outras empresas a replicarem as análises para a sua realidade operacional.

Palavras-chave: Transporte Público, comportamento dos condutores de transporte público, acidentes de trânsito.

ABSTRACT

---

Camargo, D. **Analysis of bus drivers behavior and its relation with the traffic accidents.**  
- Case study in Ribeirão Preto city. 97p

This work presents a study about the traffic accidents involving urban public transport by bus and objective, especially the relation of accidents versus driver behavior. The data used from the case study in the city of Ribeirão Preto, have two sources: System operation data (number of operated hours, mileage and fleet) and data generated by the monitoring camera system. The latter's main variables driver behavior while driving the vehicle. Through levels of exposure, using the system operating variables, along with accidents per line, it was possible to identify lines with worse indicators, or which lines deserve close attention in setting up operations or campaigns to reduce the number of accidents. Approximately 72,000 data were treated and statistically processed to obtain the most significant variables in relation to accidents. The most significant variable was the use of mobile phones while driving and which has high utilization during peak hours, the order of 53% of events occur when there is a large number of passengers and traffic jam. The utilization of the cellphone while driving the bus is overwhelmingly greater than 5 minutes, i.e., 33% of the events showed that drivers use the mobile phone for more than 5 minutes. Has been created drivers behavior taxonomy, based mainly in the database and with aim to establish a theoretical basis of bus drivers behavior, helping to describe and understand them. This is how the second significant variable was discussed. The advance of the yellow sign is not totally a decided behavior (conscious decisions of the driver), but sometimes involuntary behavior (failures and lapses). This decided or involuntary behavior distinction is complex, but we know that conscious decisions are more frequent. This work identified which lines needed intervention and what are the problems with the behavior of drivers, orienting the operator of the transportation system to needed campaigns to reduce accidents, or even allowing other companies to replicate the analysis to their operational reality.

**Keywords:** Public transport, bus driver behavior, traffic accidents.



---

**SUMÁRIO**


---

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>19</b>
1.1	Caracterização do problema .....	19
1.2	Justificativa.....	20
1.3	Objetivos .....	20
1.4	Desenvolvimento do trabalho.....	21
1.5	Estrutura do trabalho.....	21
<b>2</b>	<b>REVISÃO DE LITERATURA .....</b>	<b>23</b>
2.1	Acidentes com ônibus urbano .....	23
2.2	Comportamento dos condutores.....	24
2.3	Situação atual dos estudos a respeito do condutor de transporte público .....	25
2.4	Naturalistic Driving Study (NDS) .....	27
2.5	Aplicações de sistemas de monitoramento por câmeras no transporte público urbano 28	
2.6	Taxonomias .....	30
<b>3</b>	<b>MÉTODOS E DESENVOLVIMENTO.....</b>	<b>33</b>
3.1	Índices de Acidentes .....	33
3.1.1	Índice de acidente por quilômetro percorrido .....	33
3.1.2	Índice de acidente por tempo de operação .....	33
3.1.3	Índice de acidente por número de passageiros transportados .....	34
3.2	Distribuição estatística dos dados.....	34
3.2.1	Distribuição de Poisson (DP) .....	35
3.2.2	Distribuição Binomial Negativa (DBN) .....	35
3.3	Modelos Inflacionados de Zeros.....	36
3.4	Critério de Informação de Akaike - AIC.....	37
<b>4</b>	<b>CARACTERÍSTICAS DO TRANSPORTE PÚBLICO E DO BANCO DE DADOS – ESTUDO DE CASO.....</b>	<b>39</b>
4.1	Características do sistema de transporte público.....	39
4.2	Apresentação dos dados brutos do banco de dados e a adaptação ao estudo.....	41

4.3	Resumo dos dados utilizados.....	42
4.4	Dados extraídos <i>versus</i> dados criados a partir do banco de dados .....	42
4.4.1	Quantidade de condutores envolvidos na operação de uma linha .....	42
4.4.2	Conversa durante a condução .....	43
4.4.3	Dados das linhas de operação.....	43
4.4.4	Dificuldades com o banco de dados .....	44
<b>5</b>	<b>ANÁLISE E RESULTADOS .....</b>	<b>45</b>
5.1	Análise exploratória .....	45
5.1.1	Índices de acidentes .....	45
5.1.2	Acidentes .....	48
5.1.3	Demais ocorrências.....	54
5.2	Taxonomia – Baseada nos dados.....	58
5.2.1	Do comportamento dos condutores .....	58
5.3	Resultados estatísticos dos dados .....	60
5.3.1	Distribuição dos dados.....	61
5.3.2	Processamento estatístico .....	62
5.4	Discussão e análise dos resultados.....	64
5.5	Sugestões.....	67
5.5.1	À empresa.....	68
5.5.2	Planilha eletrônica para controle/cadastro dos acidentes .....	71
5.5.3	Ao estado.....	74
5.5.4	Trabalhos futuros.....	74
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>77</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRAFICAS .....</b>	<b>79</b>
<b>A.</b>	<b>APÊNDICE – INFORMAÇÕES DAS LINHAS DE OPERAÇÃO .....</b>	<b>81</b>
<b>B.</b>	<b>APÊNDICE – RESULTADOS DO PROCESSAMENTO ESTATÍSTICO .....</b>	<b>83</b>

<b>C. APÊNDICE – RELATÓRIO DE PADRONIZAÇÃO DE ATUALIZAÇÃO E PADRONIZAÇÃO DOS DADOS DE IMAGENS.....</b>	<b>85</b>
--	-----------



## LISTA DE FIGURAS

---

<b>Figura 1</b> - Estrutura do desenvolvimento dos dados deste trabalho. ....	21
<b>Figura 2</b> - Taxonomia dos atos inseguros (Reason, 1990 - Adaptado por Simões 2013).....	25
<b>Figura 3</b> - As imagens das quatro câmeras instaladas no interior do veículo, para o projeto NDS.....	28
<b>Figura 4</b> - Mapa das agências que fizeram parte do questionário em 2015 .....	29
<b>Figura 5</b> - Respostas para as Motivações/Propostas de implantação do sistema de monitoramento.....	30
<b>Figura 4</b> - Imagens geradas pelas câmeras de monitoramento.....	39
<b>Figura 5</b> - Organização das regiões do município de Ribeirão Preto .....	40
<b>Figura 8</b> - Criação de novo dado a partir de dados existentes.....	42
<b>Figura 9</b> - Eventos que compõem a criação de um novo dado.....	43
<b>Figura 10</b> - Taxonomia do comportamento do condutor .....	59
<b>Figura 11</b> - Taxonomia dos acidentes, quanto à gravidade. ....	70
<b>Figura 12</b> - Taxonomia dos acidentes, quanto à característica do local.....	70
<b>Figura 13</b> - Planilha eletrônica para cadastro dos acidentes ocorridos .....	71
<b>Figura 14</b> - Planilha eletrônica para cadastro dos acidentes ocorridos, exemplo .....	72
<b>Figura 15</b> - Relatório com exemplos dos tipos de acidentes de trânsito .....	72
<b>Figura 16</b> - Relatório com exemplos dos tipos de acidentes de trânsito, continuação .....	73



## LISTA DE GRÁFICOS

---

<b>Gráfico 1</b> - Gráfico de dispersão apresentando número de acidentes pela quilometragem média por mês, por linha. ....	48
<b>Gráfico 2</b> - Gráfico de dispersão apresentando número de acidentes pela quantidade média de horas operadas por mês, por linha. ....	49
<b>Gráfico 3</b> - Gráfico de dispersão apresentando número de acidentes pela quantidade média de passageiros por mês, por linha. ....	49
<b>Gráfico 4</b> - Frequência de acidentes por faixa horária .....	49
<b>Gráfico 5</b> - Número de acidentes em linhas Radiais .....	50
<b>Gráfico 6</b> - Número de acidentes em linhas Radiais .....	50
<b>Gráfico 7</b> - Número de acidentes em linhas Diametrais.....	51
<b>Gráfico 8</b> Número de acidentes em linhas Circulares.....	51
<b>Gráfico 9</b> - Número de acidentes em linhas Perimetrais .....	52
<b>Gráfico 10</b> - Número de acidentes em linhas Noturnas .....	52
<b>Gráfico 11</b> - Número de acidentes por faixa horária, linhas perimetrais .....	53
<b>Gráfico 12</b> - Número de acidentes por faixa horária, linhas noturnas .....	53
<b>Gráfico 13</b> - Número de acidentes por faixa horária, linhas diametrais.....	53
<b>Gráfico 14</b> - Número de acidentes por faixa horária, linhas circulares.....	54
<b>Gráfico 15</b> - Número de acidentes por faixa horária, linhas radiais .....	54
<b>Gráfico 16</b> - Número de eventos por faixa horária, avanço de sinal e comendo durante a condução.....	55
<b>Gráfico 17</b> - Número de eventos por faixa horária, assalto, danos ao veículo e condutor utilizando fone .....	55
<b>Gráfico 18</b> - Número de eventos por faixa horária, brigas, condutor fuma ou discute com passageiro .....	56
<b>Gráfico 19</b> - Número de eventos por faixa horária, conversa durante a condução e utilização de celular .....	56
<b>Gráfico 20</b> - Percentual de ocorrência de acidentes e de demanda de passageiros, por faixa horária .....	65



## LISTA DE TABELAS

---

<b>Tabela 1</b> - Média dos eventos .....	45
<b>Tabela 2</b> - Dados das linhas classificados pelo $I_k$ (Índice de acidentes por quilômetro) .....	46
<b>Tabela 3</b> - Dados das linhas classificados pelo $I_h$ (Índice de acidentes por hora operada) .....	47
<b>Tabela 4</b> - Dados das linhas classificados pelo $I_p$ (Índice de acidentes por passageiros) .....	47
<b>Tabela 5</b> - Tempo de uso do celular .....	57
<b>Tabela 6</b> - Período de maior utilização do celular .....	57
<b>Tabela 7</b> - Tempo de conversa durante condução.....	57
<b>Tabela 8</b> - Período de maior frequência de conversas .....	57
<b>Tabela 9</b> - Valores do AIC para as diferentes distribuições .....	61
<b>Tabela 10</b> - Resultado da análise estatística, utilizando variáveis de comportamento .....	62
<b>Tabela 11</b> - Resultado da análise estatística, inserindo o número de condutores.....	63
<b>Tabela 12</b> - Resultado da análise estatística, inserindo apenas o número de condutores.....	63
<b>Tabela 13</b> - Resultado da análise estatística, para agrupamento por Região.....	63
<b>Tabela 14</b> - Resultado da análise estatística, para agrupamento por Condutores.....	63
<b>Tabela 15</b> - Resumo dos resultados processados.....	63
<b>Tabela 16</b> - Média de acidentes e passageiros por tipo de linha.....	66
<b>Tabela 17</b> - Informações das linhas do sistema de transporte público .....	81



## 1 INTRODUÇÃO

---

Este capítulo introduz o problema, ao qual este trabalho foi desenvolvido, e caracteriza o atual panorama de acidentes no Brasil. É apresentada também a justificativa para a realização deste estudo e os objetivos deste trabalho.

### 1.1 Caracterização do problema

Acidentes e/ou incidentes envolvendo ônibus urbano têm sido, recentemente, objeto de alguns estudos, mesmo considerado o meio de transporte mais seguro, do ponto de vista de seus usuários, comparado aos outros modos motorizados (CHIMBA *et al.*, 2010).

O ônibus sendo um dos únicos veículos pesados a transitar livremente pelo meio urbano (CHIMBA *et al.*, 2010), são causadores de acidentes perigosos, quando colidem com um pedestre, bem como indiretamente proporcionam outros tipos de acidentes devido ao seu tamanho, uma vez que por vezes o ônibus não permite a visibilidade clara para os veículos que andam atrás (BRENAC *et al.*, 2005).

Os condutores de ônibus são mais expostos aos acidentes pela quilometragem diária que percorrem, sem contar aspectos psicológicos do exercício da profissão, que podem aumentar as chances de ocorrências de acidentes. Acidentes com ônibus urbano são responsáveis por 13% do custo total de acidentes no país, contando apenas com 1% da frota total (englobada no universo do estudo), sendo, portanto, acidentes com alto custo por veículo (IPEA, 2003). Os acidentes fatais, contudo, na União Europeia, envolvendo ônibus, na última década são baixos e representam 0,36% do número dos acidentes ocorridos. (EUROPEAN UNION, 2012)

No cenário nacional, o Instituto de Pesquisas e Economia Aplicada (IPEA), lançou em 2003 um relatório quantificando, monetariamente, os acidentes em áreas urbanas. No entanto, no que diz respeito ao modal ônibus, houve apenas uma extrapolação das análises realizadas na região da cidade de São Paulo, sendo necessário maior número de estudos e maior quantidade de dados a respeito de acidentes ocorridos com o transporte público.

Segundo o IPEA (2003), “*O projeto dedicou um esforço para enquadrar também, ainda que em termos de envoltórias – ou ordem de grandeza – os custos associados a acidentes urbanos envolvendo ônibus e caminhões*”. Este documento ainda relata:

*“Com relação às parcelas de custo representadas por ônibus e caminhão, há que se levar em conta as limitações metodológicas das pesquisas com esses veículos, realizadas apenas na aglomeração urbana de São Paulo. Todavia, os resultados obtidos mostram ser necessário um enfoque específico para os acidentes envolvendo ônibus e caminhões, dada a maior exposição ao risco dessas categorias de veículos (maior quilometragem rodada por ano) e os custos envolvidos, justificando-se a necessidade do aprofundamento de pesquisas relativas aos acidentes de trânsito envolvendo essas categorias. ”.*

Outros trabalhos, com foco em transporte público, apontam que o erro humano está associado a 74% dos acidentes de ônibus (PEARCE *et al.*, 1999).

## **1.2 Justificativa**

Considerando as afirmações anteriormente apresentadas, justifica-se a realização deste estudo pelos seguintes motivos:

- 1) Pelo caráter do inventário dos acidentes, uma vez que faltam dados de acidentes relacionados com ônibus urbano;
- 2) Conhecimento do comportamento dos condutores de ônibus urbano na condução dos veículos e qual a sua relação com a ocorrência de acidentes;
- 3) Alto custo associado aos acidentes envolvendo ônibus urbano.

## **1.3 Objetivos**

### Objetivo Geral:

Esta pesquisa tem como objetivo analisar os dados do comportamento dos condutores e da operação das linhas, em detrimento as ocorrências de acidentes no transporte público urbano, assim como instituir uma base teórica a respeito do comportamento dos condutores, baseado nos dados, e propor melhorias para o sistema de transporte público com visão à segurança viária.

Para alcançar o objetivo geral são definidos os seguintes objetivos específicos.

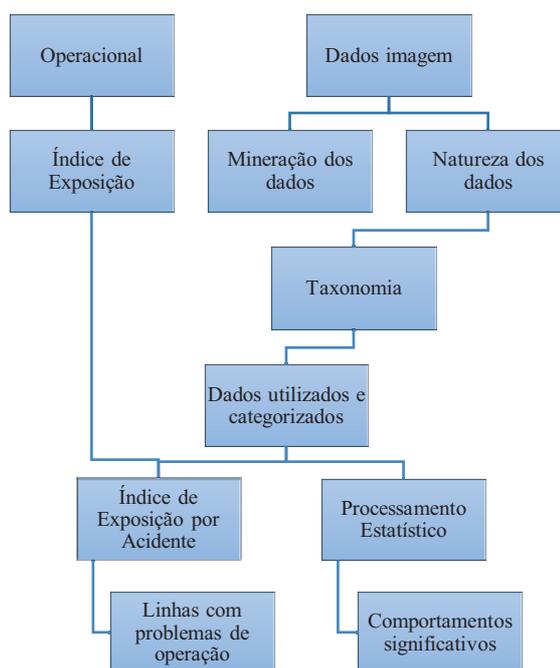
### Objetivos específicos:

- (a) Inferência sobre os dados existentes do banco de dados:

- (b) Elaboração da taxonomia do comportamento;
- (c) Análise estatística dos Acidentes *versus* Comportamento, utilizando o banco de dados disponível.

#### 1.4 Desenvolvimento do trabalho

O trabalho inicialmente contou com aproximadamente 76 mil dados, provenientes do tratamento de imagens, e dados operacionais organizados por linha de operação. A Figura 1 apresenta um resumo simplificado do fluxo e a ordem em que os dados foram trabalhados.



**Figura 1** - Estrutura do desenvolvimento dos dados deste trabalho.

No topo da Figura 1, foram inseridos os dados de imagem e os dados de operação e na sequência os processos que os dados serão submetidos para se obter o resultado final.

#### 1.5 Estrutura do trabalho

Este trabalho está dividido em 7 capítulos, incluindo este capítulo introdutório que mostra as características gerais do trabalho como os objetivos, justificativas e desenvolvimento.

O Capítulo 2 apresenta uma Revisão de Literatura a cerca do tema do tema deste trabalho, destacando os mais relevantes, embora, não exista uma quantidade expressiva de trabalhos técnico-científicos publicados abordando especificamente o tema de acidentes envolvendo transporte público urbano.

No Capítulo 3 são apresentados os métodos necessários para a realização deste trabalho e que pode ser resumido, basicamente, nos cálculos estatísticos e na formulação de equações para os cálculos dos índices.

O Capítulo 4 apresenta as características do estudo de caso analisado neste trabalho, assim como descreve a maneira como os dados foram coletados. Desenvolve-se, também, resumidamente a mineração dos dados de imagem.

Após o tratamento inicial dos dados (mineração – separação dos dados que serão utilizados), o Capítulo 5 inicia com a criação da taxonomia dos dados, criando uma padronização e, mais uma vez, preparando-os para o processamento estatístico. Logo, foi desenvolvida uma análise exploratória dos dados, com o objetivo de extrair alguma informação antes do processamento estatístico. Ainda neste capítulo temos os resultados estatísticos para os dados de comportamento dos condutores e logo após os resultados dos índices de exposição.

O Capítulo 6 faz um resumo dos resultados analíticos do trabalho, assim como, apresenta as principais inferências realizadas ao longo do mesmo. Neste Capítulo, também, são apresentadas as sugestões em todos os níveis ao qual o transporte público urbano está inserido, desde o poder público até o operacional. As sugestões têm como objetivo a redução do número de acidentes ou a redução de exposição dos condutores aos acidentes.

Finalmente, o Capítulo 7 apresenta as Referências Bibliográficas consultadas como base teórica para a realização deste trabalho.

## 2 REVISÃO DE LITERATURA

---

Este capítulo tem como principal objetivo contextualizar o problema de acidentes de trânsito e apresentar uma breve revisão da literatura disponível, no que diz respeito à investigação do comportamento dos condutores como causa de acidentes.

### 2.1 Acidentes com ônibus urbano

Segundo estudo realizado pelo IPEA (2003), a estimativa do custo anual dos acidentes em aglomerações urbanas é da ordem de R\$ 5,3 bilhões, valor que representa 0,4% do PIB do Brasil.

Embora ocorram em menor frequência, os acidentes envolvendo ônibus urbano são responsáveis por 13% do custo total estimado pelo IPEA, estando abaixo dos acidentes envolvendo automóveis e motocicletas. No entanto, o percentual de veículos é de apenas 1% da frota total apresentada no estudo do IPEA. Temos, portanto, um alto custo por veículo, quando se trata de acidentes com ônibus urbano (IPEA, 2003).

O estudo apresentado por Chimba *et al.*(2010), a respeito dos efeitos do tamanho do ônibus e da operação para a ocorrência de acidentes de trânsito, tem como um de seus resultados que pistas mais largas e medianas reduzem a probabilidade de acidentes que envolvem os ônibus, no entanto, mais faixas e maior volume de veículos por faixa aumenta a probabilidade de ocorrência acidentes.

No caso do Brasil, onde boa parte das vias urbanas são estreitas e, atualmente, com alto volume de tráfego de veículos, significa, levando em consideração os resultados de Chimba *et al* (2010), que a probabilidade de acidentes envolvendo ônibus são ainda maiores.

Independentemente dos resultados mostrados por Chimba *et al* (2010), que diz respeito principalmente às características do sistema viário por onde os ônibus transitam, Pearce *et al.* (1999) apresenta em seu estudo que 74% dos acidentes com ônibus urbanos têm como consequência o erro humano, seguido de 18% por fatores externos e 8% pelas condições do veículo, reforçando os números apresentados pelo Departamento de Transportes dos Estados Unidos (1979), que identificou, aproximadamente, 90% dos acidentes têm como prováveis causas fatores humanos.

Pearce *et al.* (1999) desenvolveu seu trabalho com foco nos acidentes envolvendo ônibus urbanos na Índia, no Nepal, na Tanzânia e no Zimbábue.

Embora os percentuais dos trabalhos de Pearce *et al.* (1999) e do Departamento de Transportes dos Estados Unidos (1979) sejam diferentes, o que é justificável pelo público, ano e país de aplicação dos estudos, é interessante concluir que o erro humano é o fator majoritariamente causador de acidentes.

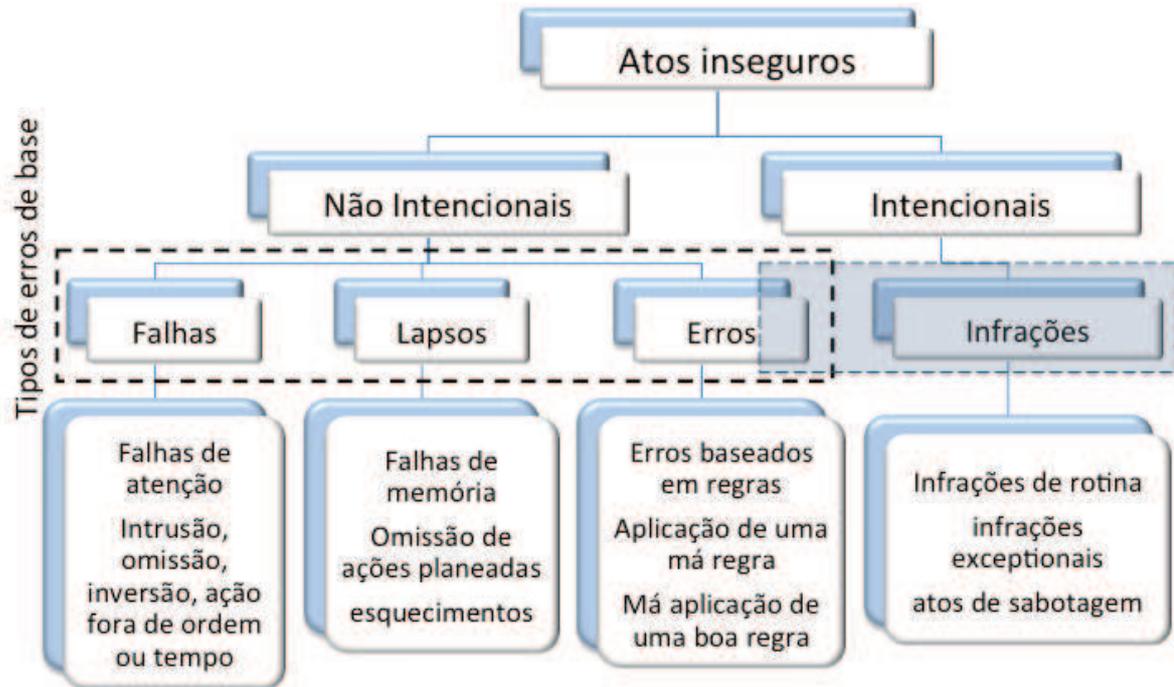
Segundo Ferraz *et al.* (2012), a respeito da exposição ao risco de acidentes, a probabilidade de ocorrência dos acidentes não depende apenas da quantidade total de exposição. Significa, portanto, que apesar de haver correlação entre a quantidade de exposição e o número de acidentes essa relação é complexa, pois depende de outros aspectos como, por exemplo, o comportamento dos condutores. Portanto, o conhecimento sobre o comportamento dos condutores é imprescindível ao entendimento das acidentalidades e suas eventuais soluções.

## **2.2 Comportamento dos condutores**

Alguns estudos têm como principal objetivo o comportamento de motoristas durante a condução de seus veículos. O Centro de Pesquisa de Segurança nas Estradas (*Highway Safety Research Center – HSRC*), da Universidade da Carolina do Norte (2003), desenvolveu um estudo apresentado como “Distrações na condução diária”. Neste documento é abordada uma ampla quantidade de distrações possíveis de ocorrer durante a condução de um veículo no dia-a-dia, desde o uso de celular até a presença de crianças recém-nascidas no banco traseiro dos automóveis.

Foram analisados os efeitos dessas distrações sob a condução do veículo, que teve como resultado, por exemplo, o ato de discar ou responder uma chamada no celular, compromete quase 70% do desempenho do condutor na direção.

James Reason (1990) desenvolveu um resumo das variedades dos atos inseguros, adaptado por Simões (2014), conforme Figura 2. Reason apresenta 4 tipos de erros base: Falhas, Lapsos, Erros e Infrações. A definição de cada um deles é descrita na Figura 2.



**Figura 2** - Taxonomia dos atos inseguros (Reason, 1990 - Adaptado por Simões 2013)

### 2.3 Situação atual dos estudos a respeito do condutor de transporte público

Os fatores de estresse apresentados por Tse *et al.* (2006) são: violência, congestionamentos, pressão para cumprir os horários, padrões de turnos, etc.. E como resultado os condutores podem apresentar; doenças cardíacas, fadiga, ansiedade, entre outros.

Alguns fatores estão diretamente relacionados, como por exemplo, congestionamento e exigências da empresa em atender os horários programados. Na medida em que o congestionamento aumenta, a responsabilidade em atender aos horários também aumenta e como consequência provoca o baixo desempenho dos profissionais (estresse, fadiga, etc.).

Wählberg (1997) realizou um estudo com condutores de ônibus urbano, que abordou a relação entre acidentes e ao atendimento do tempo de viagem da linha (operar no tempo determinado pelo planejamento da empresa), idade do condutor e a velocidade do veículo na condução. Este estudo, apontou que não houve relação significativa entre as variáveis estudadas e o risco de acidentes.

É possível apresentar alguns argumentos que justificam os estudos a respeito do comportamento do condutor de ônibus urbano: (WÄHLBERG, 2004).

1. Estudos sobre acidentes são geralmente restritos a carros, e não a veículos pesados (ônibus);

2. Há grande dificuldade em encontrar estudos sobre possíveis semelhanças e diferenças entre a maneira de conduzir carros e veículos pesados e seus respectivos acidentes;
3. As possibilidades de obter dados sobre os condutores de ônibus urbanos são maiores que um condutor de automóvel. No caso deste trabalho, há várias informações sobre o comportamento do condutor de ônibus.

Mallia *et al.* (2015) através de questionário respondido por condutores de ônibus concluiu que apenas violações (ou infrações) são capazes de prever o número de acidentes, enquanto lapsos e erros (definidos pela taxonomia de Reason) não estão associados com risco de acidentes. O questionário foi utilizado para conhecer os traços de personalidade dos condutores de ônibus e alguns fatos foram encontrados estatisticamente: Ansiedade tem correlação positiva com hostilidade e, também, com o que chamam de “*normlessness*” (ausência de regras) (MALLIA, LAZURAS, *et al.*, 2015). Significa, portanto, que o condutor inserido num ambiente de estresse e perturbação, pode demonstrar algumas das características apresentadas, comprometendo a condução do veículo e o atendimento aos passageiros.

Existe, também, a importância do conhecimento do comportamento dos condutores de ônibus urbanos, pelo simples fato de oferecer um serviço de melhor qualidade aos usuários do transporte público, que de acordo com o estudo de Ramos *et al.* (2011), o usuário tem sensibilidade quanto a maneira com que o veículo é conduzido. Portanto, a qualidade do serviço prestado está diretamente relacionada com o comportamento do condutor.

Os trabalhos relacionados ao comportamento dos condutores profissionais, de maneira geral, se limitam a questionários ou a dados operacionais disponíveis, através de tacógrafos, por exemplo, ou através de equipamentos específicos como utilizado por Wählberg (2004). Com a evolução dos equipamentos, atualmente é possível monitorar o comportamento exercido de fato pelo condutor, a exemplo o trabalho realizado pela Universidade da Carolina do Norte (2003), ou mesmo os trabalhos realizados através do ‘*Naturalistic Driving Study*’ (Estudo Naturalístico de Direção). Estes trabalhos têm como principal meio de obtenção dos dados o monitoramento, principalmente, através de câmeras no interior do veículo que capta o comportamento dos condutores durante a condução de um veículo, entretanto, são trabalhos realizados com condutores comuns, não profissionais.

Até o momento não foi possível encontrar um trabalho utilizando esse tipo de técnica para analisar o comportamento dos condutores profissionais.

#### **2.4 Naturalistic Driving Study (NDS)**

O programa chamado ‘*The Strategic Highway Research Program 2 – SHRP 2*’, ou O Programa Estratégico de Pesquisa Rodoviária 2, tem como principais desafios melhorar: segurança viária, redução de congestionamentos e melhoria dos métodos para renovação de estradas e pontes. No que diz respeito à segurança viária, um dos projetos suportados por essa área é o Projeto S31, que prevê suporte técnico aos usuários do NDS (*Naturalistic Driving Study*).

O NDS envolve o entendimento de como o condutor interage e se adapta com o veículo, o ambiente do tráfego, as características das vias, os sistemas de controle de tráfego e outras características do ambiente. Este trabalho envolve a aquisição de dados através de um sistema batizado por ‘*Data Acquisition System*’ (DAS) e que é composto por Acelerômetros, receptores GPS, quatro vídeo câmeras, radar frontal e um computador de bordo. (SHRP 2, 2014)

O SHRP2 NDS é o maior estudo conduzido desse tipo, tendo aproximadamente 2.360 participantes, todos condutores comuns, que foram selecionados em seis locais diferentes dos Estados Unidos. O estudo considera que o comportamento do condutor é um fator crítico em quase todos os acidentes de trânsito, contudo, depois do acidente ocorrido não é possível determinar precisamente qual foi o comportamento do condutor. O método oferece duas principais vantagens: a) Informações detalhadas e precisas antes dos acidentes, incluindo informações a respeito do comportamento do condutor; b) Exposição das informações, incluindo a frequência do comportamento normal na direção, assim como os fatores com maior contribuição, no contexto dos riscos estimados para acidentes. (SHRP 2, 2014)

Um total de, aproximadamente, 6.600.000 dados foi coletado durante a SHRP2. Os vídeos passaram por um processo de redução através de algoritmos criados e através de análise manual, com uma equipe de 100 pessoas treinadas. (SHRP 2, 2016)



**Figura 3** - As imagens das quatro câmeras instaladas no interior do veículo, para o projeto NDS.

Para a organização, classificação e definições dos dados coletados no projeto foram elaborados os ‘dicionários’ para todos os pesquisadores envolvidos no projeto ou que tenham interesse na obtenção dos dados é fortemente recomendado a leitura desses dicionários.

A importância de apresentar este trabalho é a criar o paralelo entre o NDS, considerado o estado da arte no estudo do comportamento do condutor, e as técnicas utilizadas pelas empresas de ônibus urbano para controle do desempenho de seus motoristas.

## **2.5 Aplicações de sistemas de monitoramento por câmeras no transporte público urbano**

Em 2001 o TRB publicou um relatório (TCRP Synthesis 38: Electronic Surveillance Technology on Transit Vehicles, 2001) uma apresentação dos sistemas de monitoramento existentes na época, assim como descreve os objetivos do documento como sendo:

Examinar, na medida do possível, os objetivos das agências em selecionar e implantar tecnologias de monitoramento;

Descrever as tecnologias que eram utilizadas para atender os objetivos selecionados;

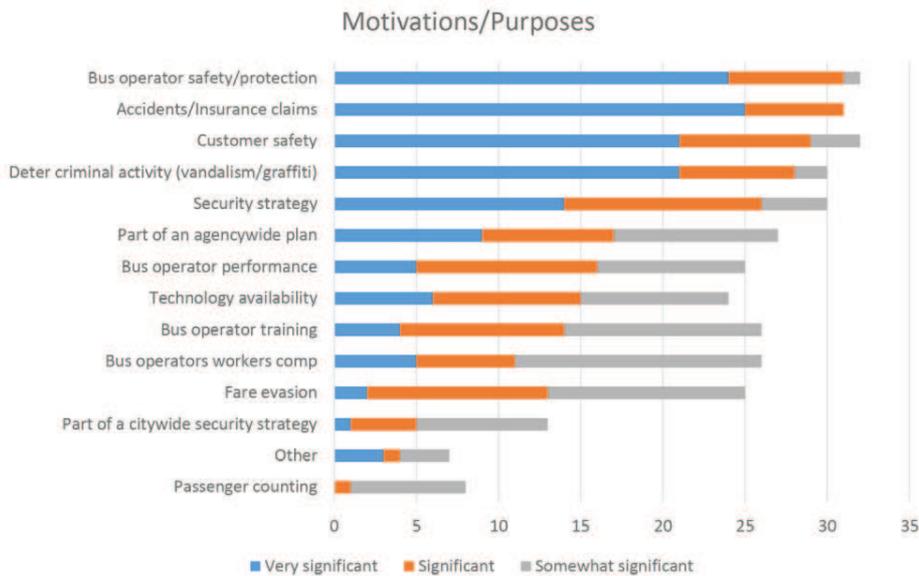
Detalhar os sistemas disponíveis, na época, os que eram escolhidos pelas agências e a evolução dos sistemas de monitoramento;

Documentar as práticas realizadas e futuras necessidades nessa área para que as agências de transporte público tivessem os procedimentos para uso do equipamento eletrônico de monitoramento.



motivador chave para a implantação de câmeras de monitoramento nas frotas de ônibus do país. (TRB, 2016)

Atualmente mais da metade das agências (18), participantes do questionário, tem 100% de sua frota com câmeras instaladas e 85% das agências que não tem sua frota totalmente equipada com o sistema de monitoramento pretende fazê-lo em breve.



**Figura 5** — Respostas para as Motivações/Propostas de implantação do sistema de monitoramento  
 Fonte: (TCRP SYNTHESIS 123 - Onboard Camera Applications for Buses, 2016, p. 8)

A Figura 5 apresenta o resultado das respostas dadas pelas agências quanto à motivação e proposta de implantação dos sistemas de monitoramento por câmeras. Na Figura, 100% das agências apresentam como motivação a segurança/proteção do condutor e do passageiro, e com 97% das agências (31) os acidentes/indenizações. Apenas 25 agências identificaram como de alguma forma significativa o Desempenho do Condutor.

O interessante dos resultados que embora as agências tenham classificado como ‘muito significativo’ a segurança do condutor e do passageiro, não houve a mesma preocupação com o desempenho do condutor, uma vez que pode ser fator de grande impacto na segurança da viagem.

## 2.6 Taxonomias

Como definição básica para taxonomia, temos: ciência ou técnica de classificação. Com a informatização de processos voltados à de geração de dados tornaram as técnicas de

classificação e ordenação ainda mais importantes, caso contrário seria criado uma quantidade de dados sem objetividade ou de maneira desordenada.

A caracterização de um acidente é necessária para determinar a natureza de sua ocorrência, por isso Wåhlberg (2001) elaborou uma taxonomia dos acidentes ocorridos com ônibus urbanos na Suécia, sendo seu principal objetivo categorizar os acidentes no cenário existente.

Os padrões de acidentes podem variar desde a natureza física do evento (por exemplo, qual o local do veículo foi danificado) até com relação à gravidade do evento (com vítimas ou sem vítimas). Segundo Wåhlberg (2001), é comum a utilização de taxonomias para categorizar os acidentes de trânsito, para os mais variados fins e tipos de pessoas e/ou entidades, como por exemplo; pesquisadores, companhias de seguros, polícia e empresas de tráfego.

O principal objetivo é determinar um padrão nos eventos e agregá-los pelas semelhanças, para que seja possível, também, direcionar a coleta ou separação de dados. No caso de Wåhlberg (2001), a taxonomia serviu para delimitar os dados utilizados em sua pesquisa, de maneira que com os próprios dados ele validou e demonstrou a confiabilidade e adaptação da taxonomia.

O HSRC (*Highway Safety Research Center*) (2003) desenvolveu, também, uma taxonomia (taxonomia das distrações) para orientar seu estudo e organizar os dados em categorias. A taxonomia apresentada na Fase II do projeto é um aperfeiçoamento da taxonomia inicialmente proposta, na Fase I, que foi baseada na análise dos dados. O objetivo principal da criação de uma taxonomia é a classificação e definição dos dados que irão compor o estudo, ou a organização dos dados já existentes.

Para Wåhlberg (2001), a criação de uma taxonomia deve ter um limite máximo de categorias, sendo que muitas categorias resultariam em pequenas amostras de dados.

No caso de Reason (REASON, 1990), já apresentado no item 2.2, a taxonomia tem aspecto mais teórico, visa entender e definir os atos inseguros.

Este capítulo apresenta os trabalhos realizados com objetivos semelhantes, para introduzir o estado da arte com relação ao problema, assim como desenvolve uma base teórica dos processos utilizados nesta pesquisa. Para o trabalho de NDS foi apresentado com o objetivo de apresentar um programa que trabalha com tratamento de imagens e monitoramento de

condutores como é a proposta deste trabalho, distinguindo-se apenas pelo fato dos condutores do NDS não serem profissionais.

### 3 MÉTODOS E DESENVOLVIMENTO

---

Este capítulo apresenta quais os métodos utilizados para trabalhar os dados, assim como, as técnicas utilizadas para o desenvolvimento teórico.

#### 3.1 Índices de Acidentes

Os índices utilizados neste trabalho foram previamente apresentados por Ferraz *et al.* (2012, p. 88), no entanto foram adaptados para a realidade e dados deste projeto de pesquisa.

##### 3.1.1 Índice de acidente por quilômetro percorrido

As empresas operadoras de transporte público urbano utilizam, para fins de planejamento das linhas de operação, os quilômetros totais percorridos pelas linhas, ou seja, este valor é expresso em km.frota. Para o índice de acidentes, portanto, temos:

$$I_k = \frac{N * 10^3}{KP} \quad (3.1)$$

Onde:

$I_k$  = Índice de Acidente por km, por linha (Acidentes/km.frota);

$N$  = Número de acidentes ocorridos por linha de operação;

$KP$  = Quilômetros percorridos por linha de operação (km.frota).

##### 3.1.2 Índice de acidente por tempo de operação

O valor para as horas operadas é similar ao quilômetro percorrido, leva em consideração a frota que opera em cada linha e é obtido o total de horas operadas em cada linha.

$$I_h = \frac{N * 10^2}{HO} \quad (3.2)$$

Onde:

$I_h$  = Índice de Acidente por hora operada, por linha (Acidentes/horas.frota);

$N$  = Número de acidentes ocorridos por linha de operação;

$HO = \text{Horas operadas por linha de operação (horas.frota)}$ .

### 3.1.3 Índice de acidente por número de passageiros transportados

Este índice exprime a relação dos quocientes, acidentes e número de passageiros transportados por linha.

$$I_p = \frac{N * 10^3}{P} \quad (3.3)$$

Onde:

$I_p = \text{Índice de Acidente por passageiros transportados, por linha (Acidentes/pass.frota)}$ ;

$N = \text{Número de acidentes ocorridos por linha de operação}$ ;

$P = \text{Número de passageiros transportados por linha de operação}$ .

## 3.2 Distribuição estatística dos dados

Dados de contagem não têm distribuição normal, ou seja, o estudo da frequência de ocorrência de eventos é atribuído à Distribuição de Poisson. Contudo, quando ocorre superdispersão (variabilidade maior que o esperado) é necessário utilizar a distribuição binomial negativa, que se ‘encaixa’ melhor nos dados.

O programa utilizado para os cálculos estatísticos foi o *R*, com o objetivo de determinar a qual distribuição os dados pertencem e quais são as variáveis mais significativas.

No programa *R* existe um pacote GLM (*Generalized Linear Models*), que fornece como resposta um modelo linear dos dados inseridos, assim como os parâmetros e índices. Um desses índices é o AIC (*Akaike Information Criterion*) que foi utilizado como critério de escolha do modelo, utilizando o menor valor de AIC. Segundo Kuha (2004), a utilização de mais índices e mais informações são necessários para ajustamento de modelos, no entanto, para simples comparação o AIC é capaz de apresentar bons resultados. Entende-se, portanto, que a utilização de mais índices para este trabalho não seja necessário, uma vez que o objetivo do processamento dos dados é de apenas obter os eventos mais significantes e não ajustamento dos modelos.

As duas distribuições foram testadas; Distribuição Negativa Binomial (*Negative Binomial*) e Distribuição de Poisson. Ambas foram testadas considerando excessos de zeros também, sendo: Distribuição Binomial Negativa com Zero Inflado (*Negative Binomial Zero-Inflated*) e Distribuição de Poisson com Zero Inflado (*Poisson Zero-Inflated*).

### 3.2.1 Distribuição de Poisson (DP)

A Distribuição de Poisson surgiu no ano de 1837, quando Simeon Denis Poisson publicou uma aproximação limite da distribuição binomial, dando origem a essa distribuição que herdaria o seu nome. Seus princípios básicos consistem na independência e mesma probabilidade de ocorrência entre os eventos. Essa distribuição também pode ser usada nos casos em que o número de tentativas de um evento é muito grande e a probabilidade de ocorrência é muito baixa (JONHSON *et al.*, 1969).

$$f(k; \lambda) = \frac{e^{-\lambda} * \lambda^k}{k!} \quad (3.4)$$

Sendo:

$e$  = base do logaritmo natural ( $e = 2,7182 \dots$ );

$k$  = número de ocorrências ( $k$  é inteiro e não negativo);

$\lambda$  = taxa de ocorrências é um número real ( $\lambda > 0$ ).

É possível obter a Distribuição de Poisson utilizando o Processo de Poisson, para isto, considere um dado intervalo de números reais e suponha que as contagens ocorram através desse intervalo. Subdividindo esse intervalo em subintervalos de comprimentos suficientemente pequenos, pode-se propor que (FUMES, 2009):

1. A probabilidade de mais de uma contagem em um subintervalo seja nula;
2. A probabilidade de uma contagem em um subintervalo seja a mesma para todos os subintervalos;
3. A contagem em cada subintervalo seja independente de outros subintervalos.

### 3.2.2 Distribuição Binomial Negativa (DBN)

A origem da Distribuição Binomial Negativa ocorreu no ano de 1679 com Pascal e Fermat, sendo também conhecida como Distribuição de Pascal (JONHSON *et al.*, 1969). A distribuição binomial negativa surgiu como uma generalização da distribuição geométrica,

que é definida como o tempo de espera do primeiro sucesso numa sequência de ensaios de Bernoulli com probabilidade  $\theta$ . De fato, a distribuição binomial negativa nada mais é do que a distribuição do tempo de espera do  $k$ -ésimo sucesso (MEYER, 1974).

$$b(n; k; \theta) = \binom{n-1}{k-1} \theta^k (1-\theta)^{n-k} \quad (3.5)$$

Sendo:

$k = \text{número de tentativas};$

$n = \text{número de experimentos de Bernoulli};$

$\theta = \text{Probabilidade de sucesso}.$

A distribuição indica o número de tentativas necessárias para obter  $k$  sucessos de igual probabilidade  $\theta$  ao fim de  $n$  experimentos de Bernoulli, sendo a última tentativa um sucesso.

### 3.3 Modelos Inflacionados de Zeros

A geração de modelos de previsão de eventos, que se enquadra em determinadas distribuições, tem como alternativa, quando da ocorrência de muitos zeros, a utilização de modelos inflacionados de zeros. Estes modelos têm como estrutura duas partes, ou duas equações (FUMES, 2009):

- 1) Os zeros constantes nos dados podem ser entendidos como:
  - a. Zeros estruturais: quando há impossibilidade de ocorrer determinado evento;
  - b. Zeros amostrais: quando a resposta é nula, contagem igual à zero.
- 2) O modelo considera a contagem não nula na qual se enquadra em uma distribuição (Poisson ou Binomial Negativa).

Um exemplo para ilustrar os zeros estruturais, apresentado por Fumes (2009) é a quantidade de mulheres de uma região que não são mães, que podem ocorrer por duas razões distintas: a mulher não é mãe porque não pode ter filhos (zeros estruturais) ou porque a mulher ainda não teve filhos (zeros amostrais).

Os modelos inflacionados de zeros, portanto, têm como resultado duas equações, uma representa os valores de contagens pertencentes à distribuição e a outra é responsável pelos zeros constantes nos dados (sejam estruturais ou amostrais).

### 3.4 Critério de Informação de Akaike - AIC

O critério de AIC é amplamente utilizado para comparação de modelos estatísticos (KUHA, 2004).

Este critério tem a seguinte equação:

$$AIC = 2[l(\widehat{\theta}_2) - l(\widehat{\theta}_1) - 2(p_2 - p_1)] \quad (3.6)$$

Sendo:

$l(\widehat{\theta}_2)$  e  $l(\widehat{\theta}_1)$  = *logaritmos das funções de verossimilhança;*

$p_1$  e  $p_2$  = *graus de liberdade dos respectivos modelos;*

Temos, portanto,  $l(\widehat{\theta}_2)$  e  $l(\widehat{\theta}_1)$  que são logaritmos das funções de verossimilhança de dois modelos ( $M_1$  e  $M_2$ , por exemplo) e  $p_1$  e  $p_2$  são os graus de liberdade dos respectivos modelos. É importante destacar que o AIC leva em consideração o número de parâmetros  $p$ .

Para comparação entre modelos, com mesmo conjunto de dados, o menor número de AIC é considerado o modelo mais representativo ou o modelo em que os dados se encaixam melhor.

A aplicação desses métodos garante a extração de resultados adequados para os dados do estudo de caso, que são apresentados no capítulo seguinte.



## 4 CARACTERÍSTICAS DO TRANSPORTE PÚBLICO E DO BANCO DE DADOS – ESTUDO DE CASO

---

Este capítulo tem o objetivo de apresentar as características do sistema de transporte público urbano, objeto deste estudo de caso, expõe como os dados empregados foram gerados e demonstra o tratamento em que os dados foram submetidos.

### 4.1 Características do sistema de transporte público

A cidade de Ribeirão Preto-SP é objeto do estudo de caso deste trabalho. A cidade está localizada no interior do estado de São Paulo e conta com, aproximadamente, 650 mil habitantes.

O transporte público urbano da cidade de Ribeirão Preto-SP oferece para os serviços aproximadamente 320 veículos (ônibus e micro-ônibus) para atender 115 linhas com 750 condutores, percorrendo da ordem de 2.000.000 quilômetros por mês e transporta 3.400.000 passageiros/mês, aproximadamente.

Todos os ônibus do sistema de transporte público desta cidade são equipados com câmeras de vídeos que filmam tanto o ambiente interno quanto o externo do veículo. Os vídeos são gravados durante toda a operação, desde a saída dos veículos da garagem da empresa (início da operação), até o seu retorno (final da operação). A Figura 6 apresenta a imagem de uma gravação das câmeras e como são disponíveis para o tratamento.



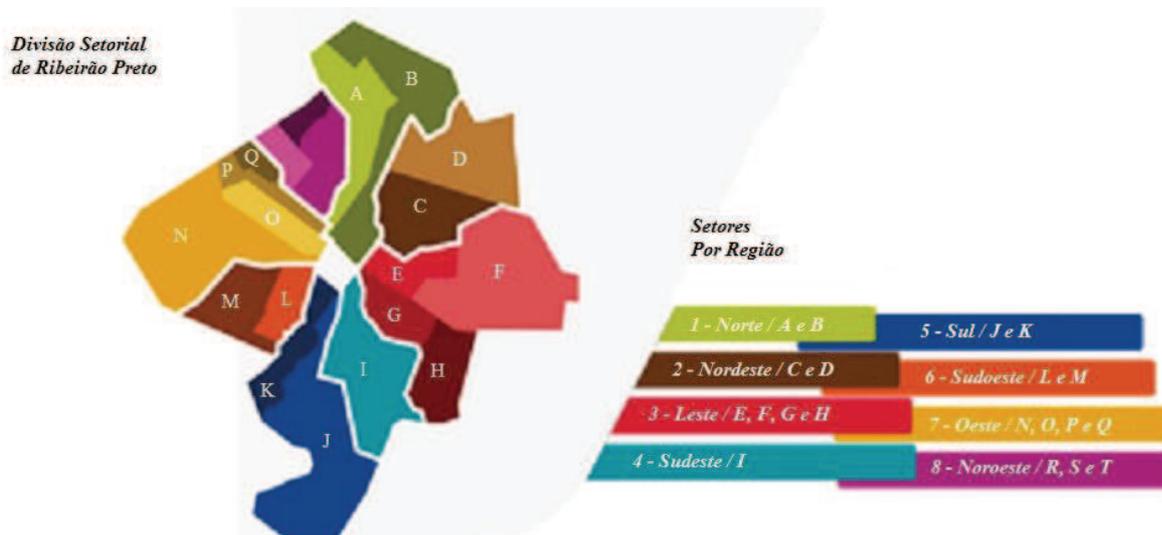
Figura 6- Imagens geradas pelas câmeras de monitoramento

O tratamento das imagens é realizado por uma equipe que trabalha diariamente gerando dados que correspondem aos diferentes eventos, sejam estes com respeito ao comportamento dos condutores, sejam quanto ao comportamento dos usuários.

Com as imagens, é possível identificar, por exemplo, se o condutor do ônibus avança o semáforo na banda vermelha (imagem no canto inferior direito), assim como visualizar qual o comportamento do condutor enquanto dirige (imagem no canto superior esquerdo). O tratamento das imagens, portanto, ocorre através da identificação de infrações dos condutores que podem ser visualizadas nos vídeos.

O total de dados fornecidos para este trabalho é da ordem de 76 mil ocorrências, que correspondem a um período, aproximado, de quinze meses (11/02/2013 à 26/05/2014 - 469 dias).

O sistema de transporte público local foi subdividido em áreas geográficas, sendo a cidade dividida em 22 regiões e cada linha de operação é inserida em uma região.



**Figura 7** - Organização das regiões do município de Ribeirão Preto

A estrutura do banco de dados tem como principais campos:

- Dia do evento;
- Horário do evento;
- Tipo do evento (acidente, utilização de celular, avanço de sinal, etc.);
- Identificação do veículo em operação;
- Linha de operação;
- Área da cidade em que a linha pertence;

- Observações (descrição dos eventos).

O sistema atualmente é utilizado para segurança dos veículos, condutores e passageiros. O acompanhamento dos eventos gerados pelo tratamento dos dados é realizado, principalmente, para identificar assaltos, inibir fraude (venda clandestina de passes) e monitorar o comportamento dos condutores e passageiros.

#### **4.2 Apresentação dos dados brutos do banco de dados e a adaptação ao estudo**

Os dados fornecidos pela empresa dizem respeito a qualquer evento ocorrido no interior do veículo e que faz parte do processo de geração de dados do monitoramento por câmeras. Muitos, no entanto, não são aplicáveis para este trabalho e foram, portanto, não considerados nas análises. Todo o processo de mineração dos dados foi realizado com o *software MySQL* que é direcionado a gerenciamento de banco de dados.

Inicialmente foi fornecido pela empresa um total de 75.896 dados. A redução do banco de dados foi de aproximadamente 40,5% o que corresponde a 30.722 eventos. Após esta redução, foram considerados nas análises iniciais deste trabalho um total de 45.174 dados. Os eventos retirados são, por exemplo, aqueles que constam no banco de dados como ‘Não identificados’, ‘Limpeza do ônibus’, ‘Corte de gravação’, entre outros que não tem aplicação no contexto deste trabalho.

Outra redução realizada foi quanto ao tratamento dos dados, por exemplo, um mesmo evento era descrito em dois dados diferentes. Isso ocorria, geralmente, quando a gravação do vídeo iniciava em outro cartão (durante a gravação dos vídeos são disponibilizados 2 *slots* de cartão de memória, quando o espaço disponível de um acaba, automaticamente passa a ser gravado no outro), portanto, eram gerados dois eventos. Esses eventos foram unificados. A duplicidade de dados também foi identificada e excluída do banco de dados. Houve, também, a exclusão de dados que não tinham o critério estabelecido como ‘Conversa durante a condução’. Este critério é apresentado no item 4.3.

Os dados efetivamente utilizados para este estudo foi de 33.038, uma redução de 56,5% do total dos dados.

### 4.3 Resumo dos dados utilizados

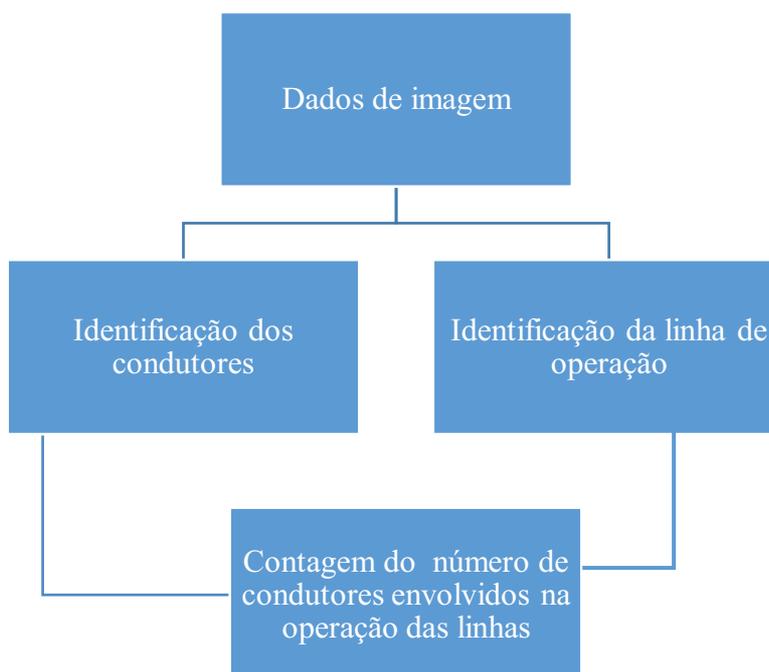
De acordo com o descrito anteriormente, caracterizando o sistema de transporte público, são aproximadamente 115 linhas, considerando que, os micro-ônibus que operam algumas linhas não contêm as câmeras embarcadas, e por isso, não entraram no universo de estudo.

O total de linhas apresentadas no estudo é de 89 e as tabelas com as informações das linhas, como por exemplo, quilometragem e quantidade de horas de operação, estão no Apêndice A, separadas por tipo de linha de operação (Radial, Diametral, etc.).

### 4.4 Dados extraídos *versus* dados criados a partir do banco de dados

Alguns processos realizados gerou a criação de novos dados, diferenciando, por exemplo, daqueles já existentes no banco de dados. A quantidade de condutores envolvidos na operação de cada linha, por exemplo.

#### 4.4.1 Quantidade de condutores envolvidos na operação de uma linha



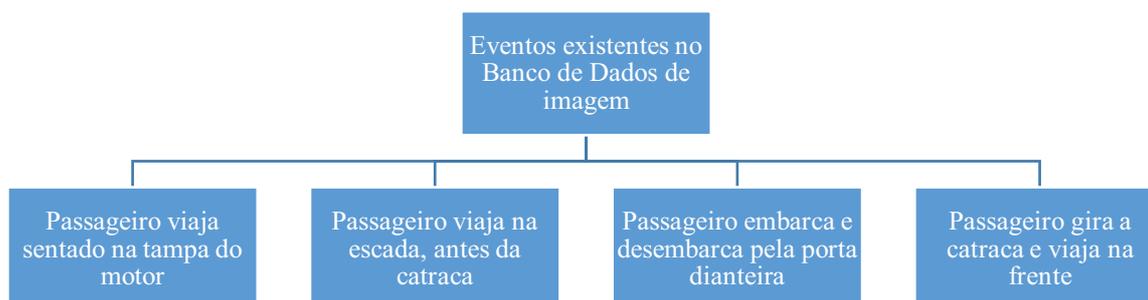
**Figura 8** - Criação de novo dado a partir de dados existentes

A Figura 8 ilustra o processo de criação da contagem do número de condutores por linha de operação, através da quantidade de vezes que um condutor aparecia operando determinada linha. Esta informação, portanto, representa apenas o número de condutores que estão inseridos no banco de dados, em outras palavras, os que se envolveram em qualquer evento pertencente ao bando de dados.

#### 4.4.2 Conversa durante a condução

Este processo foi realizado baseado na taxonomia criada para estes dados, conforme descrito no Capítulo 5, contudo nesta seção será apresentada apenas de onde foram extraídas as informações para a composição deste evento, deixando a explicação para o próximo capítulo, onde será apresentada a taxonomia.

A Figura 9 apresenta 4 eventos existentes originalmente no banco de dados que, eventualmente, têm uma característica em comum: passageiro conversando com o condutor durante a viagem. Portanto, os dados que tinham a informação adicional de que houve conversa durante a condução foram selecionados, extraídos e compostos em um novo evento. Estes eventos previamente existentes no banco de dados deixaram de existir (foram excluídos), passando a ter uma única denominação “Conversa durante a condução”.



**Figura 9** - Eventos que compõem a criação de um novo dado

#### 4.4.3 Dados das linhas de operação

Os dados de demanda de passageiros existentes neste estudo diz respeito a quantidade média de usuários transportados por linha. O cálculo foi realizado a partir de dias típicos, através do banco de dados de bilhetagem e, finalmente, estimado uma média por linha. Essa média apresentada não condiz com o período de estudo, mas sim uma média geral.

As quantidades de quilômetros percorridos e horas de operação foram disponibilizadas por linha e é uma média mensal geral também.

#### **4.4.4 Dificuldades com o banco de dados**

Os dados enviados pela empresa passaram por um processo de tratamento, necessário para compatibilizar as informações.

A partir de 11/02/2013 todo o sistema de transporte público urbano foi modificado, principalmente no que diz respeito a descrição das linhas de operação. Desta data em diante, portanto, as linhas passaram a ter denominações diferentes. O banco de dados apresenta as descrições nos dois formatos (antigo e atual), apresentando uma quantidade de linhas maior do que o existente e com descrições diferentes para a mesma linha.

A ausência de padronização na criação dos nomes das linhas também apresentaram problemas para a realização do trabalho com os dados. O excesso de espaços, por exemplo, existentes na denominação das linhas não permitia a correta agregação dos dados. Ocorria, também com relação a abreviações (Jardim Paulista ou JD. Paulista) ou até mesmo com utilização de letras maiúsculas e minúsculas. Este problema não ocorreu apenas na denominação das linhas, mas também na descrição dos Eventos.

Eventos sem a correta definição tornavam-se ambíguos, no caso dos acidentes, por exemplo, o banco de dados apresentava 4 eventos descritos como “Acidente”, “Acidente sem vítima”, “Acidente com vítima” e “Colisões”. Não havendo distinção, pela empresa, dos eventos nomeados como “Colisão” e “Acidente”. Havia outros eventos com o mesmo problema.

A resolução para os problemas apontados foi através do *MySQL* inserindo as linhas de programação para atualização das informações. No caso das linhas com as denominações antigas foi necessário ter acesso a essas denominações e fazer identificar quais eram as denominações atuais.

## 5 ANÁLISE E RESULTADOS

Este item tem como objetivo apresentar as análises realizadas a partir dos dados, assim como os resultados provenientes da aplicação dos métodos.

### 5.1 Análise exploratória

A Tabela 1 contém os dados médios por dia, mês e linha, dos eventos em estudo do banco de dados.

Os dados referentes aos acidentes não fazem distinção da gravidade do evento, estão inclusos desde colisões até mesmo atropelamentos seguidos de morte, por exemplo. Esta decisão permite o entendimento do evento em si, independentemente da classificação da gravidade, afinal em circunstâncias idênticas pode existir um acidente com ou sem gravidade.

**Tabela 1** - Média dos eventos

Descrição	Média por dia *	Média por mês **	Média por linha ***
ACIDENTES	1,1	31,3	5,6
ASSALTO	0,3	8,4	1,5
BRIGAS	0,0	1,3	0,2
DANOS_VEICULO	0,2	6,9	1,2
DESENTENDIMENTO_MOTORISTA_E_PASSAGEIRO	0,0	0,4	0,1
AVANCO_SINAL_VERMELHO	13,3	389,4	70,0
AVANCO_SINAL_AMARELO	9,3	273,9	49,2
COMENDO_DIRIGINDO	2,6	74,6	13,4
DIRIGINDO_FALANDO_CELULAR	11,3	330,8	59,5
FUMANDO_DIRIGINDO	0,3	9,6	1,7
USANDO_FONE_DE_OUVIDO	0,3	8,8	1,6
VIOLANDO_AS_REGRAS_EMPRESA	0,0	1,2	0,2
CONVERSA_CONDUCAO	31,1	911,1	163,8
QUEDA_ACIDENTAL	0,1	3,7	0,7

\* 469 dias

\*\* 16 meses

\*\*\* 89 linhas de operação

#### 5.1.1 Índices de acidentes

Os índices foram calculados de acordo com as fórmulas constantes no Capítulo 3, as Tabelas 2, 3 e 4 apresentam os valores dos índices.

As informações utilizadas para o cálculo dos índices estão incluídas nas Tabelas, contudo os valores de quilômetros e horas operadas foram estimados para o período em estudo (15,3 meses). Estas informações expressam também a frota que opera a linha, os dados são representados pela unidade ‘km\*frota’, ‘horas\*frota’ e ‘passageiros\*frota’.

A Tabela 2 os dados estão classificados pelo índice de acidentes por quilômetro as 30 primeiras linhas de operação, do maior para o menor, nota-se que a linha 52R é uma Radial com pouca quantidade de quilômetros percorridos e apenas 2 acidentes, com características similares a linha 66N (terceira posição). Estas linhas mantêm posição para a Tabela 3, que os dados estão classificados pelo índice de acidentes por hora operada.

**Tabela 2** – Dados das linhas classificados pelo  $I_k$  (Índice de acidentes por quilômetro)

Cód. Linha	Acidentes	Área da linha	Tipo da linha	km/Mês	Horas Oper./Mês	Demanda total/Mês	$I_k$	$I_h$	$I_p$
52R	2	GRUPO R	RADIAL	1638,00	46,20	3658	0,783	2,775	0,350
63R	21	GRUPO M	RADIAL	27128,60	1858,58	64575	0,496	0,724	0,208
66N	2	GRUPO J	NOTURNA	2637,00	124,00	1564	0,486	1,034	0,820
88R	4	GRUPO T	RADIAL	5319,30	147,70	11766	0,482	1,736	0,218
69R	7	GRUPO M	RADIAL	10832,40	559,50	30561	0,414	0,802	0,147
40R	20	GRUPO R	RADIAL	33493,20	2044,67	58392	0,383	0,627	0,220
75D	5	GRUPO D	DIAMETRAL	8892,00	554,50	6534	0,360	0,578	0,491
10R	6	GRUPO B	RADIAL	10692,00	493,50	9629	0,360	0,779	0,399
82R	14	GRUPO T	RADIAL	28327,50	1683,88	32073	0,317	0,533	0,280
9R	20	GRUPO R	RADIAL	40904,40	2488,00	98778	0,313	0,515	0,130
84N	1	GRUPO Q	NOTURNA	2088,00	124,00	256	0,307	0,517	2,504
48R	8	GRUPO L	RADIAL	18172,00	1080,17	39038	0,282	0,475	0,131
65C	11	GRUPO U	CIRCULAR	25321,00	1187,00	29436	0,278	0,594	0,240
26R	10	GRUPO N	RADIAL	23685,60	1698,30	73128	0,271	0,377	0,088
38R	8	GRUPO L	RADIAL	19261,80	1088,42	37163	0,266	0,471	0,138
44R	7	GRUPO D	RADIAL	19116,50	1035,82	39920	0,235	0,433	0,112
47D	18	GRUPO B	DIAMETRAL	49946,73	2677,90	81913	0,231	0,431	0,141
51R	12	GRUPO R	RADIAL	33732,20	2027,87	57475	0,228	0,379	0,134
24R	9	GRUPO K	RADIAL	25443,90	1522,92	76246	0,227	0,379	0,076
8R	4	GRUPO O	RADIAL	11677,50	755,55	25971	0,220	0,339	0,099
13D	7	GRUPO I	DIAMETRAL	20573,90	1505,42	38706	0,218	0,298	0,116
43D	17	GRUPO J	DIAMETRAL	50667,39	2427,00	84128	0,215	0,449	0,130
25R	7	GRUPO L	RADIAL	20945,10	2126,42	48053	0,214	0,211	0,093
79R	9	GRUPO T	RADIAL	28973,22	1853,92	56267	0,199	0,311	0,103
64R	9	GRUPO G	RADIAL	29645,16	1670,05	48493	0,195	0,345	0,119
6R	3	GRUPO K	RADIAL	9975,00	559,00	18878	0,193	0,344	0,102
5R	5	GRUPO I	RADIAL	16964,04	1002,85	61649	0,189	0,320	0,052
53C	7	GRUPO U	CIRCULAR	23753,99	1128,13	27773	0,189	0,398	0,162
60R	5	GRUPO C	RADIAL	17442,00	1110,00	24643	0,184	0,289	0,130
56R	14	GRUPO D	RADIAL	48872,16	2480,92	86718	0,184	0,362	0,103

**Tabela 3** – Dados das linhas classificados pelo  $I_h$  (Índice de acidentes por hora operada)

Cód. Linha	Acidentes	Área da linha	Tipo da linha	km/Mês	Horas Oper./Mês	Demanda total/Mês	$I_k$	$I_h$	$I_p$
52R	2	GRUPO R	RADIAL	1638,00	46,20	3658	0,783	2,775	0,350
88R	4	GRUPO T	RADIAL	5319,30	147,70	11766	0,482	1,736	0,218
66N	2	GRUPO J	NOTURNA	2637,00	124,00	1564	0,486	1,034	0,820
69R	7	GRUPO M	RADIAL	10832,40	559,50	30561	0,414	0,802	0,147
10R	6	GRUPO B	RADIAL	10692,00	493,50	9629	0,360	0,779	0,399
63R	21	GRUPO M	RADIAL	27128,60	1858,58	64575	0,496	0,724	0,208
49R	2	GRUPO N	RADIAL	9891,00	194,75	14172	0,130	0,658	0,090
40R	20	GRUPO R	RADIAL	33493,20	2044,67	58392	0,383	0,627	0,220
65C	11	GRUPO U	CIRCULAR	25321,00	1187,00	29436	0,278	0,594	0,240
75D	5	GRUPO D	DIAMETRAL	8892,00	554,50	6534	0,360	0,578	0,491
82R	14	GRUPO T	RADIAL	28327,50	1683,88	32073	0,317	0,533	0,280
84N	1	GRUPO Q	NOTURNA	2088,00	124,00	256	0,307	0,517	2,504
9R	20	GRUPO R	RADIAL	40904,40	2488,00	98778	0,313	0,515	0,130
48R	8	GRUPO L	RADIAL	18172,00	1080,17	39038	0,282	0,475	0,131
38R	8	GRUPO L	RADIAL	19261,80	1088,42	37163	0,266	0,471	0,138
43D	17	GRUPO J	DIAMETRAL	50667,39	2427,00	84128	0,215	0,449	0,130
44R	7	GRUPO D	RADIAL	19116,50	1035,82	39920	0,235	0,433	0,112
47D	18	GRUPO B	DIAMETRAL	49946,73	2677,90	81913	0,231	0,431	0,141
67R	14	GRUPO A	RADIAL	53235,00	2229,75	88034	0,169	0,402	0,102
53C	7	GRUPO U	CIRCULAR	23753,99	1128,13	27773	0,189	0,398	0,162
51R	12	GRUPO R	RADIAL	33732,20	2027,87	57475	0,228	0,379	0,134
24R	9	GRUPO K	RADIAL	25443,90	1522,92	76246	0,227	0,379	0,076
26R	10	GRUPO N	RADIAL	23685,60	1698,30	73128	0,271	0,377	0,088
21R	5	GRUPO C	RADIAL	20760,96	866,82	19096	0,154	0,370	0,168
56R	14	GRUPO D	RADIAL	48872,16	2480,92	86718	0,184	0,362	0,103
39R	2	GRUPO N	RADIAL	8034,00	356,25	14754	0,160	0,360	0,087
64R	9	GRUPO G	RADIAL	29645,16	1670,05	48493	0,195	0,345	0,119
61R	13	GRUPO D	RADIAL	50782,64	2413,57	77012	0,164	0,345	0,108
6R	3	GRUPO K	RADIAL	9975,00	559,00	18878	0,193	0,344	0,102
8R	4	GRUPO O	RADIAL	11677,50	755,55	25971	0,220	0,339	0,099

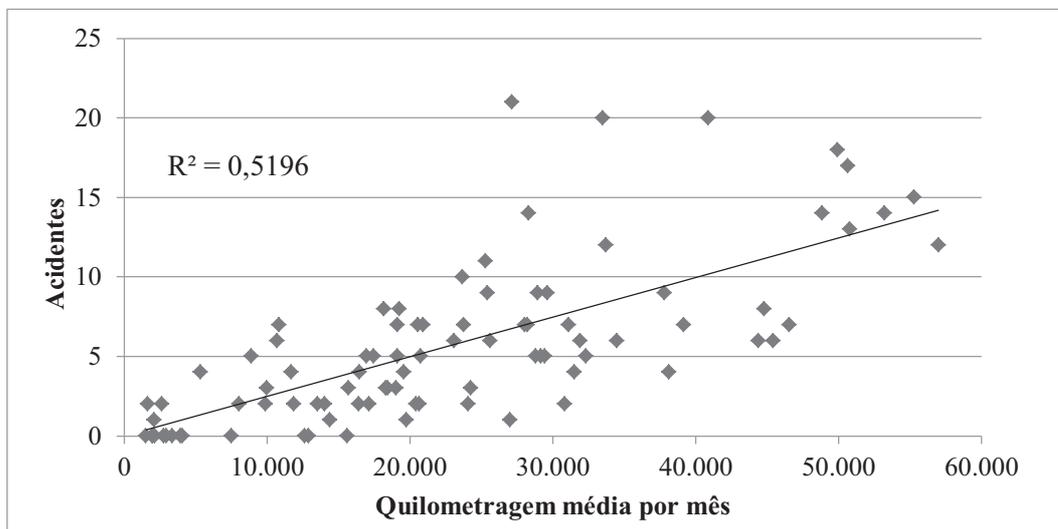
**Tabela 4** – Dados das linhas classificados pelo  $I_p$  (Índice de acidentes por passageiros)

Cód. Linha	Acidentes	Área da linha	Tipo da linha	km/Mês	Horas Oper./Mês	Demanda total/Mês	$I_k$	$I_h$	$I_p$
84N	1	GRUPO Q	NOTURNA	2088,00	124,00	256	0,307	0,517	2,504
66N	2	GRUPO J	NOTURNA	2637,00	124,00	1564	0,486	1,034	0,820
75D	5	GRUPO D	DIAMETRAL	8892,00	554,50	6534	0,360	0,578	0,491
10R	6	GRUPO B	RADIAL	10692,00	493,50	9629	0,360	0,779	0,399
52R	2	GRUPO R	RADIAL	1638,00	46,20	3658	0,783	2,775	0,350
82R	14	GRUPO T	RADIAL	28327,50	1683,88	32073	0,317	0,533	0,280
65C	11	GRUPO U	CIRCULAR	25321,00	1187,00	29436	0,278	0,594	0,240
40R	20	GRUPO R	RADIAL	33493,20	2044,67	58392	0,383	0,627	0,220
88R	4	GRUPO T	RADIAL	5319,30	147,70	11766	0,482	1,736	0,218
63R	21	GRUPO M	RADIAL	27128,60	1858,58	64575	0,496	0,724	0,208
21R	5	GRUPO C	RADIAL	20760,96	866,82	19096	0,154	0,370	0,168
53C	7	GRUPO U	CIRCULAR	23753,99	1128,13	27773	0,189	0,398	0,162
31R	4	GRUPO C	RADIAL	19580,80	818,23	16558	0,131	0,313	0,155
69R	7	GRUPO M	RADIAL	10832,40	559,50	30561	0,414	0,802	0,147
47D	18	GRUPO B	DIAMETRAL	49946,73	2677,90	81913	0,231	0,431	0,141

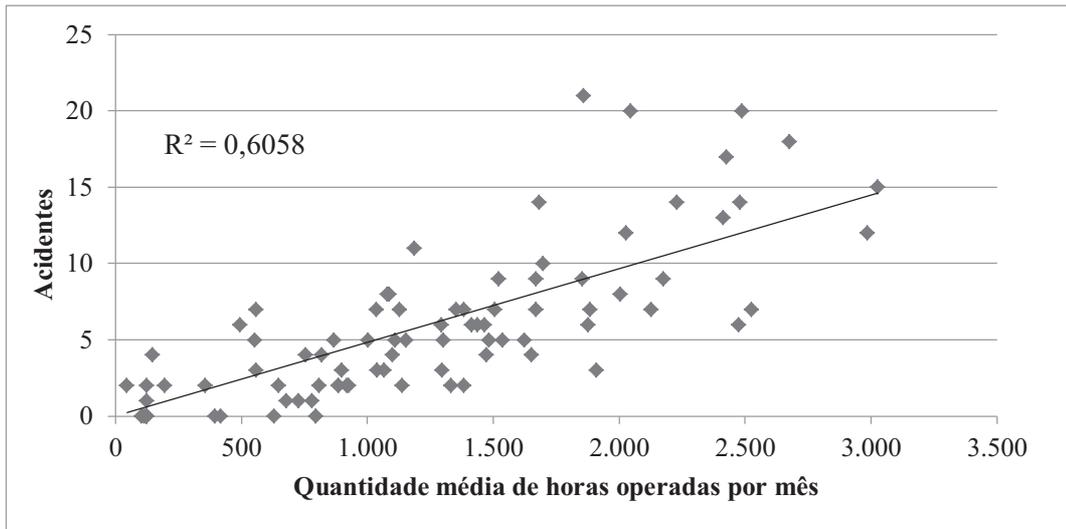
38R	8	GRUPO L	RADIAL	19261,80	1088,42	37163	0,266	0,471	0,138
62D	5	GRUPO E	DIAMETRAL	19126,80	1153,50	23772	0,168	0,278	0,135
51R	12	GRUPO R	RADIAL	33732,20	2027,87	57475	0,228	0,379	0,134
17P	2	GRUPO V	PERIMETRAL	11880,00	648,50	9717	0,108	0,198	0,132
48R	8	GRUPO L	RADIAL	18172,00	1080,17	39038	0,282	0,475	0,131
20R	7	GRUPO A	RADIAL	31125,00	1383,25	34199	0,144	0,324	0,131
60R	5	GRUPO C	RADIAL	17442,00	1110,00	24643	0,184	0,289	0,130
9R	20	GRUPO R	RADIAL	40904,40	2488,00	98778	0,313	0,515	0,130
43D	17	GRUPO J	DIAMETRAL	50667,39	2427,00	84128	0,215	0,449	0,130
30P	8	GRUPO V	PERIMETRAL	44805,60	2004,00	39869	0,114	0,256	0,129
64R	9	GRUPO G	RADIAL	29645,16	1670,05	48493	0,195	0,345	0,119
33C	15	GRUPO U	CIRCULAR	55297,60	3027,97	82343	0,174	0,318	0,117
13D	7	GRUPO I	DIAMETRAL	20573,90	1505,42	38706	0,218	0,298	0,116
44R	7	GRUPO D	RADIAL	19116,50	1035,82	39920	0,235	0,433	0,112
86R	7	GRUPO T	RADIAL	39164,40	1885,40	40654	0,115	0,238	0,110

### 5.1.2 Acidentes

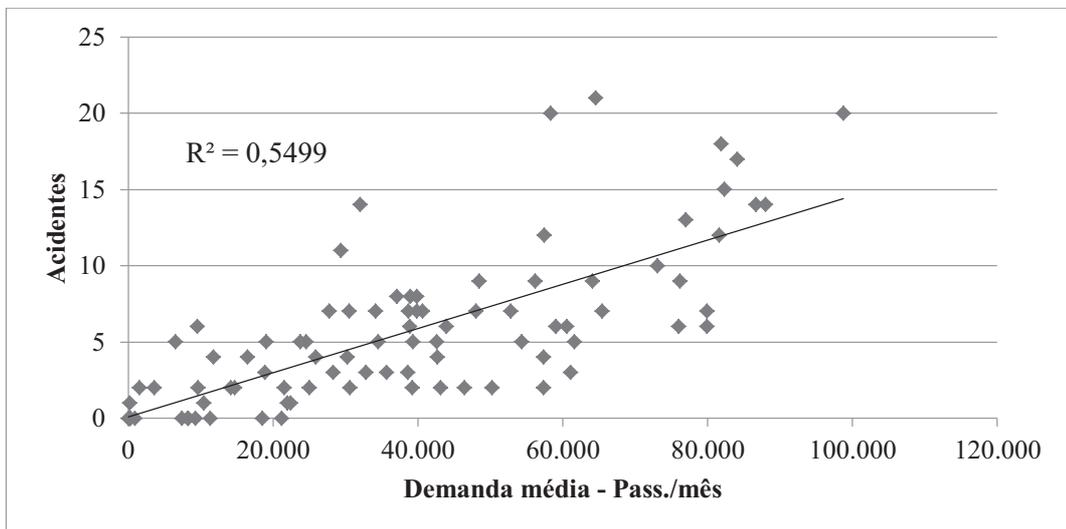
Os Gráficos 1 e 2 apresentam os dados de acidentes ‘*versus*’ quilometragem e quantidade de horas de operação, e com valores de 72% e 78% de correlação, respectivamente. É possível notar maior linearidade para estes dados, reafirmando que quanto maior a exposição dos veículos, maior a probabilidade de se envolver em acidentes. O Gráfico 3, no entanto, demonstra a relação Acidentes ‘*versus*’ Demanda média, que tem correlação de 74%.



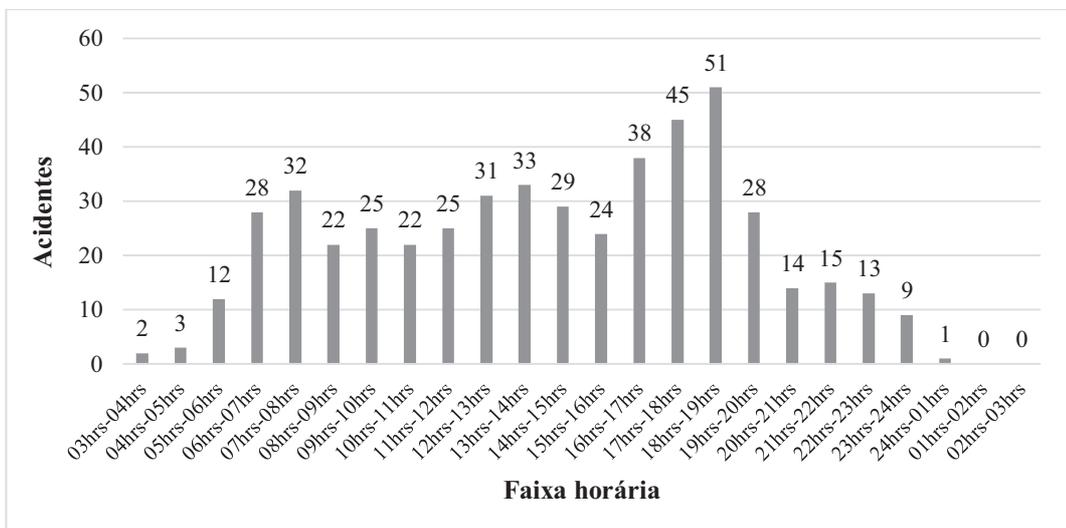
**Gráfico 1** - Gráfico de dispersão apresentando número de acidentes pela quilometragem média por mês, por linha.



**Gráfico 2** - Gráfico de dispersão apresentando número de acidentes pela quantidade média de horas operadas por mês, por linha.

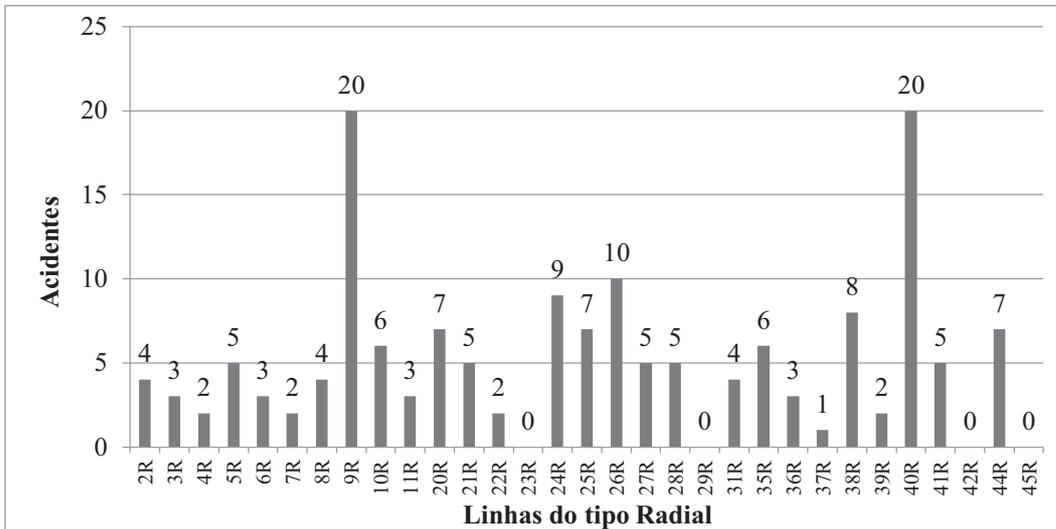


**Gráfico 3** - Gráfico de dispersão apresentando número de acidentes pela quantidade média de passageiros por mês, por linha.

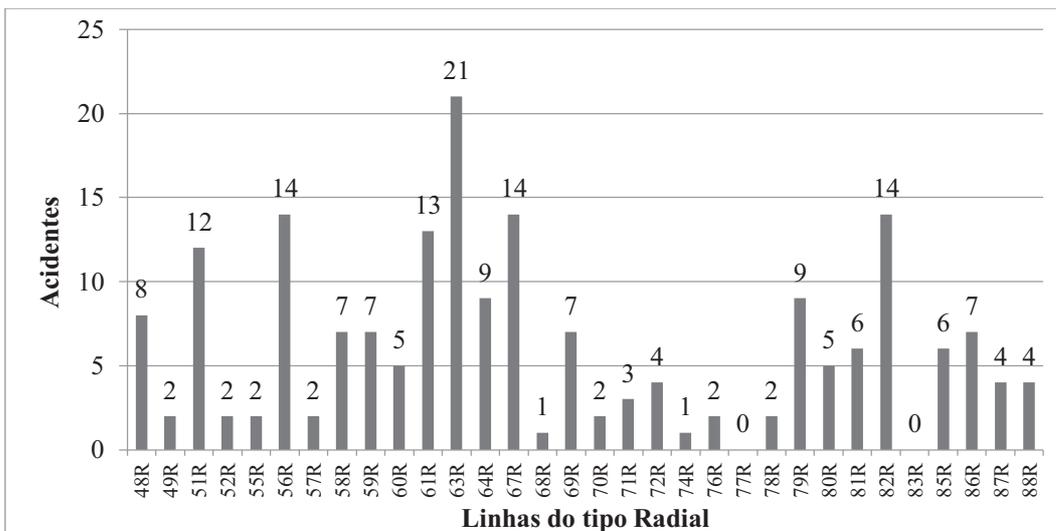


**Gráfico 4** - Frequência de acidentes por faixa horária

O Gráfico 4 apresenta o número de acidentes por faixa horária (ao longo do período em estudo), onde é possível observar que ocorrem com maior frequência durante os horários de pico (06 às 09 horas da manhã, 11 às 14 horas da tarde e das 16 às 19 horas).

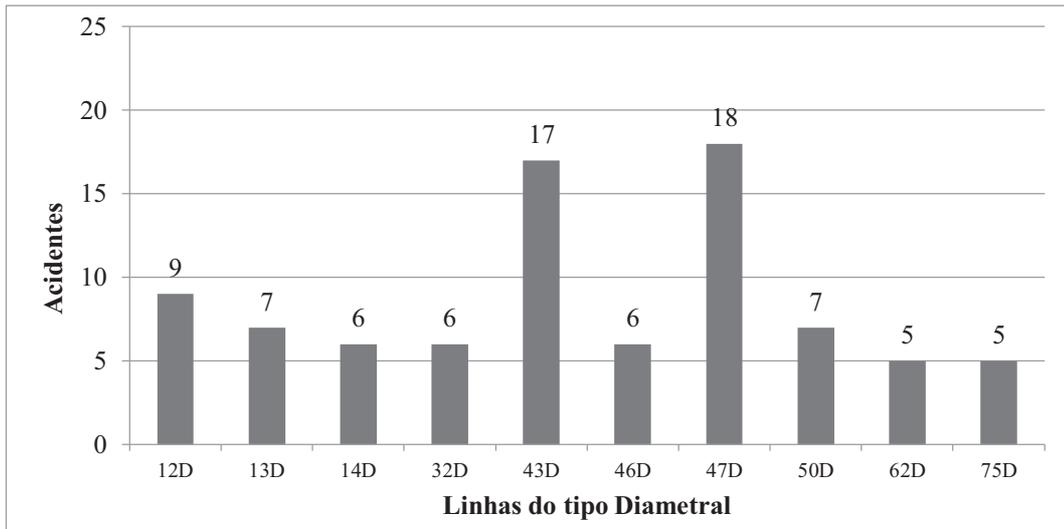


**Gráfico 5 - Número de acidentes em linhas Radiais**

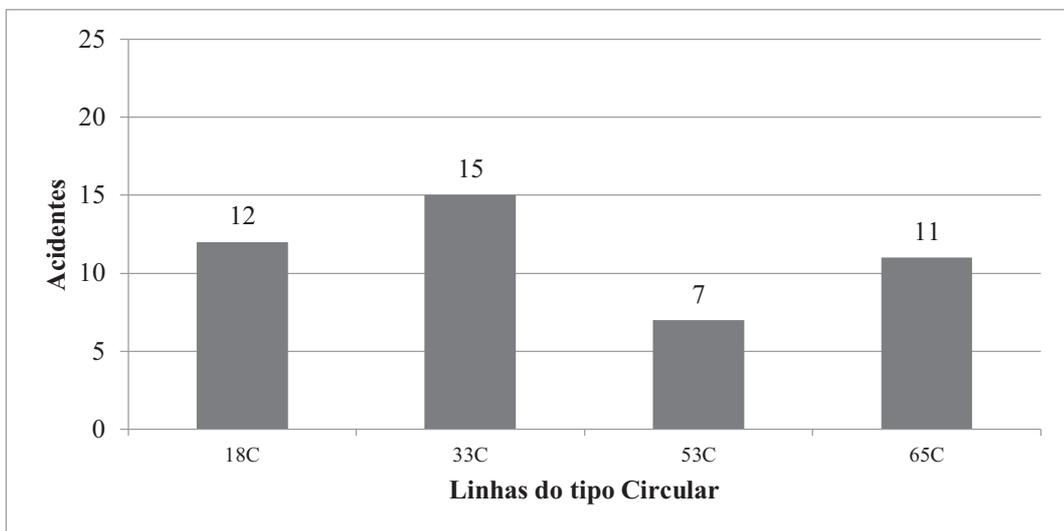


**Gráfico 6 - Número de acidentes em linhas Radiais**

Os Gráficos 5 e 6 ilustram a quantidade de acidentes por linha do tipo Radial. É possível perceber uma variação grande do número de acidentes, uma vez que o sistema de transporte público é majoritariamente com linhas Radiais.

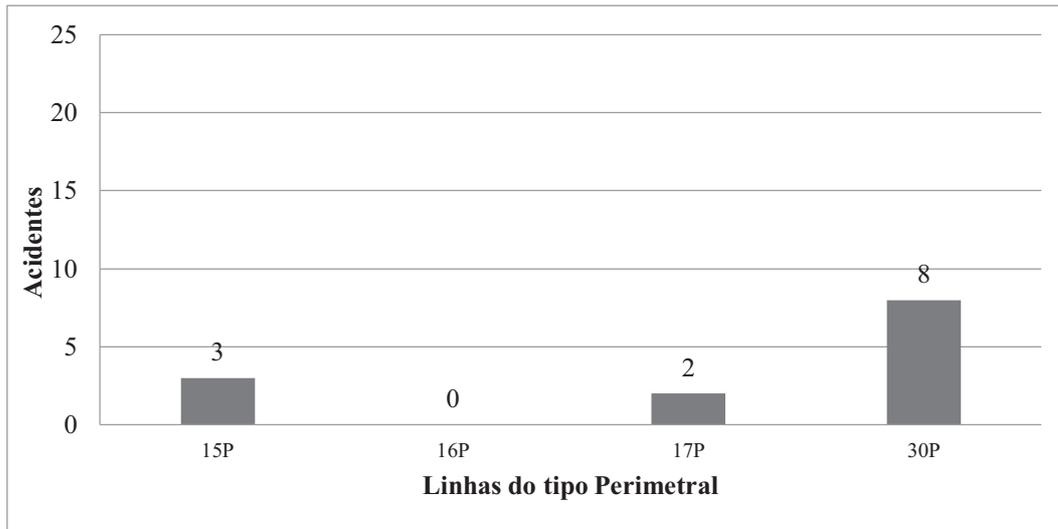


**Gráfico 7** - Número de acidentes em linhas Diametraais



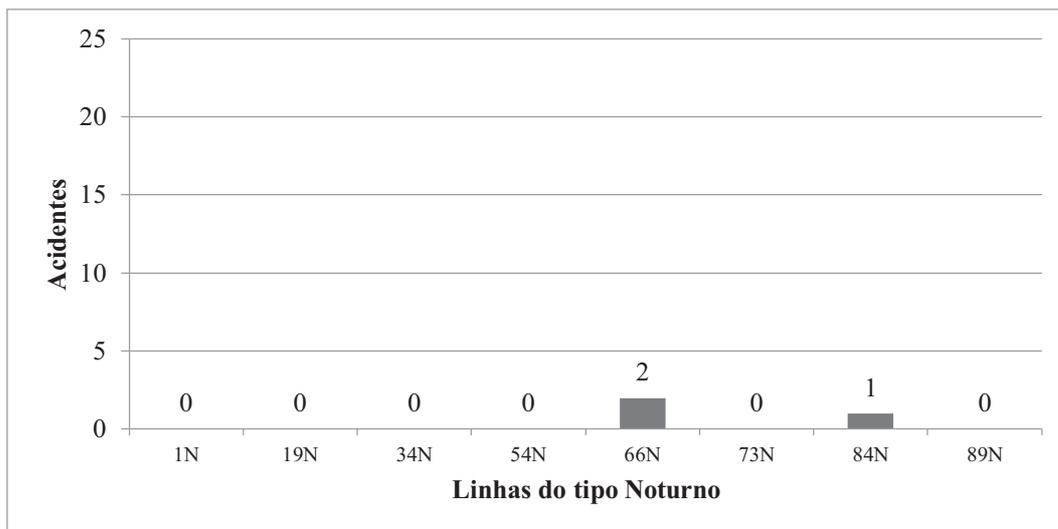
**Gráfico 8** Número de acidentes em linhas Circulares

Há uma considerável contribuição do número de acidentes pelas linhas circulares, Gráfico 8, assim como das linhas do tipo Diametral, Gráfico 7. Nas linhas noturnas, que têm características quase idênticas entre si, há um destaque para a linha 66N, que têm 2 acidentes, diante das demais com 0 e 1 acidentes.



**Gráfico 9** - Número de acidentes em linhas Perimetrais

A partir do Gráfico 11 até o Gráfico 15, são apresentados os valores dos acidentes por faixa horária, separada por tipo de linha.



**Gráfico 10** - Número de acidentes em linhas Noturnas

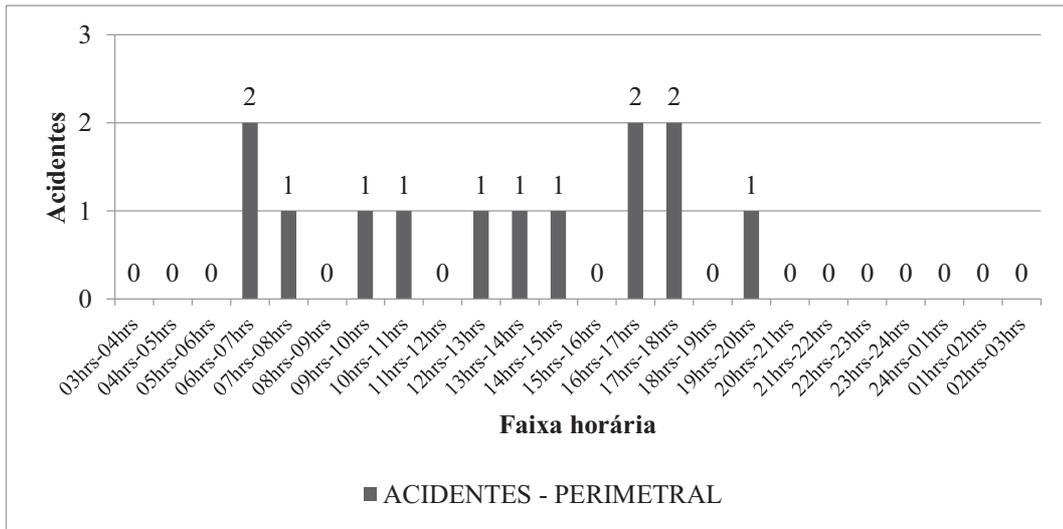


Gráfico 11 - Número de acidentes por faixa horária, linhas perimetrais

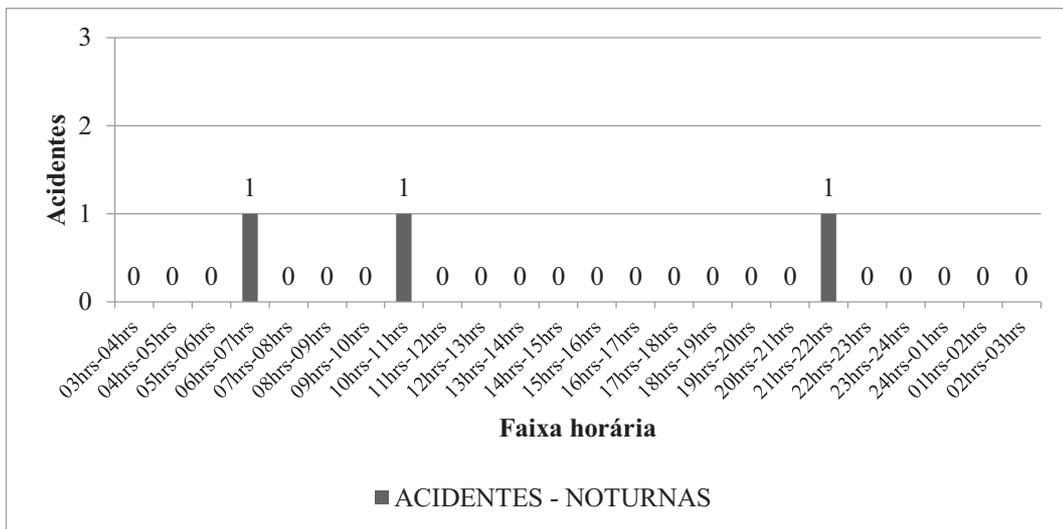


Gráfico 12 - Número de acidentes por faixa horária, linhas noturnas

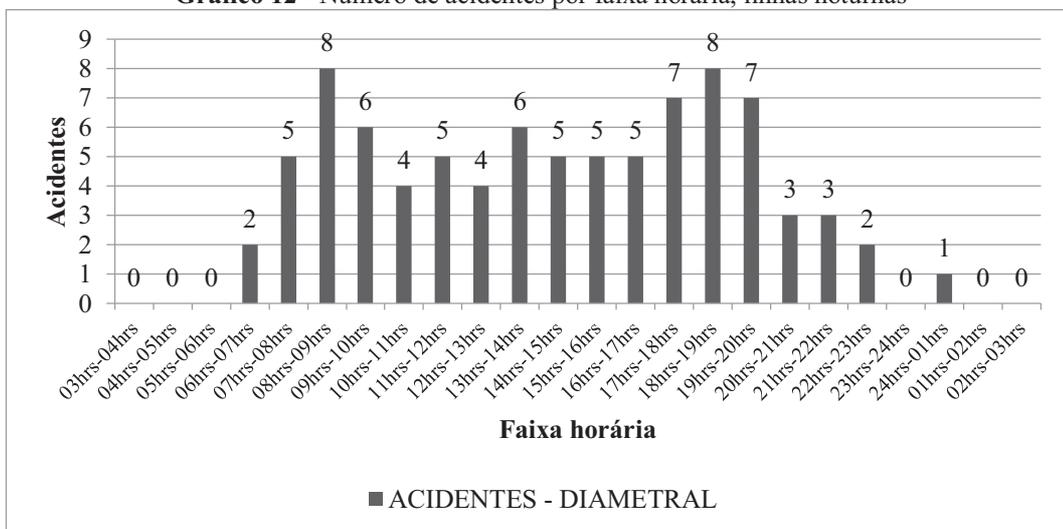


Gráfico 13 - Número de acidentes por faixa horária, linhas diametrais

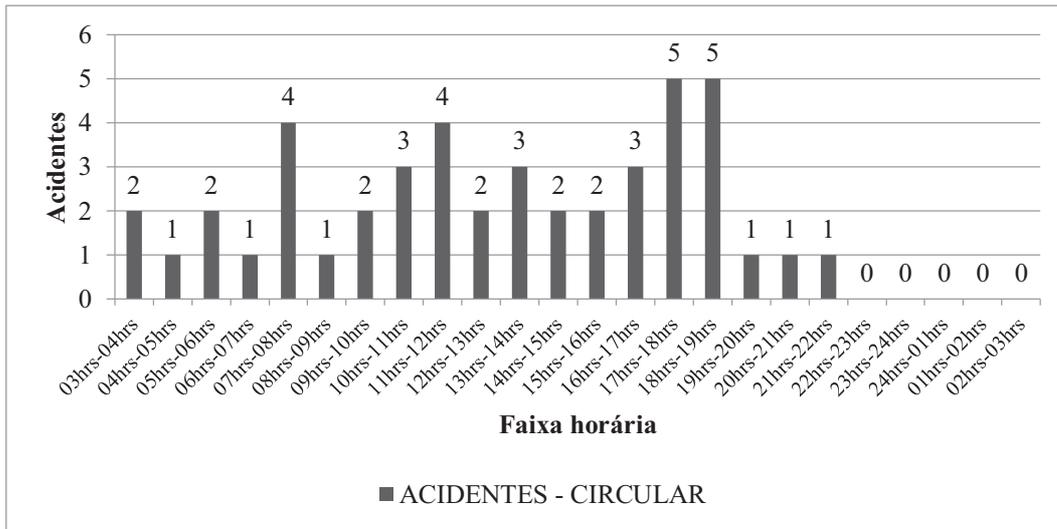


Gráfico 14 - Número de acidentes por faixa horária, linhas circulares

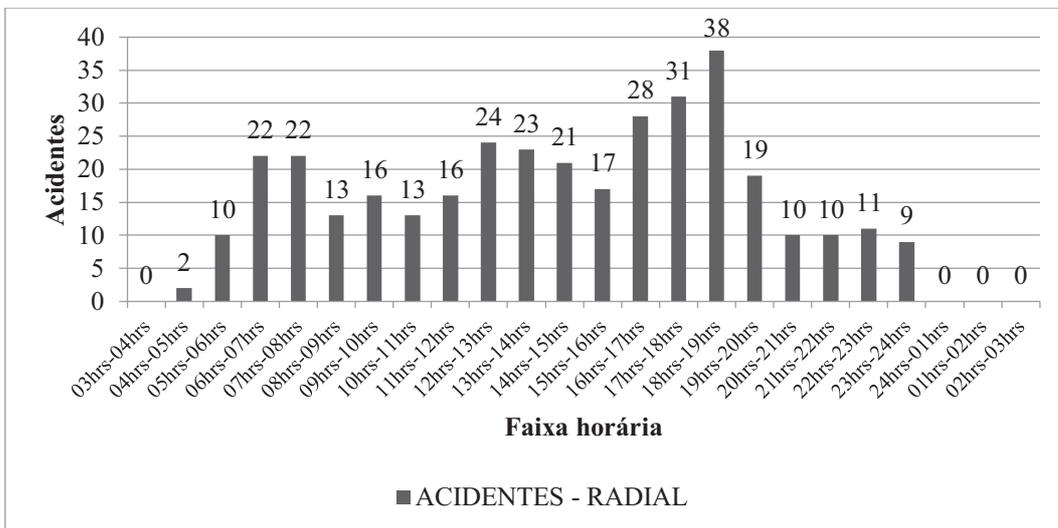
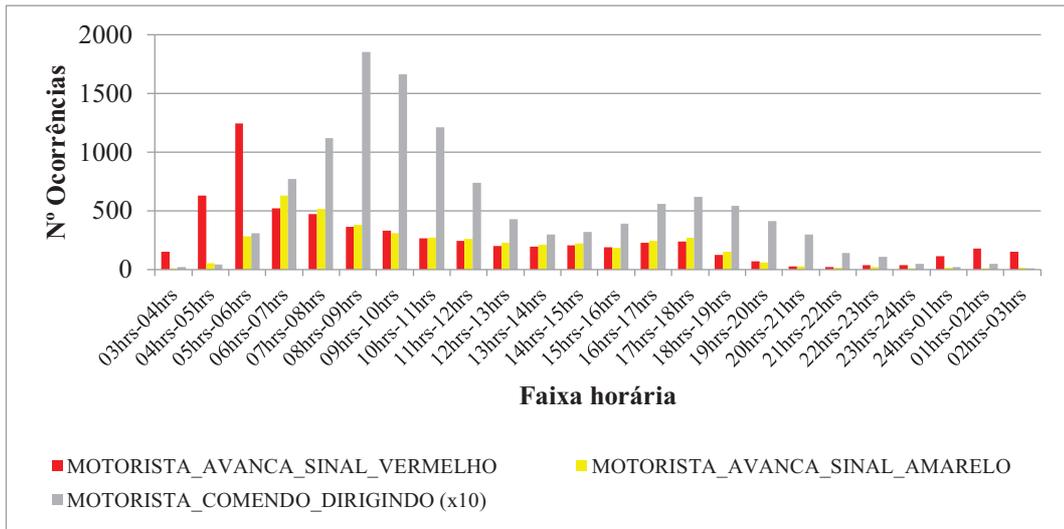


Gráfico 15 - Número de acidentes por faixa horária, linhas radiais

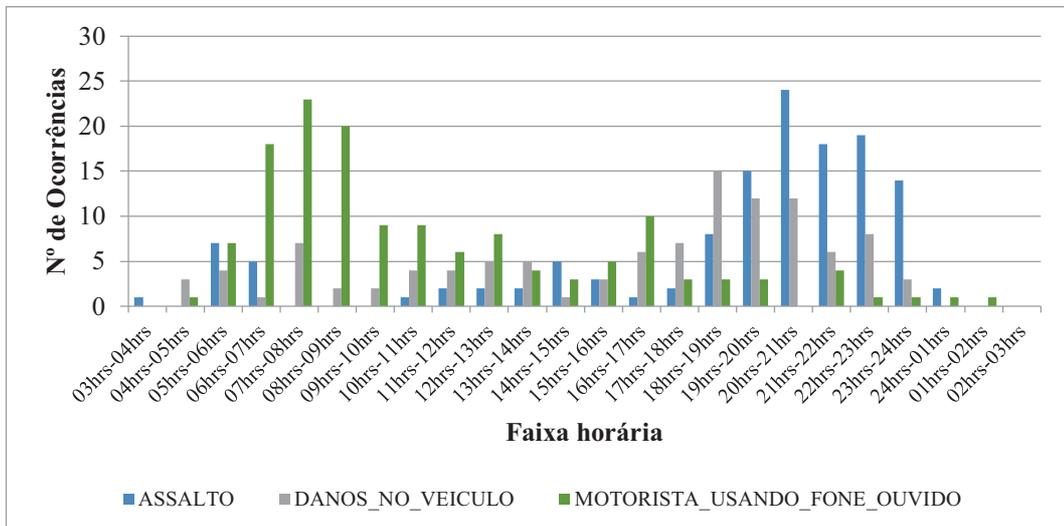
### 5.1.3 Demais ocorrências

Alguns eventos têm comportamento similar aos acidentes, tendem a aumentar nos horários de pico, conversa e condução, comendo e dirigindo, avanço de sinal amarelo, conforme Gráfico 16.

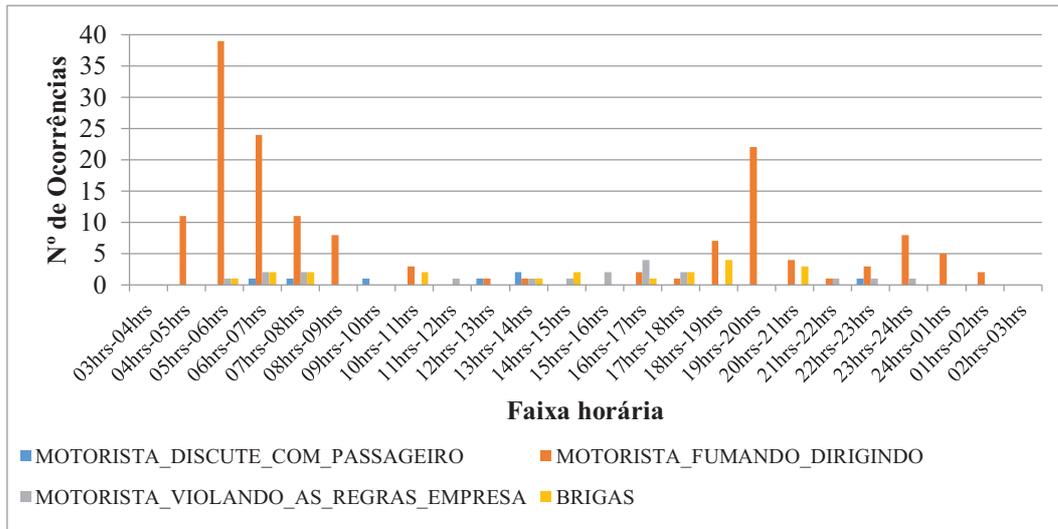


**Gráfico 16** - Número de eventos por faixa horária, avanço de sinal e comendo durante a condução.

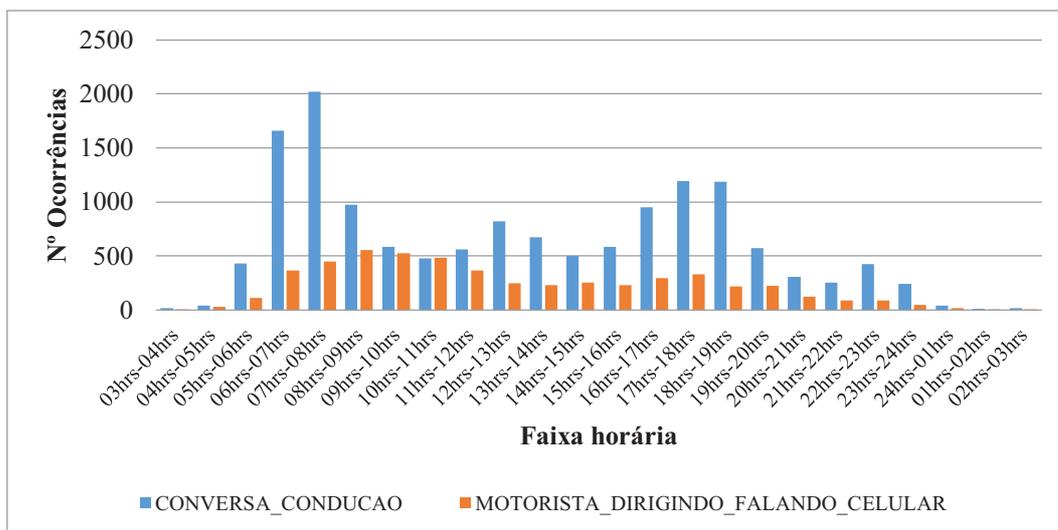
A ocorrência do avanço de sinal vermelho, no entanto, tem alta concentração no período das 04 às 06 horas da manhã. É possível explicar, essa alta frequência de avanços de sinal vermelho, pelo fato de não existir fiscalização nesse horário e ser uma prática comum, utilizar o ‘sinal vermelho’ como o sinal de ‘Pare’, por exemplo, caso em que a pessoa para, olha e segue em frente.



**Gráfico 17** - Número de eventos por faixa horária, assalto, danos ao veículo e condutor utilizando fone



**Gráfico 18** - Número de eventos por faixa horária, brigas, condutor fuma ou discute com passageiro



**Gráfico 19** - Número de eventos por faixa horária, conversa durante a condução e utilização de celular

Para a utilização do celular e conversa durante a condução foi possível extrair algumas informações adicionais, como duração do evento, através do campo ‘Observações’ onde são inseridos os horários de início e término do evento.

De acordo com a Tabela 5, 33% do total dos eventos de uso de celular durante a condução ultrapassam 5 min, tornando-se predominante. Na sequência, até 2 min e 3 min de utilização do celular, com 21% e 17%, respectivamente. Para extração desses dados houve perda de 3% dos dados que não tinham informações completas a respeito do tempo de utilização do celular.

**Tabela 5** – Tempo de uso do celular

Tempo	Quantidade de eventos	Percentual
Até 1 Min	735	14%
Até 2 Min	1.139	21%
Até 3 Min	911	17%
Até 4 Min	570	11%
Mais de 5 Min	1774	33%
TOTAL	5129	97%

**Tabela 6-** Período de maior utilização do celular

Período	Horário	Quantidade de eventos	Percentual
Pico da manhã	Das 06 às 09 hrs	1.373	26%
Entre pico	Das 11 às 13 hrs	613	12%
Pico da tarde	Das 16 às 19 hrs	847	16%
TOTAL		2.833	53%

A Tabela 6 mostra que 53% do uso de celular ocorrem nos horários de pico. Do mesmo modo, as Tabelas 7 e 8 apresentam informações a respeito da conversa durante a condução, onde há um percentual de 43% de conversas que ultrapassam 25 minutos.

**Tabela 7** - Tempo de conversa durante condução

Tempo	Quantidade de eventos	Percentual
Até 5 Min	990	7%
Até 10 Min	2155	15%
Até 15 Min	2347	16%
Até 20 Min	2278	16%
Até 25 Min	6244	43%
TOTAL	14014	96%

**Tabela 8** - Período de maior frequência de conversas

Período	Horário	Quantidade de eventos	Percentual
Pico da manhã	Das 06 às 09 hrs	4.657	32%
Entre pico	Das 11 às 13 hrs	1.383	9%
Pico da tarde	Das 16 às 19 hrs	3.331	23%
TOTAL		9.371	64%

A frequência de conversa durante a condução aumenta naturalmente no horário de pico, uma vez que estes horários são os que têm maior concentração de passageiros.

## **5.2 Taxonomia – Baseada nos dados**

Neste item é apresentada uma proposta de taxonomia, que tem como principal objetivo estabelecer uma base teórica para a classificação dos comportamentos existentes no banco de dados. Esta base teórica servirá para entender melhor os resultados da análise estatística, apresentada mais adiante.

### **5.2.1 Do comportamento dos condutores**

A criação e adaptação de uma taxonomia condizente com o problema em estudo são importantes para o conhecimento dos níveis em que ocorrem as más decisões.

Entende-se que o comportamento do condutor tem inicialmente 3 categorias, que o separam em ‘induzido’, ‘decidido’ e ‘involuntário’. O comportamento discutido neste trabalho diz respeito aos que têm como consequência acidentes ou incidentes, ocorrendo simultaneamente ou não.

*Induzido:* Comportamento produzido pelas adversidades do ambiente de trabalho do condutor. É relacionado, principalmente, com o estresse gerado pelos eventos.

O condutor é diariamente submetido ao exercício de conduzir o ônibus em circunstâncias desconhecidas e que podem influenciar seu comportamento, contudo estes fatores externos estão fora do controle do condutor, induzindo-o a uma distração.

*Decidido:* Más decisões tomadas pelos condutores. Estes comportamentos são, em sua maioria, infrações (transgressões) das leis de trânsito, portanto, de conhecimento de todos os condutores, contudo, são realizados durante a condução;

*Involuntário:* Está categoria foi adaptada da taxonomia dos atos inseguros de James Reason (1990), onde o autor descreve como ‘Não intencionais’: Falhas e lapsos. Neste trabalho, são considerados como comportamentos involuntários, atos não planejados ou realizados de maneira involuntária (falta de atenção e esquecimentos, por exemplo).

A Figura 10 ilustra a estrutura elaborada para a taxonomia, isso não quer dizer, no entanto, que alguns comportamentos não possam ocorrer simultaneamente, ou que um comportamento

não seja causado por consequência de outro (avanço de sinal enquanto conversa, por exemplo).

Existe a ocorrência de avanço de sinal em duas categorias, Comportamento decidido e involuntário. Onde um ocorre por uma simples distração, ou esquecimento, e o outro com consciência. A causa da primeira pode decorrer de distrações externas que são difíceis de identificar, no entanto, a segunda pode ser causada, principalmente, pelo atraso na operação.

COMPORTAMENTOS			
Taxonomia dos Comportamentos dos Condutores	Imposições do ambiente	Brigas	Desentendimentos que ocorrem entre os passageiros e/ou envolvendo os condutores.
		Assaltos	Assaltos realizados no interior do veículo, envolvendo o condutor e/ou os passageiros.
		Danos ao veículo	Atos de vandalismo proferidos ao veículo, durante a operação do transporte público de passageiros.
		Tráfego de veículos	Para os ônibus que compartilham as vias com automóveis o tráfego intenso de veículos pode dificultar o cumprimento de horários de uma linha, por exemplo.
	Decisões do Condutor	Uso de celular	Utilização de celular durante a condução, independente do uso.
		Comer	Condutor se alimenta durante a condução do veículo
		Fumar	Condutor fuma durante a condução do veículo.
		Uso de fone de ouvido	Utilização de fone de ouvido durante a condução, independente do uso.
		Conversa	Condutor conversa com passageiros durante a condução do veículo.
		Avanço de sinal	Condutor avança sinal, na banda amarela ou vermelha.
	Falhas Lapsos	Avanço de sinal	Condutor avança sinal, na banda amarela ou vermelha.

**Figura 10** - Taxonomia do comportamento do condutor

O condutor, portanto, apenas conduzindo o veículo já tem grandes chances de acidente, uma vez que a própria estrutura de cognição permite a ele falhas e lapsos (comportamentos

involuntários), associados às imposições do ambiente externo, podemos dizer que suas chances, de se envolver em um acidente, aumentam consideravelmente. Quando somado aos desvios de atenção durante a condução, temos baixo grau cognitivo associado à condução do veículo. Aumento da chance de Erro.

É preciso, no entanto, estabelecer o perfil do condutor de ônibus urbano, uma vez que não se trata de um condutor comum. Sendo um profissional e que realiza suas atividades diariamente, a tarefa de dirigir se torna automática, ou mecânica. Em outras palavras, o trabalho de conduzir não exige deste condutor elevada atenção (cognição), pelo menos, não quando comparado com um condutor comum (que seu trabalho não é conduzir nenhum tipo de veículo).

O número de linhas que são designadas, geralmente, para um único condutor não excede 5. Ou seja, o condutor sempre faz os mesmos trajetos, itinerários, tornando ainda mais fácil o poder de automatização da tarefa.

### **5.3 Resultados estatísticos dos dados**

A análise realizada não considerou os comportamentos como causa dos acidentes, mas sim a quantidades de eventos praticados ao longo do período em estudo e a relação desses eventos com a ocorrência dos acidentes. Trata-se, portanto, da verificação da quantidade de eventos existentes *versus* quantidade de acidentes.

Para análise e processamento dos dados foi utilizado o *software* 'R', junto com os pacotes de processamento disponíveis no programa.

O método de seleção das variáveis foi realizado manualmente, conforme nível de significância. Foram inseridas todas as variáveis e retiradas uma a uma (método *backward elimination*).

O processamento se dividiu em 3 partes:

- 1) Os dados agregados por linha de operação;
- 2) Os dados agregados por condutores;
- 3) Os dados agregados por setor geográfico da cidade (divididos em 22 áreas).

Dados agregados por condutor tem uma contagem baixa para cada indivíduo (são aproximadamente 750 condutores), deixando a análise individualizada, portanto, os resultados são de significância baixa. Para essa análise seria importante ter uma base de dados maior, com um período de tempo maior, no entanto, necessitando também que os condutores fossem os mesmos ao longo do tempo e na prática não ocorre, a dinâmica de entrada e saída de profissionais é intensa.

Dados agregados por setor geográfico há uma alta concentração de eventos por área, contendo resultados altamente significativos, contudo leva a uma análise simplista e muito heterogênea (são englobados muitos condutores e linhas de operação em cada região).

Os dados agregados por linha de operação, diferentemente dos outros já apresentados, dispõem de características mais centrais, pois se tem sensibilidade suficiente para a análise dos resultados. O número de condutores, por exemplo, que operam as linhas é reduzido, atribuindo o evento, realizado naquela linha, a um grupo menor de condutores. De maneira similar, as linhas operam sempre na mesma região geográfica, os eventos ocorridos em determinada linha é conferido a uma única região.

### 5.3.1 Distribuição dos dados

Os dados foram testados a fim de se obter a distribuição que melhor se adequa aos dados. O critério utilizado foi o AIC (*Akaike Information Criterion*), amplamente utilizado para comparação de modelos.

**Tabela 9** - Valores do AIC para as diferentes distribuições

Distribuição	Linhas	Condutores	Área
Poisson	529,77	1705,40	157,85
ZIP	516,73	1686,69	S/N
NB	<b>461,62</b>	<b>1683,30</b>	<b>154,57</b>
ZINB	476,60	1687,31	S/N

A Tabela 9 apresenta os valores de AIC para cada uma das distribuições, assim como para os três diferentes agrupamentos dos dados.

O resultado, portanto, é a distribuição Binomial Negativa (valores em negrito), onde ocorre o menor valor numérico, comparado entre as outras distribuições de um único agrupamento (linhas).

É possível identificar, ainda na Tabela 9, que se realizar a comparação entre os 3 agrupamentos (colunas), o de menor valor seria o agrupamento por Área, conforme já comentado.

### 5.3.2 Processamento estatístico

As informações dos resultados dos cálculos estatísticos, realizados pelo programa *R*, estão apresentadas, integralmente, no Apêndice B.

O objetivo da análise se limita a extrair os comportamentos mais significativos com relação aos acidentes. Portanto, foi avaliado o número de ocorrências de acidentes pelo número dos eventos, já apresentados no item 5.1.

A seguir estão os resultados, separados por agrupamento, utilizando a distribuição Binomial Negativa.

#### 1. Agrupamento por linhas de operação:

Resultados:

##### a) Utilizando dados de comportamento:

**Tabela 10-** Resultado da análise estatística, utilizando variáveis de comportamento

Variável	Coef. estimado	Significância	AIC
<b>Uso de celular</b>	0.007796	4.99e-08 ***	461,62
<b>Avanço de Sinal Amarelo</b>	0.005928	0.00114 **	

A Tabela 10 mostra o resultado do processamento para o agrupamento por linhas de operação e trás informações do coeficiente estimado para cada variável, significância de cada variável e o valor de AIC para o modelo gerado.

##### b) Utilizando dados de comportamento e quantidade de condutores (conforme descrito no item 4.4.1):

O método *backward* retira progressivamente as variáveis menos significativas do resultado estatístico, contudo, a Tabela 11 apresentou uma resposta inadequada do ponto de vista técnico. O valor representando o coeficiente estimado significa que à medida que o número de acidentes aumenta, o número de avanço de sinal vermelho reduz, o inverso é verdadeiro. Portanto, a variável Avanço de Sinal Vermelho foi retirada e reprocessado o resultado, apresentado na Tabela 11. Esse procedimento foi adotado para os demais processamentos.

**Tabela 11** - Resultado da análise estatística, inserindo o número de condutores

Variável	Coef. estimado	Significância	AIC
<b>Número de condutores</b>	0.030685	< 2e-16 ***	430,36
<b>Avanço de Sinal Vermelho</b>	-0.004170	0.000566 ***	

**Tabela 12** - Resultado da análise estatística, inserindo apenas o número de condutores

Variável	Coef. estimado	Significância	AIC
<b>Número de condutores</b>	0.002544	< 2e-16 ***	438,49

## 2. Agrupamento por Área/Região Geográfica

**Tabela 13** - Resultado da análise estatística, para agrupamento por Região

Variável	Coef. estimado	Significância	AIC
<b>Uso de celular</b>	0.0026599	6.63e-09 ***	159.96

## 3. Agrupamento por Condutores

**Tabela 14** - Resultado da análise estatística, para agrupamento por Condutores

Variável	Coef. estimado	Significância	AIC
<b>Avanço de Sinal Amarelo</b>	0.025855	0.001185 **	1686.4
<b>Assalto</b>	0.343289	0.000147 ***	

## 4. Resumo

Nota-se que os resultados das Tabelas 13 e 14 apresentam variáveis existentes na Tabela 9, ou seja, repetição das variáveis para diferentes agrupamentos. A Tabela 15 resume as variáveis que se repetiram no processamento.

**Tabela 15** – Resumo dos resultados processados

Variável
Uso de celular
Avanço de Sinal Amarelo

#### 5.4 Discussão e análise dos resultados

A grande diferença do transporte público dos demais transportes é que não se pode mudar demasiadamente a rota dos veículos (por questões de demanda e por questões viárias) e não é possível reduzir o número de ônibus nas ruas (a menos que haja pouca demanda), pois o objetivo principal é garantir a mobilidade da população. Diferentemente do automóvel, por exemplo, que o cidadão pode escolher não utilizá-lo, o transporte público deve funcionar, pois representa o único modo motorizado acessível para qualquer pessoa, independente de idade, renda, etc.. Não é uma opção retirar o transporte público para reduzir o índice de acidentes ou mesmo prejudicar seu funcionamento.

Variáveis como quilometragem, horas de operação e demanda de passageiros, embora altamente correlacionadas com os acidentes, não foram considerados na estatística dos dados por serem variáveis fixas. Em outras palavras, nem sempre é fácil reduzir a quilometragem do sistema e esta tem uma variação muito baixa por mês. Contudo, para reduzir a quilometragem é necessária uma análise profunda na rede de transporte público e na demanda. Se simplesmente reduzir a quilometragem parte dos usuários serão sacrificados. Analogamente à quantidade de horas de operação.

Através das variáveis quilometragem, horas de operação e demanda de passageiros, foi calculado os índices, permitindo uma análise comparativa entre as linhas e, inclusive, entre os próprios índices, além de identificar linhas com possíveis problemas.

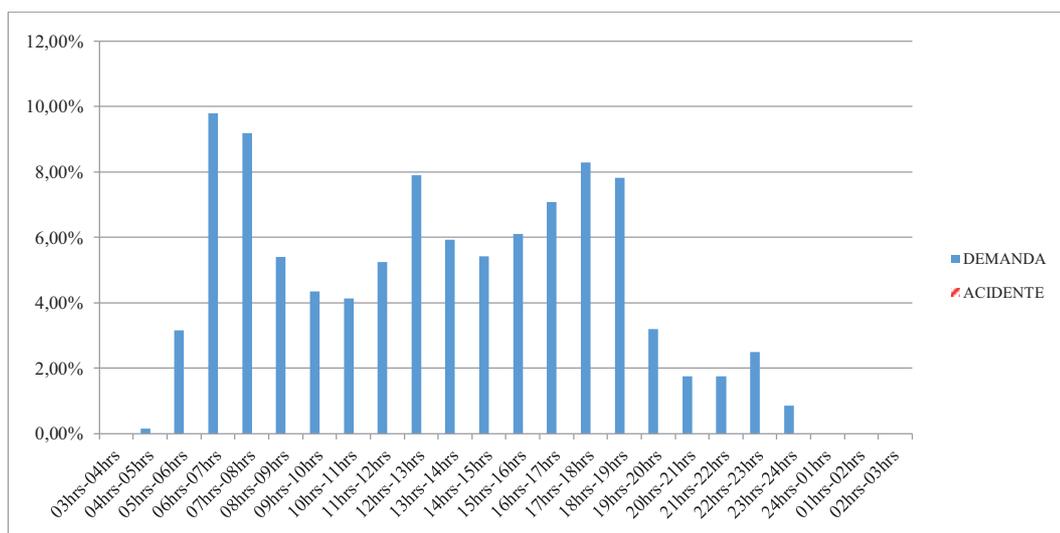
Os índices foram aplicados de acordo com a pesquisa de literatura, contudo estes são índices genéricos, no caso da utilização para o transporte público urbano, que a comparação está sendo realizada por linha de operação, deve ser considerada a frota máxima operante para cada linha, uma vez que quanto maior o número de veículos nas vias maior a probabilidade de acidente. De acordo com a Tabela 17 (Apêndice A) algumas linhas com a quantidade de quilômetros próximos tem uma frota máxima operante bem diferente, por exemplo, as linhas 12D e 40R com, aproximadamente, 33 mil km e 37 mil km, respectivamente, no entanto, cada uma das linhas tem uma frota máxima de operação (frota de pico) de 4 veículos e 7 veículos (12D e 40R, concomitantemente).

Semelhante ao exposto para o índice de quilometragem se aplica ao índice de horas operadas. A facilidade, no entanto, em obter os dados das linhas e o que cada veículo opera pode ajudar na equalização dos índices.

Os índices expostos, finalmente, apresentam a relação entre as variáveis (km, horas oper. e passageiros) com os acidentes, tornando possível a comparação entre as linhas e relacionar as com piores desempenhos, de acordo com os critérios expostos.

A demanda de passageiros, no entanto, há mais elementos para se analisar com relação aos acidentes. O Gráfico 20 contém o percentual de ocorrência de acidentes e o número de passageiros transportados, por faixa horária. Nota-se que o comportamento dos acidentes é igual ao da demanda, ocorrendo a maior concentração nos horários de pico.

Diante dos resultados do processamento estatístico que não apontaram a conversa do condutor com o passageiro como significativa, mesmo 64% do total das conversas ocorrerem nos horários de pico, considera-se que a presença do passageiro não é um fator determinante para a ocorrência do acidente.



**Gráfico 20** - Percentual de ocorrência de acidentes e de demanda de passageiros, por faixa horária

Em suma, temos mais duas opções para representar esse fenômeno: o volume de tráfego nesse horário também é intenso, provocando mais acidentes; ou a quantidade de paradas nos pontos de ônibus é maior para embarque e/ou desembarque de passageiros. Esta última foi abordada por Wåhlberg (1997) e concluiu que o número de paradas do ônibus é causa de acidentes. Embora este fator dependa bastante das características do sistema de transporte em operação, no caso deste estudo de caso os ônibus não são segregados e nem sempre têm baias para o veículo encostar durante o embarque e desembarque dos passageiros.

Quanto ao volume de tráfego que, também, ocorre no horário de pico é natural que os veículos disputem por espaço físico ou mesmo ocorram choques na troca de faixas. Os Gráficos 11 ao 15 apresentam os acidentes por faixa horárias, por tipo de linha. Independente

do tipo da linha, os acidentes ocorrem com maior frequência nos horários de pico, inclusive para as linhas perimetrais, que atendem áreas periféricas da cidade e que em tese são áreas menos movimentadas.

Em números absolutos a quantidade de acidentes por tipo de linha é maior nas linhas radiais, no entanto, a quantidade de linhas radiais é maior. Foi calculado o valor médio de acidentes por tipo de linha e os valores indicam que as circulares têm maior número de acidentes por linha, conforme Tabela 16.

Os valores médios de passageiros por tipo de linha também são apresentados na Tabela 16 e mostra que a maior concentração de passageiros está entre as linhas circulares e diametrais, assim como os acidentes.

Para a análise da operação da linha, quanto ao comportamento dos condutores, foi realizado o processamento estatístico dos dados, conforme descrito no item 5.3.2. O resultado para o agrupamento das linhas de operação foi Uso de Celular e Avanço de Sinal Amarelo, sendo esses comportamentos altamente significativos em relação aos acidentes. Os outros resultados do processamento (agrupamento por área e condutor) reafirmam os resultados obtidos por linha (conforme Tabela 15).

**Tabela 16** - Média de acidentes e passageiros por tipo de linha

<b>Tipo da linha</b>	<b>Quantidade de linhas</b>	<b>Quantidade de acidentes</b>	<b>Quantidade de passageiros</b>	<b>Média de acidentes</b>	<b>Média de passageiros</b>
Circular	4	45	221.210	11,25	55.302,50
Diametral	10	86	594.241	8,60	59.424,10
Radial	63	353	2.448.287	5,60	38.861,70
Perimetral	4	13	122.099	3,30	30.524,75
Noturna	8	3	4.000	0,38	500,00

Há também o processamento incluindo o número de condutores envolvidos na operação das linhas e que, inclusive, foi o resultado com menor valor de AIC. O principal motivo para que tenha gerado um novo cálculo é que não é uma variável de comportamento, mas um item de planejamento (escala de condutores). Portanto, a utilização de muitos condutores na operação de uma linha pode comprometer a segurança do sistema.

A partir da taxonomia criada para os comportamentos contidos no banco de dados e os resultados da análise estatística, podemos concluir que o comportamento decidido (tomado

pelo condutor) é definitivamente nocivo, uma vez que utilizar o celular e, muitas vezes, avançar o sinal é uma decisão tomada pelo condutor.

O avanço de sinal amarelo embora seja uma situação de atenção, não há exigência de parar, contudo o bom senso e a direção defensiva requer que o condutor pare o veículo, se for possível. Neste caso, em que se trata de um sistema de transporte público por ônibus, o tamanho deste veículo pode causar acidentes mais graves para com os outros agentes envolvidos (pedestres, ciclistas, etc.). Portanto é importante estabelecer critérios internos na empresa para o comportamento do condutor, quanto ao avanço de sinal amarelo.

O resultado estatístico mostra, principalmente, que há relação significativa entre o número de acidentes e o número de avanços de sinal amarelo, ou seja, a sucessiva ocorrência de avanços de sinal amarelo pode acabar em um acidente de trânsito.

Embora haja explicações, por parte dos condutores, da razão pela qual avançam o semáforo, não há dúvidas de que a decisão é sempre do condutor. Dirigir o veículo de maneira segura para os passageiros e para terceiros é uma responsabilidade que deve ser imposta aos condutores, acima de qualquer outra tarefa. Há exceções nos casos em que ocorrem lapsos de memória, contemplado na taxonomia da Figura 10, e produzem ações involuntárias. Esse tipo de comportamento é produto do estresse e fadiga, por exemplo, e condutores de ônibus são rotineiramente submetidos a esse tipo de ambiente.

A utilização de celular é infração de trânsito, não é permitido para qualquer tipo de condutor e o rigor da lei deveria ser mais bem aplicado à condutores profissionais. Neste caso a prática da intolerância deve ser cultivada pela empresa operadora e pela fiscalização de trânsito. Outros trabalhos, conforme apresentado na revisão de literatura, apontam o uso do celular como principal fator de causa de acidente, nem sempre diretamente (ocorrência de acidente no momento do uso), mas também indiretamente, em momentos antes ou depois da utilização de aparelhos celulares.

## **5.5 Sugestões**

Através dos resultados e das análises foi possível desenvolver algumas soluções, que seguem como sugestões na ordem do operacional (à empresa), de campanhas estaduais e melhoria nos cursos para aquisição de carteira de habilitação profissional (ao estado) e, finalmente, para trabalhos futuros.

### 5.5.1 À empresa

A utilização de celular durante a condução é um comportamento que deveria ser erradicado da empresa através de uma campanha onde contemple: informação, advertências e penalidades. Informar que não se trata de uma regra da empresa, mas sim de uma Lei de Trânsito. Advertir que embora a fiscalização estadual não seja presente, a empresa zela pelo cumprimento da Lei, pelo bom andamento do transporte público e pelo transporte seguro dos passageiros e demais indivíduos presentes no trânsito urbano. Penalizar após informar e advertir.

O sistema de análise das imagens de vídeo, atualmente em funcionamento na empresa, se assemelha ao esquema montado para o *Naturalistic Driving Study*, respeitando as diferenças tecnológicas dos dois. Contudo, é possível transformar o atual sistema, do estudo de caso, em um NDS simplificado voltado ao transporte público, uma vez que parte dos procedimentos de coleta de dados já é realizada. SHRP2 (2016) descreve como *Longitudinal Deceleration* e *Longitudinal Acceleration* as variáveis que representam se o veículo realizou uma desaceleração ou aceleração muito forte. As câmeras instaladas nos ônibus, ao menos na última visita técnica, tinham esse sistema de identificação, denominado pelo SHRP2 (2014) por Acelerômetro, e até então não haviam sido utilizados.

Conclusão, temos um potencial NDS, simplificado, em operação e que precisariam apenas de algumas diretrizes para ser utilizado.

Outras sugestões são listadas a seguir.

- a) Boa parte do esforço aplicado a este trabalho foi em tratar os dados, deste modo, padronizá-los para que estes estejam prontos para utilização é muito importante, os resultados são obtidos mais facilmente. Portanto, instituir uma padronização na geração dos dados é uma tarefa importante para se obter dados de qualidade e posteriormente respostas rápidas e objetivas;
- b) Programar 3 níveis de gestão das ocorrências, sendo curto, médio e longo prazo. Não é possível determinar padrões recorrentes do comportamento dos condutores se não realizar a análise de longo prazo destes dados. Ajuda a fiscalizar e acompanhar a evolução das ocorrências;
- c) Tornar a segurança viária parte do planejamento da operação do sistema, por exemplo, reduzir o número de condutores por linha, ou pelo menos não exigir excessivamente de condutores que não conhecem o itinerário de determinada linha. Deve haver um

comprometimento da empresa para o bem estar dos condutores. Lembrando que ações como essas elevam a qualidade do transporte público e reduzem custos;

- d) Determinar as características comportamentais dos condutores e identificar as deficiências individuais para que haja, inclusive, um planejamento em função do perfil comportamental de cada indivíduo;
- e) Investir na formação dos condutores, tornando a ocorrência de acidentes cada vez menor;
- f) Para o banco de dados apresentado é de extrema importância a constituição de definições dos eventos, especialmente no que diz respeito aos acidentes. É necessário excluir a redundância de informações, para que seja possível gerar dados mais precisos e objetivos. Este, portanto, é o objetivo de desenvolver a taxonomia dos acidentes, pois os dados de acidentes devem ter a maior quantidade de detalhes possível, e por outro lado não serem redundantes ou ambíguos.

Esta taxonomia dos acidentes é desenvolvida com base naquelas já apresentadas nas referências bibliográficas (Walberg, 2001 e HSRC, 2003), no entanto, compreendendo que o levantamento dos dados é realizado através de vídeos.

A construção das categorias de acidentes, quanto à gravidade do evento e quanto à característica do local onde o mesmo ocorreu, vem de encontro, principalmente, para organizar a geração dos dados. Embora exista no banco de dados uma “categoria” dos acidentes, estes não são bem delimitados, como por exemplo, ‘Colisão’ e ‘Acidente’. Não há uma definição formal da diferença destes eventos.

Na Figura 11 é apresentada a categorização dos acidentes levando em consideração a gravidade e o ambiente em que ocorre (externo e interno).

Inicialmente foram categorizados em 3 partes, Com vítima, Sem vítima e No interior do ônibus, para os dois primeiros são considerados apenas acidentes ocorridos no ambiente externo do ônibus e, finalmente, o terceiro diz respeito aos decorrentes no interior do ônibus.

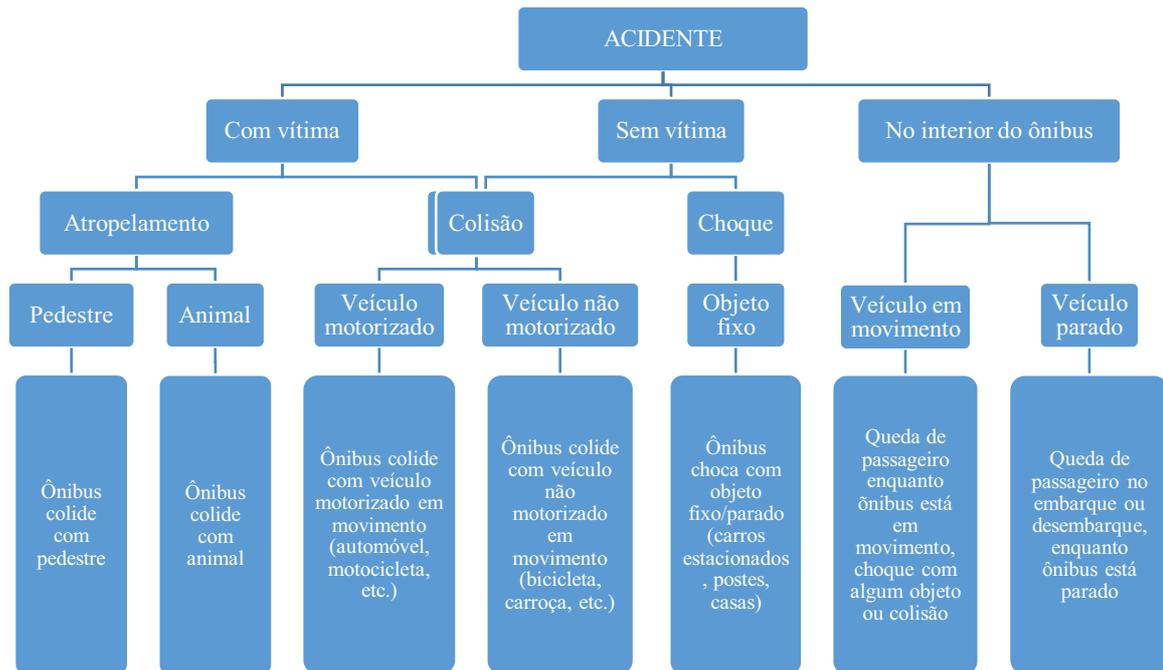


Figura 11 - Taxonomia dos acidentes, quanto à gravidade.

*Com vítima:* Acidentes em que há danos humanos;

*Sem vítima:* Acidentes em que há apenas danos materiais, avarias.

A Figura 12 apresenta algumas definições quanto ao local em que o acidente ocorreu. Este foi baseado no que é possível identificar através dos vídeos gravados. A intenção é que a equipe que gera os dados dos vídeos (conforme apresentado no item 3) seja capaz de captar mais detalhes dos acidentes.

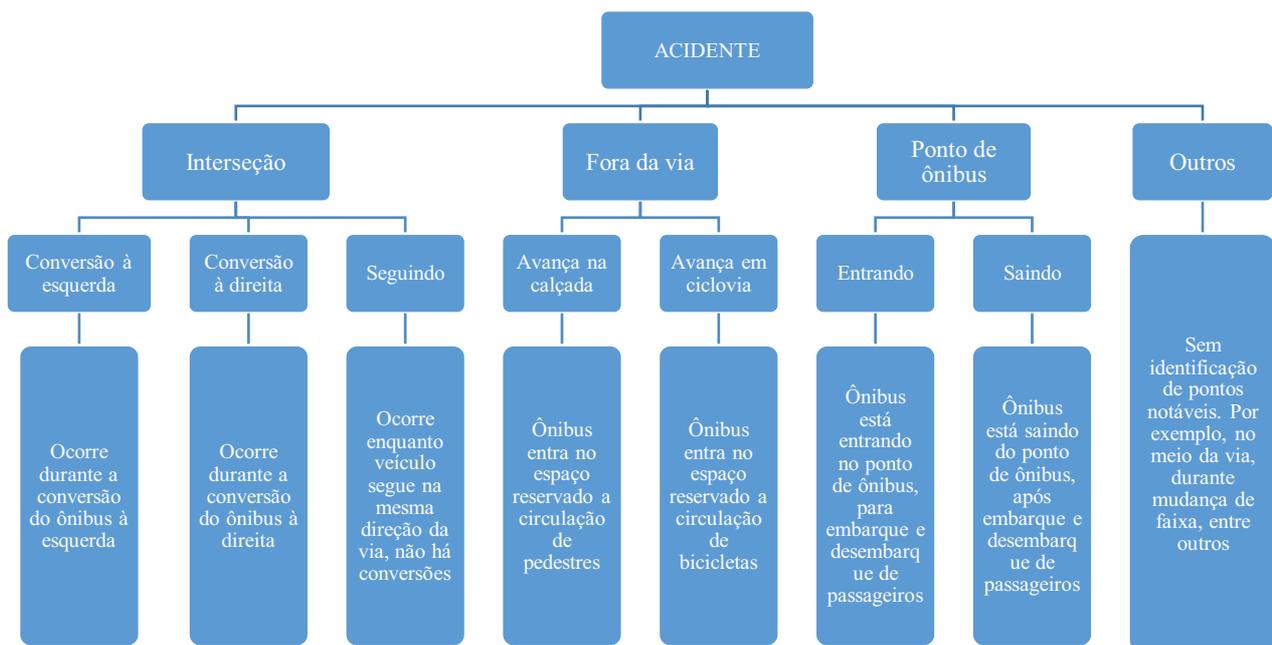


Figura 12 - Taxonomia dos acidentes, quanto à característica do local.

A sugestão das taxonomias, apresentadas nas Figuras 11 e 12, tem como objetivo estruturar os dados gerados a partir do sistema de monitoramento por câmeras, portanto, as taxonomias foram construídas permitindo a categorização apenas assistindo aos vídeos, não necessitando de qualquer outra fonte de informações.

### 5.5.2 Planilha eletrônica para controle/cadastro dos acidentes

Alguns documentos e arquivos foram encaminhados à empresa operadora do transporte público, na tentativa de propor uma coleta de dados mais precisa, quanto aos acidentes. Um destes arquivos foi uma planilha eletrônica automatizada com elementos pré-definidos (através do *Visual Basic Application*), conforme Figura 13 e 14. O objetivo desta planilha eletrônica era coletar os dados dos acidentes, de maneira organizada, para então cruzar, também, com os dados do comportamento dos condutores.

Além da planilha eletrônica, mais dois arquivos digitais foram disponibilizados para definir as variáveis constantes na planilha, ou seja, para orientar o preenchimento da mesma.

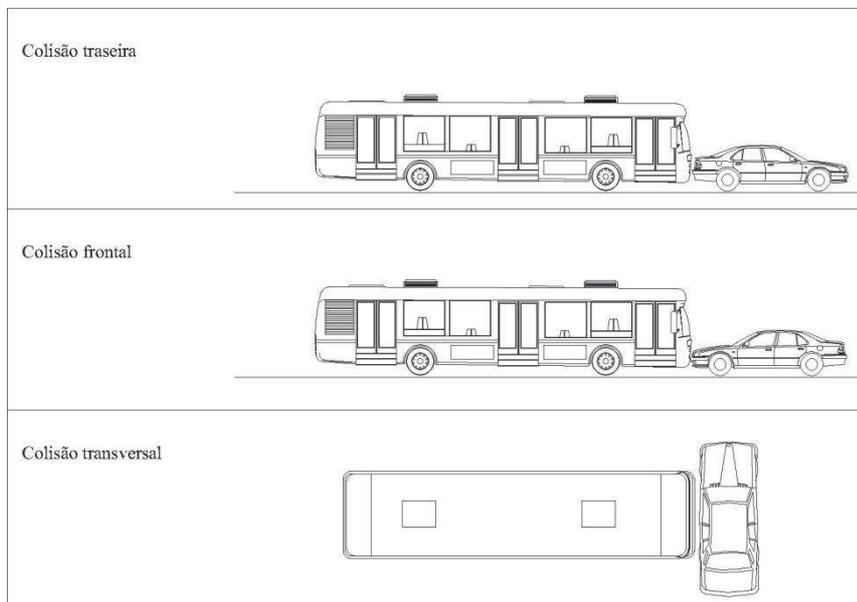
The image shows a screenshot of an Excel spreadsheet with a dialog box titled "Inclusão Cliente" overlaid on it. The spreadsheet has a header row with the following columns: Código, Data da ocorrência, Hora da ocorrência, Nº do carro, Linha, Tipo de acidente, Possível causa, Possível culpado, Tipo do veículo envolvido, Motorista, Situação da escala de trabalho do motorista, and Tempo de empresa. The dialog box contains the following fields and controls:

- Código: 1
- Data da ocorrência: 21/06/2016
- Hora da ocorrência: [text box]
- Nº do Carro: [text box]
- Linha: [dropdown menu]
- Tipo acidentes: [dropdown menu]
- Avaria: [dropdown menu]
- Possível causa: [text box]
- Possível culpado: [dropdown menu]
- Tipo do veículo envolvido: [dropdown menu]
- Motorista: [text box]
- Situação da escala do Motorista: [dropdown menu]
- Tempo de empresa: [dropdown menu]
- Buttons: Gravar, Cancelar

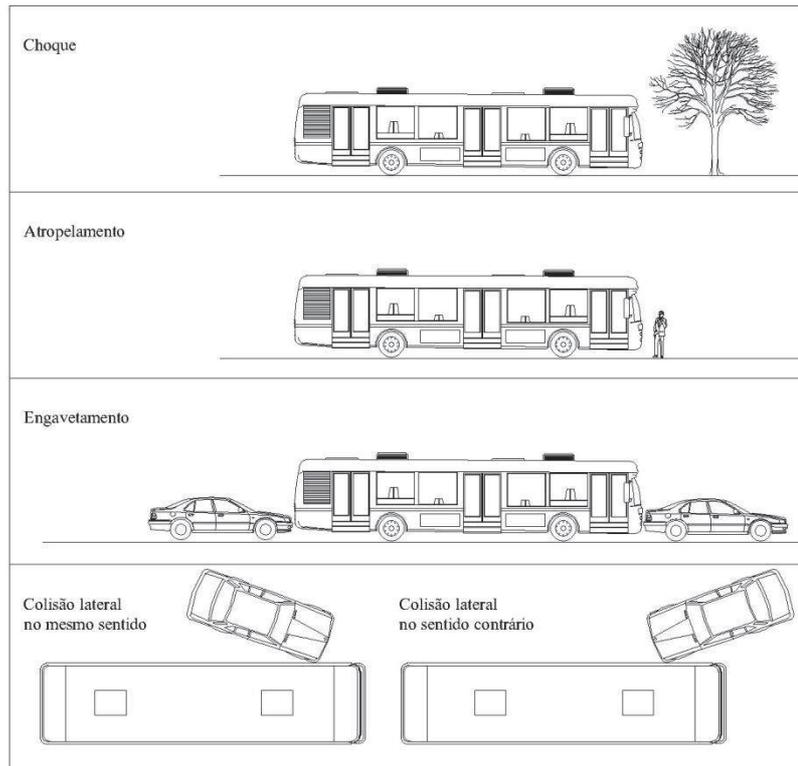
**Figura 13** - Planilha eletrônica para cadastro dos acidentes ocorridos

**Figura 14** - Planilha eletrônica para cadastro dos acidentes ocorridos, exemplo

O relatório, para orientar o preenchimento da planilha, apresenta, principalmente, as definições de cada uma das possíveis acidentes de trânsito, conforme Figura 15 e 16.



**Figura 15** - Relatório com exemplos dos tipos de acidentes de trânsito



**Figura 16** - Relatório com exemplos dos tipos de acidentes de trânsito, continuação

Além das imagens que exemplificam as ocorrências, também foram definidas em texto.

### ***Colisão traseira***

*Exemplo: Veículo colide na parte traseira do ônibus (ou inverso).*

### ***Colisão frontal***

*Exemplo: Veículo colide na parte dianteira do ônibus (ou inverso).*

### ***Colisão Transversal***

*Exemplo: Ônibus avança sinal vermelho e veículo colide diretamente na lateral do ônibus (ou inverso).*

### ***Choque***

*Exemplo: Colisão entre o ônibus com qualquer objeto fixo (árvore, poste, residência, etc.).*

### ***Atropelamento***

*Exemplo: Ônibus colide contra pessoas ou animais.*

### ***Engavetamento***

*Exemplo: Colisão frontal e traseiro.*

***Colisão lateral no mesmo sentido***

*Exemplo: Ônibus saindo do ponto de embarque e outro veículo colide na lateral (ou inverso).*

***Colisão lateral no sentido contrário***

*Exemplo: Em uma via simples de mão dupla o veículo colide na lateral do ônibus.*

**5.5.3 Ao estado**

Este trabalho apresenta as deficiências no comportamento do condutor de transporte público, especificamente quanto ao local do estudo de caso. Pontos particulares devem ser levados em conta pelo Departamento Estadual de Trânsito de São Paulo, para que sejam abordados estes temas de maneira mais eficiente nos cursos de formação de condutores profissionais. Ou mesmo na reciclagem desses profissionais, tendo como objetivo principal a redução de comportamentos que aumentam o risco de acidentes.

**5.5.4 Trabalhos futuros**

- a) Não foram considerados fatores espaciais, embora os dados agrupados por regiões tenham sido processados. Seria interessante conhecer a localização exata desses acidentes para identificar problemas com a via, com sinalização ou com determinados cruzamentos. Embora inicialmente este trabalho visasse a análise espacial dos acidentes, através do cruzamento dos dados de acidentes com os dados do GPS embarcado, não foi possível, pois a diferença de horário entre os dados (um dos campos responsável pela integração) interferia demais na localização dos acidentes. Portanto, para a realização deste estudo seria necessário utilizar outro método;
- b) Acompanhamento das ocorrências após a implantação de campanhas com os condutores, com o objetivo de dimensionar a real relação do comportamento dos condutores com os acidentes;
- c) Com a finalização da construção das vias segregadas do sistema de Transporte Público Urbano em estudo, avaliar as mudanças ocorridas nos itinerários das linhas e quais as melhoras, quanto aos acidentes antes e depois da implantação;
- d) Outros trabalhos utilizam os dados de acidentes em que a culpa do condutor do ônibus é atestada, não havendo variação na culpa dos acidentes. Isso implicaria em maior certeza nos resultados, já que não existe a possibilidade de atribuir o erro do acidente

no outro envolvido. Portanto, sugere-se que sejam submetidos aos mesmos cálculos apenas os dados de acidentes em que os condutores do ônibus são de fato responsáveis;

- e) Os veículos do transporte público ficam em constante operação, contudo, será que existe variação temporal no número de acidentes? A análise temporal destes acidentes pode determinar se os eventos são flutuantes ou não;
- f) Desenvolver um método de contagem ou análise do tráfego a jusante do veículo, uma vez que os vídeos gravados por uma das câmeras mostra exatamente a situação da via à frente do ônibus. Considerando que existem vários ônibus e que estes passam pelos mesmos pontos várias vezes por dia é possível extrair informações a respeito do tráfego dos veículos em horários diferentes, além da possibilidade de verificar qual o peso desse fenômeno na ocorrência de acidentes, será gerado dados de trânsito, para outros fins;
- g) Utilização dos vídeos para contagem de passageiros, lotação ao longo da linha em movimento ou pesquisa “Sobe-Desce”, através das imagens que mostram o interior do veículo (porta de entrada, corredor e porta de saída);



## 6 CONCLUSÕES

---

O sistema de transporte público, objeto do estudo de caso para este trabalho, foi analisado através dos dados de operação e dos dados gerados pelo sistema de monitoramento por câmeras a fim de estabelecer uma relação dessas variáveis com os acidentes ocorridos, no período em estudo.

O objetivo foi atingido, uma vez que os resultados levantaram os pontos mais importantes com relação aos acidentes, permitindo direcionar a empresa no planejamento de campanhas de redução de acidentes ou mesmo no planejamento da operação e fiscalização dos condutores.

Concluiu-se, primeiramente, que a utilização do celular é a variável mais significativa em detrimento aos acidentes e que, também, há um alto índice de utilização em horários de pico (53% do total de eventos de uso de celular), sendo estes horários de maior frequência de passageiros e de veículos. Outro ponto apresentado é que 33% das vezes utilizaram o telefone mais que 5 minutos.

A segunda variável significativa, avanço de sinal amarelo, é difícil definir qual a origem do comportamento, decidido ou involuntário, de qualquer maneira boa parte dos eventos devem ocorrer por meio de decisão do condutor. No entanto, trata-se de uma situação mais complexa, todavia, deve-se atentar para o planejamento da operação considerando que o ato de avançar o sinal amarelo pode ocorrer por atraso no horário de operação, consequência de intervalos pequenos entre as viagens. A questão central é que independente das circunstâncias é necessário treinar o condutor a dirigir sempre a favor da segurança, de maneira responsável.

Houve também o resultado obtido a partir do processamento estatístico com a variável do número de condutores envolvidos na operação da linha. Este resultado em particular não diz respeito ao comportamento dos condutores, mas sim ao planejamento da operação. Este resultado, inclusive, teve melhor estimativa que os outros modelos, através do valor de AIC. Com estas informações, é possível concluir que o fator de planejamento da operação é mais significativo, em prejuízo aos acidentes, uma vez que escalar muitos condutores em uma linha (muitas vezes não conhecem o itinerário, ou não são acostumados com o horário da operação) mostrou ser mais prejudicial que o próprio mau comportamento dos condutores. Em contrapartida, o resultado dessa contagem dos condutores se deve ao fato do condutor ter se

envolvido em alguma das ocorrências constantes no banco de dados, por tanto, algumas linhas podem ter o número de condutores baixo não necessariamente por existir poucos, mas que poucos tiveram mau comportamento. Partindo dessa última premissa, podemos concluir, também, que talvez haja uma concentração de condutores com comportamento ruim em determinadas linhas.

A discussão quanto às linhas com piores índices de acidentes também foram apresentadas e a conclusão, ainda que os índices sejam para fins comparativos entre as linhas do sistema de transporte público, devem ser objeto de estudo mais aprofundado, uma vez que sugere claramente que as linhas mais expostas aos acidentes não são as que têm maior número de acidentes. Entretanto, os índices devem ser sempre atualizados com o objetivo de acompanhar as linhas após mudanças no comportamento dos condutores, por exemplo, e/ou identificar algum erro de planejamento nas mesmas (tempo de ciclo insuficiente, escala de condutores, problemas viários, etc.).

Os resultados discutidos, no capítulo anterior, quanto aos acidentes por faixa horária trouxe alguns questionamentos: quantidade de passageiros, alto volume de tráfego e maior quantidade de parada dos veículos. De acordo com os resultados da análise estatística, a variável ‘Conversa durante condução’ não teve significância alta com os acidentes, assim sendo, conclui-se que a presença de passageiros conversando com o condutor não tenha relação com os acidentes. Neste caso, temos como prováveis causas o volume de tráfego e a quantidade de paradas dos ônibus, no entanto, ambos têm como solução compartilhada a segregação do transporte público por ônibus. Conclui-se, portanto, que após separar os ônibus de transporte público dos veículos particulares haja uma redução dos acidentes.

O desenvolvimento deste trabalho garante a sua replicação, ou seja, é perfeitamente possível utilizar os procedimentos apresentados aqui em outro sistema de transporte público, desde que os dados estejam disponíveis.

Este trabalho se faz importante não só para a empresa, no sentido de identificar as características dos acidentes, mas também para o poder público, precisamente o Departamento Estadual de Trânsito de São Paulo, pois permite avaliar os problemas atuais com os condutores profissionais, uma vez que este departamento é o responsável pela emissão da Carteira Nacional de Habilitação.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

---

BRENAC, T.; CLABAUX, N. The indirect involvement of buses in traffic accident processes. **Safety Science**, 2 Fevereiro 2005. 835-843.

CHIMBA, D.; SANDO, T.; KWIGIZILE, V. Effect of bus size and operation to crash occurrences. **Accident Analysis and Prevention**, 01 Março 2010. 2063-2067.

FERRAZ, A. C. P.". et al. **Segurança Viária**. São Carlos: Suprema Gráfica e Editora, 2012.

GOH, K. et al. Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. **Accident Analysis and Prevention**, 25 dez. 2013. 20-26.

HAIR, J. F. et al. **Análise Multivariada de Dados**. 5ª. ed. São Paulo: Bookman, 2006.

HIGHWAY SAFETY RESEARCH CENTER. **Distractions in everyday driving**. University of North Carolina. Chapel Hill, p. 105. 2003.

HO, R. **Handbook of Univariate and Multivariate Data Analysis with IBM SPSS**. Boca Raton, FL: CRC Press, 2014.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEA, ASSOCIAÇÃO NACIONAL DE TRANSPORTES PÚBLICOS - ANTP. **Impactos sociais e econômicos dos acidentes de trânsito nas aglomerações urbanas brasileiras**. Brasília. 2003.

KUHA, J. AIC and BIC: Comparisons of Assumptions and Performance. **Sociological Methods Research**, Novembro 2004. 188–229.

MALLIA, L. et al. Crash risk and aberrant driving behaviors among bus drivers: The role of personality and attitudes towards traffic safety. **Accident Analysis and Prevention**, 28 Março 2015. 145-151.

RAMOS, M. W.; OLIVEIRA, L. K. D. Preferência dos usuários de ônibus de Belo Horizonte em relação à conduta dos motoristas. **Revista dos Transportes Públicos - ANTP**, São Paulo, v. Ano 34, n. 3º quadrimestre, p. 39-49, 2011.

REASON, J. **Human Error**. Nova York: Cambridge University Press, 1990. 302 p.

SHRP 2. **The SHRP 2 Naturalistic Driving Study**. Transportation Research Board of the National Academies. Washington, p. 8. 2014.

SHRP 2. **Description of the SHRP2 naturalistic database and the crash, near-crash, and baseline data sets**. Transportation Research Board of The National Academies. Blacksburg, p. 58. 2016.

STUTTS, J. et al. **Distractions in Everyday Driving**. University of North Carolina, AAA Foundation for Traffic Safety. Washington , D.C., p. 105. 2003.

TSE, J. L. M.; FLIN, R.; MEARNNS, K. Bus driver well-being review: 50 years of research. **Transportation Research Part F 9**, p. 89-114, Março 2006.

U.S. DEPARTMENT OF TRANSPORTATION. **Tri-level study of causes of traffic accidents**. National Highway Traffic Safety Administration. Washington, D.C., p. 82. 1979. (DOT HS-805 099).

WÅHLBERG, A. A. et al. Development and validation of a self-report measure of bus driver behaviour measure of bus driver behaviour. **Ergonomics**, Bedfordshire, v. 53, p. 1420-1433, 2010.

WÅHLBERG, A. E. A. **Time pressure, age and driving speed among bus drivers: A pilot study**. Uppsala, Sweden: Uppsala University. 1997.

WÅHLBERG, A. E. A. Characteristics of low speed accidents with buses in public transport. **Accident Analysis and Prevention**, 1 Maio 2001. 637–647.

WÅHLBERG, A. E. A. Characteristics of low speed accidents with buses in public transport: part II. **Accident Analysis and Prevention**, 17 Outubro 2004. 63-71.

WEST, R. **Accident Script Analysis**. Transport Research Laboratory. Crowthorne, p. 25. 1997.

## A. APÊNDICE – INFORMAÇÕES DAS LINHAS DE OPERAÇÃO

**Tabela 17 -** Informações das linhas do sistema de transporte público

Cód. Linha	Acidentes	Área da linha	Tipo da linha	Quilômetros/Mês	Horas Oper./Mês	Demanda total/Mês	$I_k$	$I_h$	$I_p$	Frota Máxima
10R	6	B	RADIAL	10692,00	493,50	9629	0,360	0,779	0,399	1
11R	3	G	RADIAL	15703,50	899,02	38609	0,122	0,214	0,050	3
20R	7	A	RADIAL	31125,00	1383,25	34199	0,144	0,324	0,131	3
21R	5	C	RADIAL	20760,96	866,82	19096	0,154	0,370	0,168	2
22R	2	G	RADIAL	14058,00	808,25	39254	0,091	0,159	0,033	3
23R	0	I	RADIAL	7522,20	394,00	18543	0,000	0,000	0,000	1
24R	9	K	RADIAL	25443,90	1522,92	76246	0,227	0,379	0,076	5
25R	7	L	RADIAL	20945,10	2126,42	48053	0,214	0,211	0,093	3
26R	10	N	RADIAL	23685,60	1698,30	73128	0,271	0,377	0,088	4
27R	5	R	RADIAL	28832,80	1624,65	42602	0,111	0,197	0,075	4
28R	5	A	RADIAL	32332,00	1484,60	39375	0,099	0,216	0,081	3
29R	0	A	RADIAL	3351,60	121,45	9277	0,000	0,000	0,000	2
2R	4	B	RADIAL	31526,72	1651,42	42745	0,081	0,155	0,060	4
31R	4	C	RADIAL	19580,80	818,23	16558	0,131	0,313	0,155	2
35R	6	A	RADIAL	31944,00	1466,78	38917	0,120	0,262	0,099	3
36R	3	D	RADIAL	18335,90	1039,88	32837	0,105	0,185	0,059	5
37R	1	E	RADIAL	14393,56	781,20	22040	0,045	0,082	0,029	2
38R	8	L	RADIAL	19261,80	1088,42	37163	0,266	0,471	0,138	3
39R	2	N	RADIAL	8034,00	356,25	14754	0,160	0,360	0,087	3
3R	3	C	RADIAL	19060,80	1911,03	35754	0,101	0,101	0,054	3
40R	20	R	RADIAL	33493,20	2044,67	58392	0,383	0,627	0,220	4
41R	5	A	RADIAL	29452,50	1537,25	34550	0,109	0,208	0,093	4
42R	0	A	RADIAL	4067,70	112,70	8223	0,000	0,000	0,000	2
44R	7	D	RADIAL	19116,50	1035,82	39920	0,235	0,433	0,112	2
45R	0	E	RADIAL	12915,34	796,20	21252	0,000	0,000	0,000	2
48R	8	L	RADIAL	18172,00	1080,17	39038	0,282	0,475	0,131	3
49R	2	N	RADIAL	9891,00	194,75	14172	0,130	0,658	0,090	1
4R	2	G	RADIAL	16435,75	1137,90	50307	0,078	0,113	0,025	6
51R	12	R	RADIAL	33732,20	2027,87	57475	0,228	0,379	0,134	8
52R	2	R	RADIAL	1638,00	46,20	3658	0,783	2,775	0,350	1
55R	2	A	RADIAL	17150,00	925,58	25077	0,075	0,139	0,051	2
56R	14	D	RADIAL	48872,16	2480,92	86718	0,184	0,362	0,103	9
57R	2	H	RADIAL	20666,40	885,03	46462	0,062	0,145	0,028	3
58R	7	M	RADIAL	28039,30	1352,58	65486	0,160	0,332	0,069	6
59R	7	O	RADIAL	28248,60	1670,05	52869	0,159	0,269	0,085	5
5R	5	I	RADIAL	16964,04	1002,85	61649	0,189	0,320	0,052	3
60R	5	C	RADIAL	17442,00	1110,00	24643	0,184	0,289	0,130	4
61R	13	D	RADIAL	50782,64	2413,57	77012	0,164	0,345	0,108	9
63R	21	M	RADIAL	27128,60	1858,58	64575	0,496	0,724	0,208	3
64R	9	G	RADIAL	29645,16	1670,05	48493	0,195	0,345	0,119	5
67R	14	A	RADIAL	53235,00	2229,75	88034	0,169	0,402	0,102	10
68R	1	H	RADIAL	19759,52	726,37	22493	0,032	0,088	0,028	2
69R	7	M	RADIAL	10832,40	559,50	30561	0,414	0,802	0,147	1
6R	3	K	RADIAL	9975,00	559,00	18878	0,193	0,344	0,102	2
70R	2	C	RADIAL	20453,00	918,58	21583	0,063	0,140	0,059	2
71R	3	S	RADIAL	18468,00	1297,63	28305	0,104	0,148	0,068	4
72R	4	S	RADIAL	16474,70	1099,75	30331	0,156	0,233	0,085	3
74R	1	B	RADIAL	26993,96	678,92	10470	0,024	0,094	0,061	2

76R	2	F	RADIAL	24066,40	1385,00	57415	0,053	0,093	0,022	4
77R	0	M	RADIAL	12622,50	419,25	7462	0,000	0,000	0,000	1
78R	2	P	RADIAL	30842,66	1333,12	43194	0,042	0,096	0,030	3
79R	9	T	RADIAL	28973,22	1853,92	56267	0,199	0,311	0,103	4
7R	2	L	RADIAL	13532,45	885,63	30632	0,095	0,145	0,042	2
80R	5	F	RADIAL	29160,00	1301,37	54333	0,110	0,246	0,059	5
81R	6	C	RADIAL	25631,60	1293,22	44002	0,150	0,297	0,087	3
82R	14	T	RADIAL	28327,50	1683,88	32073	0,317	0,533	0,280	4
83R	0	T	RADIAL	3931,20	103,95	8374	0,000	0,000	0,000	2
85R	6	F	RADIAL	34510,00	1436,42	60560	0,111	0,268	0,064	8
86R	7	T	RADIAL	39164,40	1885,40	40654	0,115	0,238	0,110	4
87R	4	F	RADIAL	38163,30	1473,95	57401	0,067	0,174	0,045	5
88R	4	T	RADIAL	5319,30	147,70	11766	0,482	1,736	0,218	3
8R	4	O	RADIAL	11677,50	755,55	25971	0,220	0,339	0,099	2
9R	20	R	RADIAL	40904,40	2488,00	98778	0,313	0,515	0,130	8
15P	3	V	PERIMETRAL	24273,60	1068,25	61152	0,079	0,180	0,031	8
16P	0	V	PERIMETRAL	15594,00	631,00	11361	0,000	0,000	0,000	2
17P	2	V	PERIMETRAL	11880,00	648,50	9717	0,108	0,198	0,132	1
30P	8	V	PERIMETRAL	44805,60	2004,00	39869	0,114	0,256	0,129	6
19N	0	D	NOTURNA	2763,00	124,00	364	0,000	0,000	0,000	1
1N	0	A	NOTURNA	2025,00	124,00	201	0,000	0,000	0,000	1
34N	0	F	NOTURNA	2943,00	124,00	927	0,000	0,000	0,000	1
54N	0	I	NOTURNA	2151,00	124,00	53	0,000	0,000	0,000	1
66N	2	J	NOTURNA	2637,00	124,00	1564	0,486	1,034	0,820	1
73N	0	M	NOTURNA	1530,00	124,00	285	0,000	0,000	0,000	1
84N	1	Q	NOTURNA	2088,00	124,00	256	0,307	0,517	2,504	1
89N	0	T	NOTURNA	1944,00	124,00	350	0,000	0,000	0,000	1
12D	9	E	DIAMETRAL	37810,20	2175,92	64125	0,153	0,265	0,090	7
13D	7	I	DIAMETRAL	20573,90	1505,42	38706	0,218	0,298	0,116	4
14D	6	I	DIAMETRAL	23092,80	1413,53	59103	0,167	0,272	0,065	4
32D	6	H	DIAMETRAL	44391,60	2474,08	79937	0,087	0,155	0,048	9
43D	17	J	DIAMETRAL	50667,39	2427,00	84128	0,215	0,449	0,130	10
46D	6	F	DIAMETRAL	45433,30	1876,72	76064	0,085	0,205	0,051	10
47D	18	B	DIAMETRAL	49946,73	2677,90	81913	0,231	0,431	0,141	10
50D	7	F	DIAMETRAL	46578,90	2526,45	79959	0,096	0,178	0,056	8
62D	5	E	DIAMETRAL	19126,80	1153,50	23772	0,168	0,278	0,135	3
75D	5	D	DIAMETRAL	8892,00	554,50	6534	0,360	0,578	0,491	1
18C	12	U	CIRCULAR	57020,60	2987,08	81658	0,135	0,258	0,094	8
33C	15	U	CIRCULAR	55297,60	3027,97	82343	0,174	0,318	0,117	8
53C	7	U	CIRCULAR	23753,99	1128,13	27773	0,189	0,398	0,162	7
65C	11	U	CIRCULAR	25321,00	1187,00	29436	0,278	0,594	0,240	5

## B. APÊNDICE – RESULTADOS DO PROCESSAMENTO ESTATÍSTICO

---

Abaixo os resultados do processamento estatístico, apresentado pelo *software* R. Dados Utilizando Número De Motoristas

```
> summary(mNB <- glm.nb(ACIDENTES ~ N_MOTORISTA + AVANCO_SINAL_VERMELHO, data = df))
```

Call:

```
glm.nb(formula = ACIDENTES ~ N_MOTORISTA + AVANCO_SINAL_VERMELHO,
       data = df, init.theta = 6.536546634, link = log)
```

Deviance Residuals:

```
   Min      1Q  Median      3Q      Max
-2.2714 -0.9231 -0.2703  0.5496  2.4640
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.347621  0.149103  2.331 0.019731 *
N_MOTORISTA     0.030685  0.002908 10.553 < 2e-16 ***
AVANCO_SINAL_VERMELHO -0.004170  0.001210 -3.447 0.000566 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.5365) family taken to be 1)

Null deviance: 230.53 on 88 degrees of freedom  
Residual deviance: 106.88 on 86 degrees of freedom  
AIC: 430.36

Number of Fisher Scoring iterations: 1

```
      Theta: 6.54
    Std. Err.: 2.37
```

2 x log-likelihood: -422.362

+++++

```
> summary(mNB <- glm.nb(ACIDENTES ~ N_MOTORISTA , data = df))
```

Call:

```
glm.nb(formula = ACIDENTES ~ N_MOTORISTA, data = df, init.theta = 4.870435343,
       link = log)
```

Deviance Residuals:

```
   Min      1Q  Median      3Q      Max
-2.1471 -1.1237 -0.1281  0.5840  2.3477
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.314342  0.156759  2.005  0.0449 *
N_MOTORISTA  0.025557  0.002544 10.046 < 2e-16 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(4.8704) family taken to be 1)

Null deviance: 204.67 on 88 degrees of freedom  
 Residual deviance: 105.88 on 87 degrees of freedom  
 AIC: 438.49

Number of Fisher Scoring iterations: 1

Theta: 4.87  
 Std. Err.: 1.53

2 x log-likelihood: -432.489

+++++

summary(mNB\_COMP <- glm.nb(ACIDENTES ~ AVANCO\_SINAL\_AMARELO + DIRIGINDO\_FALANDO\_CELULAR, data = df))

Call:  
 glm.nb(formula = ACIDENTES ~ AVANCO\_SINAL\_AMARELO + DIRIGINDO\_FALANDO\_CELULAR, data = df, init.theta = 3.050125137, link = log)

Deviance Residuals:  
 Min 1Q Median 3Q Max  
 -2.2069 -1.0185 -0.2647 0.5149 2.2020

Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
 (Intercept) 0.809451 0.133216 6.076 1.23e-09 \*\*\*  
 AVANCO\_SINAL\_AMARELO 0.005928 0.001822 3.253 0.00114 \*\*  
 DIRIGINDO\_FALANDO\_CELULAR 0.007796 0.001430 5.452 4.99e-08 \*\*\*

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(3.0501) family taken to be 1)

Null deviance: 164.06 on 88 degrees of freedom  
 Residual deviance: 105.88 on 86 degrees of freedom  
 AIC: 461.62

Number of Fisher Scoring iterations: 1

Theta: 3.050  
 Std. Err.: 0.795

2 x log-likelihood: -453.624

+++++

**C. APÊNDICE – RELATÓRIO DE PADRONIZAÇÃO DE  
ATUALIZAÇÃO E PADRONIZAÇÃO DOS DADOS DE IMAGENS**

---

**RELATÓRIO I -  
ATUALIZAÇÃO E PADRONIZAÇÃO DOS DADOS DE IMAGENS**

RIBEIRÃO PRETO

NOV/2014

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>87</b>
1.1 Objetivo.....	87
1.2 Justificativa.....	87
<b>2. APRESENTAÇÃO DOS DADOS CONTÍDOS NO BANCO DE DADOS.....</b>	<b>88</b>
<b>3. PADRONIZAÇÃO E NOMENCLATURAS .....</b>	<b>92</b>
<b>4. DEMAIS PADRONIZAÇÕES.....</b>	<b>93</b>

## INTRODUÇÃO

---

A padronização de informações tem um papel importante para evitar retrabalhos e agilizar a análise de dados.

### **Objetivo**

Apresentar os campos que são passíveis de modificações, assim como propor melhorias para os mesmos.

### **Justificativa**

Algumas dificuldades foram encontradas na manipulação dos dados, diante desse problema é de interesse apresentar todas as deficiências do banco de dados para conhecimento de todos e aprimorar as informações com vista à trabalhabilidade dos dados.

## APRESENTAÇÃO DOS DADOS CONTÍDOS NO BANCO DE DADOS

### Nomenclatura das linhas

A totalidade de linhas apresentadas no banco de dados é de 205, sendo o total de linhas do sistema de transporte 94.

A existência de duplicidade de dados e nomenclaturas erradas elevam os números de linhas, assim como, impossibilitam a análise imediata dos dados.

A Tabela 1 apresenta a lista das linhas que compõem o banco de dados.

TABELA 1 – COMPOSIÇÃO DAS LINHAS NO BANCO DE DADOS

CÓD - NOME		
001 - quintino corujao	136- castelo branco/adão do carmo	203-ribeirania
08 - noturno noroeste	136-castelo branco-adao do carmo	204 - jardim joao rossi
101 - quintino facci i	137 - lagoinha / ipiranga	204-city ribeirão
101- parque avelino	147 - santa cruz / ipiranga	205 - vila virginia
101- pq avelino	147- jd. Irajá- monte alegre	205 jd.joão rossi
101-parque avelino	147- jd.iraja monte alegre	205- joão rossi
102- jd.independencia	147-jd.iraja-monte alegre	205-jd.joao rossi
102-jd independencia	148- santa cruz ipiranga	206 - hospital das clinicas
103 - forum	148-santa cruz-ipuranga	206- v.virginia
103- igutemi	156- parque ribeirão/shopping	206-v.virginia
103-iguatemi	156- pq. Ribeirão shopping	207- hosp. Das clinicas
104 - shopping / via joao rossi	170 - planalto verde	207- hospital das clinicas
104- jd.canada	178- dom miele-hc	207-hospital das clinicas
104-jd. Canada	178-dom mielle /hc	208- vila albertina
105- rib shopping	187- heitor rigo/ hc	208-v. Albertina
105-ribeirao shopping	187- heitor rigon / hc	210- simioni
106- d euboux	199- circular 1	210-simione
106-d elboux	199- circular i	211-expresso simione
107- sumarezinho	199-circular 01	217 quintino hc
107-sumarezinho	199-circular i	217-quintino/hc
108- jd.presidente dutra	1-noturno norte	220- parque de exposições
108-jd.presidente dutra	201 - quintino facci ii	220-parque de exposições
110- quintino i	201 quintini ii	236- são jose adão do carmo
110-quintino i	201-quintino ii	236-sao jose-adao do carmo
114 - higienopolis	202 - jardim independencia	270 - eugenio lopes
130- forum	202- jardim iara	299- circular ii
130-forum	202- jd.iara	299-circular 02
135 - adao carmo / castelo branco	202-jd.iara	299-circular ii
136 - castelo branco / vila amelia	203 - iguatemi	2-noturno nordeste
	203 ribeirania	301 - avelino palma
		301- avelino palma
		301-avelino palma

302- jardim aeroporto	373-v. Abranches ipiranga	508-jd.procopio
302 jd aeroporto	373-v. Abranches-ipuranga	509 - parque ribeirao / shopping
302-jardim aeroporto	380 geraldo carvalho	580-jd.alexandre balbo
303 - ribeirania	380-geraldo de carvalho	5-noturno sul
303 bom pastor	388 expresso via norte	601-distrito empresarial
303-bom pastor	388- expresso via norte	602- aeroporto
305 - parque ribeirao preto	388-expresso via norte	602-aeroporto
305- jardim nova aliança	399- circular iii	603-jd.juliana
305- jd.nova aliança	3-noturno leste	606 - jd. Paulo gomes
306 - cidade universitaria	401 - simioni	606-fazenda experimental
306- jd. Marchesi	401 vila carvalho	607 - jose sampaio
306-jd.marchesi	401-v.carvalho	607-jd.eugenio lopes
307 - marincek	402-ribeirão verde i	608-jd.amalia
307- c. Universitaria	403 jd manojel pena	630- pq. São sebastião
307-cidade universitaria	403-jd.manoel penna	630-pq.sao sebastiao
308 marincek	406 - jardim recreio	670-jd.jamil cury
308-marincek	406- jd maria da graça	680-jd.orestes lopes
309 - jardim independencia / hc	406-jd.maria da graça	688- expresso jd. Amalia
310 quintino/ avelino	407 - geraldo de carvalho	6-noturno sudoeste
310-quintino-avelino	407- jd paulo gomes	702 - ribeirao verde
311 - expresso avelino palma	407-jd.paulo gomes	703 - bom pastor
311-expresso avelino	408- jose sampaio	703-parque dos servidores
315- bonfim c. Eliseos	408-jose sampaio	707 - jardim procopio
315-campos eliseos -bonfim	411 - expresso simioni	708-jd.heitor rigon
320 jd.salgado filho i	420-ribeirão verde ii	730 - parque portinari
320-jd.salgado filho i	437 castelo branco vila amelia	730-parque portinari
330-jd. Palmares	437-castelo branco-v.amelia	770 - alexandre balbo
335 - sao jose / adao carmo	460-parque ribeirão	788-expresso heitor rigon
337 lagoinha ipiranga	470-jd.paiva	7-noturno oeste
337-lagoinha -ipuranga	499 circular 4	801 - regatas caicara
351- jardim irajá/monte alegre	499- circular iiiii	830 - parque servidores
351-lapa-bonfim	4-norte sudeste	8-noturno noroeste
360 jd. Centenario	501 - quintino avelino	901 - parque avelino
360-jd.centenario	501- higienopolis	901- regatas caicara
370- jamil cury	501-higienopolis	Jd. Salgado filho i
370 jd recreio	503-recreio das acacias	Não identificada
370-jd.recreio	506 - jardim paiva i	Não identificado
373- v.abranches	506-jd.progresso	
	507-planalto verde	

TABELA 2 –NOMENCLATURA CORRETA DAS LINHAS

CÓD	NOME	503	Recreio das Acácias	680	Jd. Orestes Lopes
<b>CIRCULAR</b>				<b>688</b>	Expresso Jd. Amália
199	Circular 1	104	Jd. Canadá	708	Jd. Heitor Rigon
299	Circular 2	204	City Ribeirão	788	Expresso Rigon
399	Circular 3				<b>DIAMETRAL</b>
499	Circular 4	305	Jd. Nova Aliança	315	Campos Elíseos-Bonfim
<b>DIAMETRAL</b>				351	Lapa-Bonfim
602	Aeroporto Leite Lopes	105	Ribeirão Shopping		
		205	Jd. João Rossi	136	Castelo Branco-Adão do Carmo
<b>RADIAL</b>				437	Castelo Branco-V. Amélia
201	Quintino II	106	D'Elboux		
210	Simoni	206	V. Virgínia		
211	Expresso Simoni	306	Jd. Marchesi		
301	Avelino Palma	360	Jd. Centenário	337	Lagoinha-Ipiranga
310	Quintino/Avelino			373	V. Abranches-Ipiranga
311	Expresso Avelino	406	Jd. Maria das Graças		
401	V. Carvalho	460	Pq. Ribeirão	236	São José-Adão do Carmo
501	Higienópolis	506	Jd. Progresso		
		606	Fazenda Experimental	147	Jd. Irajá-Monte Alegre
101	Pq. Avelino			148	Santa Cruz-Ipiranga
110	Quintino I				<b>PERIMETRAL</b>
601	Distrito Empresarial	207	Hospital das Clínicas	156	Pq. Ribeirão-Shopping
		307	Cidade Universitária	178	D. Mielle-Hospital das Clínicas
102	Jd. Independência	370	Jd. Recreio	187	Heitor Rigon-Hospital das Clínicas
202	Jd. Iara			217	Quintino-Hospital das Clínicas
220	Pq. Exposições	107	Sumarézinho		
		407	Jd. Paulo Gomes	901	Regatas/Caiçara
302	Jd. Aeroporto	470	Jd. Paiva	902	FEAPAM
320	Jd. Salgado Filho I			803	Estádio do Comercial
402	Ribeirão Verde I	607	Jd. Eugênio Lopes	903	Estádio do Botafogo
420	Ribeirão Verde II	670	Jd. Jamil Cury		
303	Bom Pastor	507	Planalto Verde		
330	Jd. Palmeiras				
		108	Jd. Presidente Dutra		
603	Jd. Juliana	208	V. Albertina	001	Norte
630	Pq. São Sebastião	308	Marincek	002	Nordeste
703	Pq. Servidores	380	Geraldo de Carvalho	003	Leste
730	Pq. Portinari	388	Expresso Via Norte	004	Sudeste
				005	Sul
103	Iguatemi	408	José Sampaio	006	Sudoeste
130	Fórum	508	Jd. Procópio	007	Oeste
203	Ribeirânia	580	Jd. Alexandre Balbo	008	Noroeste
403	Jd. Manoel Penna	608	Jd. Amália		

É possível ver na Tabela 1, por exemplo, quatro linhas com o código 101, porém com nomes diferentes ou padrões de escrita diferente (Quintino Facci, Pq. Avelino, Parque Avelino, etc). Há algumas sutis diferenças que são apenas espaços inseridos a mais. Utilizando o mesmo exemplo anterior, temos:

**TABELA 4 – EXEMPLO DAS LINHAS DUPLICADAS**

101 - quintino facci i
101- parque avelino
101- pq avelino
101-parque avelino

Na Tabela 4 temos duas informações iguais, segunda e quarta linha. No entanto não estão padronizadas com relação aos espaços, tornando-se diferentes para qualquer sistema de dados, uma planilha eletrônica, por exemplo.

Ao longo da Tabela 1 existem várias situações de duplicidade de informação. Há ainda outro problema com o exemplo da Tabela 4, a primeira linha (101 - quintino facci) não está correta, sendo necessário encontrar qual o seu código atual e modifica-lo.

## PADRONIZAÇÃO E NOMENCLATURAS

---

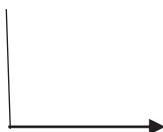
A padronização de nomes e códigos embora seja uma atividade trivial é muito fácil se perder quando há muitas informações envolvidas.

Neste contexto será sugerida uma padronização dos códigos e dos nomes.

**XXX - JD. YYYY**



3 Dígitos para o código



Logradouros simplificados/reduzidos  
terminados com um ponto (.)

No exemplo acima foram utilizados 3 dígitos para o código e a separação entre o código e o nome da linha é feito com um espaço, hífen e espaço. Por exemplo:

**001 - NORTE** ou **101 - PQ. AVELINO**

## DEMAIS PADRONIZAÇÕES

---

Outro campo que é de grande interesse é “OBSERVAÇÕES”, onde são complementadas as informações de horários, por exemplo.

São encontradas por várias vezes diferentes padrões de horários digitados, dificultando mais uma vez a utilização dos dados de maneira rápida. Visando a melhoria da informação será sugerida, também, a padronização do horário digitado.

# HH:MM

Preferencialmente utilizar o padrão 24hrs e mesmo que se tratando de horários anteriores a 12hrs, utilizar o zero à esquerda. Por exemplo:

**08:45** ou **10:00** ou **02:44**.