

**ALEXANDRE JOSÉ DE SOUSA**

**ESTATÍSTICA APLICADA À ANÁLISE DO TRANSPORTE  
FERROVIÁRIO DE PASSAGEIROS NA CIDADE DE  
SÃO PAULO**

**SÃO PAULO  
2017**

**ALEXANDRE JOSÉ DE SOUSA**

**ESTATÍSTICA APLICADA À ANÁLISE DO TRANSPORTE  
FERROVIÁRIO DE PASSAGEIROS NA CIDADE DE  
SÃO PAULO**

Monografia apresentada à Faculdade de Tecnologia de São Paulo, como parte dos requisitos para a obtenção do grau de Tecnólogo em Construção Civil na Modalidade de Movimento de Terra e Pavimentação.

Orientadora: Prof<sup>ª</sup>. Esp. Josie Pereira de Freitas Micalli

SÃO PAULO  
2017

## **DEDICATÓRIA**

Dedico à Lucila e a Adorico.

## AGRADECIMENTOS

A princípio gostaria de agradecer à Elizabete Ribeiro, a qual forneceu algumas das provisões que necessitei independentemente se elas eram relacionadas à graduação acadêmica ou não. Provisões estas fornecidas com um carinho característico das pessoas que possuem um tipo de bondade singular, especial, em seus corações, assim como ela o é.

À minha orientadora e professora, Josie Micalli, pelo tempo despendido em auxiliar na melhoria contínua deste trabalho, lendo e relendo, comentando e acrescentando notas, sem nada receber em troca. Percebe-se através da sua atenção o sentimento materno que transcende o âmbito familiar e adentra as barreiras do ofício para com seus alunos. Obrigado pela sua orientação.

Aos meus amigos, Duone Eleno, Rafael Cruz, Felipe Zanetti, Lázaro Sena, Anderson Souza, Victor Souza, Pedro Henrique, Claudio Inacio, Fernando Oliveira e Luciano Sena, pelo companheirismo ao logarmos as veredas das labutas que apresentaram-se em alguns trechos de forma tão análoga a todos, fazendo com que nossas experiências e aflições pudessem ser compreendidas mutuamente e pudéssemos auxiliar uns aos outros quando o fardo se tornava demasiado pesado para prosseguir.

À Carolina Ozaki, minha namorada e companheira, por toda sua ajuda e paciência quando por vezes necessitei sacrificar o pouquíssimo tempo que tínhamos juntos para dar término a este trabalho que aqui se apresenta, jamais demonstrando sequer traço de irascibilidade. Muito pelo contrário, sempre se revelando afável e disposta a auxiliar caso a necessidade surgisse.

Às minhas irmãs, Dulcineia e Dulcemary Sousa, especialmente à primeira, que não assistiu de braços cruzados aos óbices surgidos e por vezes auxiliou das formas que eram cabíveis a si, mesmo em momentos de adversidades particulares, fazendo o possível e de bom grado não só para mim como também para a família.

Um agradecimento póstumo ao homem que muito me ensinou em campos diversos, Adorico Sousa. À perspectiva da infância, exímio; nas últimas lembranças, sublime. Um homem de força de vontade e dedicação de magnitudes colossais. Sinto grande orgulho em possuir seu sobrenome, pai.

E finalmente a uma pessoa de extrema importância para mim, Lucila Sousa. Obrigado mãe por toda paciência e dedicação, amor e carinho, por fornecer-me a oportunidade de uma vida melhor e ser essa pessoa simples, mas de coração que tal qual não há igual. Magnânima! Alegro-me em saber que possuo tal progênie maternal.

A todos tenho imenso carinho e estima.

## EPÍGRAFE

*Se eu conversasse com Deus  
Iria lhe perguntar:  
Por que é que sofremos tanto  
Quando viemos pra cá?  
Que dívida é essa  
Que a gente tem de morrer pra pagar?  
Perguntaria também  
Como é que ele é feito  
Que não dorme, que não come  
E assim vive satisfeito.  
Por que foi que ele não fez  
A gente do mesmo jeito?  
Por que existem uns felizes  
E outros que sofrem tanto?  
Nascemos do mesmo jeito,  
Moramos no mesmo canto.  
Quem foi temperar o choro  
E acabou salgando o pranto?*

**Leandro Gomes de Barros**

## **RESUMO**

O trabalho analisa, através da estatística, o quantitativo de usuários que entram no sistema de uma linha férrea, através das estações, e averigua se essa demanda é suprida pela frota de trens que está sendo adotada. Os dados fornecidos pela empresa responsável pelo transporte foram submetidos a um tratamento estatístico, iniciado pela organização das informações e definição do íterim estudado, seguido pelos cálculos das medidas descritivas e verificação de suas precisões e terminado na disposição dos resultados e comparação entre o quantitativo de passageiros e os tipos e quantidades de trens empregados. Como resultado foram obtidas as quantidades médias de passageiros que entram em cada estação da linha, de forma ordenada, as quantidades de trens que circulam no íterim de estudo e a resposta à questão se há ou não atendimento da demanda de usuários e, conseqüentemente, se sucede a superlotação das composições e estações. Por fim, a análise diz, com certa margem de segurança, que em grande parte dos casos há a necessidade de adotar alguma medida face ao grande número de pessoas que usufruem deste modal.

**Palavras-chave: Estatística Aplicada. Transporte Ferroviário de Passageiros. Transporte Público.**

## LISTA DE ILUSTRAÇÕES

Figura 3.1 – Zoneamento – Pesquisa Origem e Destino 2007 .....	21
Figura 3.2 – Sub-Regiões da Secretaria de Estado dos Transportes Metropolitanos - 2007....	22
Figura 4.1 – Exemplo de Gráficos de Barras Verticais e <i>Pizza</i> .....	28
Figura 4.2 – Exemplo de uma Tabela de Distribuição de Frequências .....	31
Figura 4.3 – Exemplo de um Histograma junto ao Polígono de Frequência.....	32
Figura 4.4 – Percentil – Equivalência em Termos de Posicionamento .....	36
Figura 4.5 – Comparação de duas linhas com mesma média e desvios distintos.....	38
Figura 4.6 – Distribuições simétrica, assimétrica positiva e assimétrica negativa.....	41
Figura 4.7 – Curvas Platicúrtica, Mesocúrtica e Leptocúrtica .....	42
Figura 5.1 – Média de embarcados por intervalos de tempo na estação de Itaim Paulista na Linha 12 – Safira da CPTM em dias úteis em uma situação de normalidade .....	44
Figura 5.2 – Verificação dos dados primários .....	46
Figura 6.1 – Médias de embarcados por hora em dias úteis.....	48
Figura 6.2 – Matriz de disposição dos dados ordenados .....	50
Figura 6.3 – Histograma dos intervalos das classes e suas frequências absolutas .....	51

## LISTA DE TABELAS

Tabela 6.1 – Definição dos limites do intervalo onde foram gerados os números aleatórios ..	49
Tabela 6.2 – Distribuição de frequências .....	51
Tabela 6.3 – Valores encontrados no <i>Rol</i> de acordo com os percentis .....	52
Tabela 6.4 – Resumo das médias adotadas de passageiros que entram no sistema .....	53
Tabela 6.5 – Resumo da frota de material rodante para o intervalo em estudo.....	54



## **LISTA DE ABREVIATURAS E SIGLAS**

**CBTU** – Companhia Brasileira de Trens Urbanos

**CET** – Companhia de Engenharia de Tráfego

**CPTM** – Companhia Paulista de Trens Metropolitanos

**EMTU** – Empresa Metropolitana de Transportes Urbanos

**FEPASA** – Ferrovia Paulista S.A.

**IBGE** – Instituto Brasileiro de Geografia e Estatística

**METRÔ** – Companhia do Metropolitano de São Paulo

**RMSP** – Região Metropolitana de São Paulo

**SMT** – Secretaria Municipal de Transportes

**SPTrans** – São Paulo Transportes

**STM** – Secretaria de Estado dos Transportes Metropolitanos

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
1.1	Objetivo .....	14
1.1.1	Objetivo Geral .....	14
1.1.2	Objetivo Específico .....	14
1.2	Justificativa.....	14
1.3	Metodologia.....	15
<b>2</b>	<b>SÚMULA DO HISTÓRICO FERROVIÁRIO DA CIDADE DE SÃO PAULO</b>	
	<b>16</b>	
<b>3</b>	<b>PLANEJAMENTO DA MOBILIDADE URBANA EM SÃO PAULO .....</b>	<b>19</b>
3.1	Pesquisa Origem/Destino .....	19
3.1.1	Planejamento .....	20
3.1.1.1	Zoneamento .....	20
3.1.1.2	Amostra .....	22
3.1.2	Execução .....	23
3.1.2.1	Inquérito Residencial.....	23
	3.1.2.1.1 Questionário .....	24
	3.1.2.1.2 Abordagem.....	24
	3.1.2.1.3 Checagem .....	24
3.1.2.2	Tratamento dos Dados.....	25
3.2	Metodologia de Levantamento de Usuários - CPTM.....	25
<b>4</b>	<b>ESTATÍSTICA – DESCRIÇÃO E ANÁLISE.....</b>	<b>27</b>
4.1	Séries Estatísticas .....	27
4.2	Descrição Gráfica de Variáveis Quantitativas.....	28
4.3	Medidas Descritivas .....	32
4.3.1	Tendência Central.....	33
4.3.1.1	Média Aritmética.....	33
4.3.1.2	Média Aritmética Ponderada.....	33
4.3.1.3	Mediana .....	34
4.3.1.4	Moda .....	35
4.3.2	Ordenação.....	35
4.3.3	Dispersão.....	37
4.3.3.1	Amplitude Total .....	37
4.3.3.2	Variância .....	37
4.3.3.3	Desvio Padrão .....	38
4.3.3.4	Coefficiente de Variação de Pearson .....	39
4.3.3.5	Escore Padronizado .....	39
4.3.3.6	Outliers .....	40
4.3.4	Assimetria.....	41
4.3.5	Curtose .....	42
<b>5</b>	<b>METODOLOGIA.....</b>	<b>43</b>
5.1	Levantamento e Organização dos Dados .....	43
5.2	Análise e Validação das Informações.....	46
5.3	Exame da Demanda e Suprimento.....	47
<b>6</b>	<b>RESULTADOS .....</b>	<b>48</b>

<b>7</b>	<b>CONCLUSÃO.....</b>	<b>55</b>
	<b>REFERÊNCIAS .....</b>	<b>57</b>
	<b>ANEXO A.....</b>	<b>58</b>
	<b>ANEXO B.....</b>	<b>59</b>
	<b>APÊNDICE A .....</b>	<b>60</b>

## 1 INTRODUÇÃO

Hoje comumente usado por milhares de pessoas por dia, esse ramerrão faz com que por vezes passe despercebido o fato de como o transporte ferroviário tem sido um importante instrumento no desenvolvimento do Estado de São Paulo, assim como do país.

Sendo uma significativa forma de locomoção tanto de cargas como de pessoas, sua relevância pode apresentar-se, além de outras formas, através de várias vantagens em relação a outros modais de transporte e sua influência histórica. Segundo Micalli (2016) temos como exemplos de vantagens deste modal: sua grande capacidade de carga, o fato de ser adequado a grandes distâncias, elevada eficiência energética, baixo custo de transporte, baixo custo de manutenção, possui maior segurança (em relação ao modal rodoviário, visto que ocorrem poucos acidentes, furtos e roubos) e pouco poluente.

As construções das primeiras ferrovias em São Paulo estão ligadas às dificuldades que se mostraram através das tentativas de um modelo agroexportador na época em que ocorreu o êxito das plantações de café no estado. A primeira ferrovia de São Paulo teve início das suas obras em 1860 e término sete anos mais tarde, sendo ela o prelúdio da malha ferroviária paulista que tempos depois desenvolveu-se e ramificou em todos os sentidos do Estado. Entre 1880 a 1918 houve uma nova incorporação às funcionalidades dos trens. Estes começaram a ser adotados como uma forma de transporte para os trabalhadores das regiões periféricas chegarem aos seus ofícios (STEFANI, 2007).

O sistema de trens da Região Metropolitana de São Paulo (RMSP) passou por mudanças na sua administração e, em 1992, a Companhia Paulista de Trens Metropolitanos (CPTM) tomou sua gestão em substituição à Companhia Brasileira de Trens Urbanos (CBTU) e à Ferrovia Paulista S.A. (FEPASA). Mais tarde, em 2016, após a construção de novos trechos e inaugurações de novas estações, a malha ferroviária de São Paulo tornou-se presente em 22 municípios, com 260,8 km de malha e 92 estações. Doze anos mais tarde, em 2004, os trens sob a responsabilidade da CPTM transportavam 368,8 milhões de pessoas, valor este que em doze anos subiu para 819,5 milhões, apresentando um crescimento de 122% na quantidade de utentes (CPTM, 2016).

Em relação às características socioeconômicas dos habitantes da Região Metropolitana de São Paulo, segundo o METRÔ (2013), através da sua Pesquisa de Mobilidade de 2012, a população cresceu aproximadamente 20% em quinze anos, sendo que em 1997 o número aproximado de habitantes era de 16.792 milhões, passando para um valor

similar a 20.012 milhões em 2012. Quanto a emprego, de acordo com o mesmo estudo, a sub-região Centro concentra 65% de todas as vagas da RMSP.

Para as regiões mais periféricas, uma grande característica que pode determinar a escolha do modal de transporte público é sua agilidade. Considerando-se a grande aglomeração de oferta de trabalho nos centros das cidades, um modal que consiga fazer esse trajeto em um tempo relativamente aceitável sem que gere maiores custos aos passageiros é outro aspecto relevante no momento da escolha. Dentro dessas premissas, o transporte ferroviário e metroviário destacam-se quando ponderados em relação ao ônibus, que pode ser considerado bom pela sua flexibilidade e capacidade, mas para viagens dentro do mesmo distrito, ou até mesmo para distritos vizinhos.

Muitas dessas regiões periféricas não são alcançadas pelo metrô, colocando em primeira instância o transporte ferroviário, visto o alcance da sua malha. Levando-se em conta o crescimento populacional, que gera áreas com alta densidade de habitantes que utilizam o modal de transporte em horários concomitantes em função de jornadas de trabalho ou períodos letivos acadêmicos com turnos padronizados, disso acaba ocorrendo o pico de capacidade do transporte, dado a grande massa populacional que está se deslocando com o mesmo sentido. Desse tipo de situação advém vários problemas, tais como atrasos na rede, uma diminuição no período de manutenção tanto nos trens quanto nas vias, entre outros problemas mais específicos, como brigas causadas por estresse devido à demora e trens extremamente abarrotados, ou em casos mais graves onde pessoas caem nas vias por causa da superlotação das plataformas.

Surgidos esses problemas, seguidamente há reclamações dos próprios usuários, oriundas das insatisfações quanto ao modal, o que pode denegrir a imagem deste, fazendo com que os usuários tenham uma impressão negativa e até mesmo de inépcia. Para a resolução desse problema consideram-se hipóteses para melhorar a forma como serão transportados os utentes, porém, sejam quais forem elas, tais precisam basear-se em premissas sólidas sobre informações da população, tais como as quantidades de passageiros que usufruem do modal e os polos que mais geram viagens, para a partir delas arquitetar uma forma de melhorá-lo ou desenvolver novos meios de transporte para a população. Como exemplos desse tipo de pesquisa para obtenção de informações tem-se a Pesquisa Origem e Destino, que desde 1967 é executada pela Companhia do Metropolitano de São Paulo (METRÔ) em associação com outras entidades, com uma periodicidade de dez anos; a Pesquisa de Mobilidade, executada em 2012 pela mesma Companhia, com o intuito de aferição dos dados; a metodologia da CPTM, baseada na quantificação da utilização de

Bilhete Único nos bloqueios das estações; dentre outras mais específicas e pertinentes à outras companhias.

Este trabalho vem apresentar uma metodologia através da qual é feita uma análise estatística das quantidades de usuários que embarcam, com o mesmo sentido, nas estações de uma linha de transporte ferroviário, a demanda de trens utilizados e se eles atendem a necessidade exigida pelo quantitativo de utentes. Para isso são identificadas as quantidades de passageiros que estão embarcando nas estações em horários de pico, analisadas e verificadas estatisticamente as informações obtidas, averiguados os tipos de trens que estão sendo usados diante dessa situação e feito um comparativo para constatar se atendem a quantidade de usuários.

Com isso o trabalho colabora com informações resultantes de uma síntese estatística baseada em dados obtidos previamente, enunciando em quais estações embarcam grandes quantidades de passageiros e se há a ocorrência da superlotação do transporte, proporcionando um melhor entendimento de como age o fluxo de passageiros em uma linha férrea e servindo como base para futuras soluções de descongestionamento, tendo como seu maior intento o desenvolvimento de métodos para que os passageiros possam ter uma viagem com mais qualidade e dignidade nos horários de pico.

A metodologia empregada dividiu-se em três partes. Na primeira os dados, sendo estes fornecidos pela própria companhia responsável pelo transporte ferroviário de passageiros, foram obtidos, separados e organizados, possibilitando e facilitando seu tratamento. Na segunda, cálculos das medidas descritivas da série foram engendrados para que os dados fossem analisados e validados. Na terceira e última etapa foram examinadas as demandas médias de usuários, a forma de suprimento que está sendo usada para atendê-las e a verificação se são ou não suficientes.

Quanto ao que tange a estruturação deste trabalho, este foi dividido em sete capítulos. No primeiro são apresentados os objetivos geral e específicos, assim como sua justificativa e um breve resumo da metodologia. O segundo capítulo retrata, de forma concisa, o histórico ferroviário da cidade de São Paulo, iniciando em meados do século XIX e consumando-se no cenário atual, passando pelos motivos das construções das primeiras ferrovias, algumas de suas funcionalidades ao longo do tempo e outros. No terceiro o tema abordado é o planejamento da mobilidade urbana no estado, onde é contemplada a Pesquisa Origem e Destino, executada pelo METRÔ com o auxílio de outras entidades, e o método através do qual a CPTM obtém o quantitativo de passageiros que adentram o sistema do modal. No quarto são apresentadas as partes da estatística que serão utilizadas na

metodologia, mostrando como são executadas as descrições gráficas das variáveis quantitativas, as medidas descritivas, e uma breve explicação sobre séries estatísticas. O quinto é a metodologia em si, expondo paulatinamente seu desenvolvimento, as proposições adotadas e as disposições de suas etapas para a formação da metodologia integral. No sexto são apresentados os resultados da aplicação da metodologia em uma via do sistema ferroviário, usando como exemplo a estação de Itaim Paulista, assim como a Linha 12 – Safira, a qual pertence. No sétimo e último capítulo apresenta-se a conclusão, onde é feita uma reflexão sobre o desenvolvimento do trabalho, os óbices encontrados e o confronto entre as expectativas no início do trabalho e os resultados obtidos no final dele.

## **1.1 Objetivo**

### ***1.1.1 Objetivo Geral***

Analisar estatisticamente as quantidades de usuários que embarcam, com o mesmo sentido, nas estações de uma linha de transporte ferroviário, a demanda de trens utilizados e se eles atendem a necessidade exigida pelo quantitativo de utentes.

### ***1.1.2 Objetivo Específico***

Tencionando lograr o objetivo principal, torna-se necessário a execução de objetivos específicos. São eles:

- Identificar a quantidade de passageiros que estão embarcando nas estações de uma ferrovia em horários de pico;
- Empregar e analisar estatisticamente as informações obtidas;
- Verificar os tipos de trens que estão sendo usados diante dessa situação;
- Constatar se a frota de transporte adotada atenderá à quantidade de usuários.

## **1.2 Justificativa**

Este trabalho justifica-se pela sua colaboração com informações resultantes de uma síntese estatística baseada em dados obtidos previamente, enunciando em quais estações embarcam grandes quantidades de passageiros e se há a ocorrência da superlotação do transporte, proporcionando um melhor entendimento de como age o fluxo de passageiros em uma linha férrea e servindo como base para futuras soluções de descongestionamento, tendo

como seu maior intento o desenvolvimento de métodos para que os passageiros possam ter uma viagem com mais qualidade e dignidade nos horários de pico.

### **1.3 Metodologia**

O presente trabalho será elaborado através do desenvolvimento de um modelo descritivo baseado em dados fornecidos pela empresa responsável pelo transporte ferroviário de passageiros. O modelo elucida o comportamento do fluxo de usuários que entram nas estações de uma dada linha do modal e se há o atendimento da demanda de passageiros. A metodologia, em sumo, é separada em três etapas.

Na primeira os dados são separados e organizados, possibilitando a facilitando seu tratamento. Nela os dados são examinados de forma facilitar os procedimentos que serão realizados.

Na segunda etapa são calculadas as medidas descritivas das informações tratadas no primeiro ciclo, verificadas suas precisões e analisadas se são fidedignas ou não para a procedência do estudo, visando exatidão para uma simulação mais próxima da realidade da rotina do modal.

Na terceira e última etapa são examinadas as demandas médias de usuários, a forma de suprimento que está sendo usada para atendê-las e a verificação se são ou não suficientes.



## **2 SÚMULA DO HISTÓRICO FERROVIÁRIO DA CIDADE DE SÃO PAULO**

O transporte ferroviário tem mostrado um importante papel na história do desenvolvimento do país desde meados do século XIX. Antes da implantação de ferrovias que ligavam o centro de São Paulo às outras regiões, o transporte de materiais era feito por escravos ou em lombos de burros e mulas por meio de diversos “caminhos de tropas” que cortavam a serra. Muitos caminhos de terra, durante as épocas de chuvas, transformavam-se em lamaçais intransponíveis. Outra forma importante para locomoção e transporte na época, apesar das dificuldades topográficas locais, era a navegação fluvial (STEFANI, 2007).

Segundo Reis Filho (1994 apud STEFANI, 2007, p. 28): “Antes da construção das ferrovias, as cidades que não se encontravam dentro da faixa litorânea eram sempre pequenas e sua população quase que totalmente dedicada à atividade rural.”.

As construções das primeiras ferrovias em São Paulo estão essencialmente ligadas ao sucesso advindo das plantações de café no Estado. Produto esse que, por sua vez, teve rápido crescimento do seu comércio, elucidando as dificuldades que os caminhos existentes impunham ao modelo agroexportador.

Pouco mais de três décadas após a primeira ocorrência histórica do transporte de passageiros nesse modal (Inglaterra, 1825), foi implantada a primeira ferrovia do estado, batizada de São Paulo Railway, com início das obras em 1860 e término em 1867 (STEFANI, 2007). Após a construção desse trecho inicial, a malha ferroviária paulista começou a desenvolver-se e ramificar em todos os sentidos do estado.

Na época da implantação dos primeiros trilhos os responsáveis não usaram uma forma padronizada de bitolas (largura entre trilhos). A priori iniciaram as obras com bitolas de 1,60 metro, porém, essa bitola mostrou-se economicamente inviável e rapidamente passaram a ser adotadas, sem critérios, outros tipos de larguras, tais como 1,00, 0,96 e 0,60 metro (STEFANI, 2007). A falta de normatização para tal escolha deu origem a um problema com o qual convivemos até hoje, as constantes baldeações e as trocas de composições.

A partir de 1870 a cidade desenvolveu-se com grande agilidade. Fazendeiros do café fixaram residência na cidade, que passou a concentrar cada vez mais comércios, indústrias e serviços. Nessa mesma época ocorre também o loteamento de antigas chácaras ao redor da cidade, o que leva a multiplicação das residências. De acordo com Reis Filho (1994 apud STEFANI, 2007, p. 28), as primeiras indústrias, instaladas entre 1860 e 1890 na capital paulista, acompanharam o trajeto das vias férreas.

O processo de urbanização e industrialização na capital paulista começa, também, a evidenciar a diferenciação social, seja na formação de novos bairros junto à área urbana mais centralizada, quanto nos prolongamentos nas diferentes regiões da cidade. De um lado, vão surgindo os bairros “jardins”, no entorno da região da Avenida Paulista, que juntamente com a região de Higienópolis e Campos Elísios, vão instalar os membros da oligarquia cafeeira e industrial, servidos por uma extensa rede de bondes que interligava a região central da cidade, além dos primeiros automóveis importados, símbolo de prestígio e poder. De outro, os bairros periféricos, formados no entorno das indústrias, ao longo dos eixos ferroviários da Santos-Jundiaí, e mais tarde, da Sorocabana e da Central do Brasil e que vão abrigar os trabalhadores assalariados e suas famílias (STEFANI, 2007).

O período existente entre os anos de 1880 a 1918 é de grande importância para o transporte ferroviário, pois é nesse ínterim que ele sofre uma grande transformação em uma de suas funcionalidades. Até o final do século XIX ele servia como transporte de imigrantes com destino às fazendas de café, passando a servir agora aos trabalhadores das regiões periféricas da capital, dando início ao transporte suburbano de passageiros.

O reflexo dessa nova funcionalidade somado à acentuação de imigrantes que ocorre também no final do século XIX é observado na própria capital. Em 1872 a população presente em São Paulo era pouco mais de 800 mil, passando a um sobejo valor de 4,5 milhões em 1920. Em 1940 chegamos a uma população superior a 7 milhões de habitantes, onde mais de 40% já vive em áreas urbanas. (BRASIL, 1990).

Apesar de toda a contribuição do modal com o crescimento e desenvolvimento do estado, no decênio que se seguiu após 1930, o sistema ferroviário, que em alguns trechos se encontrava em situação precária, começou a entrar em um constante e acelerado processo de estagnação, porquanto que as estradas de rodagem começaram a ganhar grande impulso no sistema de locomoção. Segundo Stefani (2007, p. 128) “enquanto a rodovia assistia a um aumento de 59,5 milhões de passageiros em 1961 para 225,8 milhões em 1971, a ferrovia sofria uma redução de 165,2 milhões em 1961 para 11,9 milhões de passageiros em 1971.”.

Houve também grande divergência entre a quinta (1987/1991) e a sexta (1991/1994) gestão da Ferrovia Paulista S.A. em relação ao transporte de passageiros de longo percurso. A primeira defendia este como sendo de grande importância, tendo como uma de suas expectativas reconquistar os passageiros que foram obrigados a adotar o ônibus. Em contrapartida, sua sexta gestão dizia que era necessário uma justificativa de volume de tráfego para colocar em melhor posição o transporte de longo percurso, classificando-o até mesmo como problemático (STEFANI, 2007).

É possível observar, nessa administração, por meio de seu programa de investimentos, a urgência de adequar o sistema de cargas aos interesses da iniciativa privada. Dos 166,6 milhões de dólares investidos no sistema, 163,8 foram destinados ao transporte de carga. Apenas 2,8 milhões de dólares foram para o transporte de subúrbio. (STEFANI, 2007, p. 171)

Seguindo essa linha concretizou-se o plano de divisão total do sistema, ação decisiva para a privatização do transporte ferroviário de carga e eliminação definitiva do transporte de passageiros de longo percurso.

Em 28 de maio de 1992 é autorizada a criação da Companhia Paulista de Trens Metropolitanos, através da Lei nº 7.861, segundo a qual a nova companhia deveria assumir o sistema de trens da Região Metropolitana de São Paulo, em substituição à CBTU e à FEPASA.

No início de sua gestão eram transportados diariamente 894.000 usuários, número que foi crescendo gradualmente ao passo que os anos se passavam. Dois anos após o início da sua gestão (1994) há uma queda de aproximadamente 11% na quantidade de passageiros transportados por dia, ano em que a empresa começa a operar efetivamente as atuais linhas 7, 10, 11 e 12, onde a frota herdada se encontrava deteriorada, levando a um primeiro plano de modernização de composições (CPTM, 2016).

Em 1996 a quantidade de passageiros que usam os trens diariamente volta a subir, chegando ao valor de 835.000 passageiros. Nesse período a CPTM começa a operar as linhas 8 e 9. Nessa época os trens ainda partiam com os chamados “pingentes” e “surfistas”, pessoas que se arriscavam viajando fora dos vagões, sobre eles e suspensos, apoiando-se nas portas, que trafegavam abertas durante as viagens. Quatro anos depois há um aumento de 4% na quantidade diária de usuários, elevando o número para 869.000. Nesse mesmo ano inicia-se a operação do Expresso Leste, junto à entrega de quatro novas estações e inauguração de outras seis (CPTM, 2016).

Em 2001 a quantidade de passageiros por dia passa de 1.000.000, ano em que entra em operação a estação Vila Olímpia. Em 2008 o montante de usuários salta para 1.800.000, época em que há também a inauguração de outras 5 novas estações (CPTM, 2016).

Dois anos mais tarde a quantidade de pessoas que usufruem diariamente das linhas bate um novo recorde, chegando a 2.408.342 usuários. Em 2015 a média de passageiros transportados nos dias úteis era de 2.783.000 (CPTM, 2016).

### **3 PLANEJAMENTO DA MOBILIDADE URBANA EM SÃO PAULO**

#### **3.1 Pesquisa Origem/Destino**

O planejamento no transporte é fundamental, pois é a partir dele que se elaboram estudos para que se possa inferir situações e arquitetar soluções com uma boa margem de segurança baseando-se nos resultados, visto que muitos dos dados resultantes desses estudos projetam o quantitativo de usuários em relação ao modal em pauta. Uma das principais ferramentas no planejamento da mobilidade urbana em São Paulo é denominada Pesquisa Origem – Destino.

A Pesquisa Origem e Destino, comumente chamada de “Pesquisa O/D” é um instrumento vital para o planejamento de transporte, pois fornece dados para o conhecimento da natureza dos deslocamentos da população em um aglomerado urbano, incluindo aí sua situação socioeconômica.

Com a Pesquisa O/D é possível identificar as principais viagens diárias das pessoas, conforme o motivo e o modo de transporte utilizado. Essas viagens, quando mapeadas, produzem uma fotografia dos fluxos na cidade. A comparação desses fluxos com a rede de linhas de transporte e a rede de vias existentes permite identificar as carências no atendimento da demanda de transporte e, portanto, fundamentar as propostas para a ampliação dessas redes.

Essa pesquisa permite ainda estabelecer relações quantitativas entre as viagens realizadas e diversas outras variáveis, como características socioeconômicas da população e aspectos físicos da ocupação urbana, de forma a estabelecer projeções futuras das necessidades de viagens das pessoas. (METRÔ, 2012)

A primeira pesquisa O/D foi realizada em 1967, adotando como critério temporal uma periodicidade de dez anos a cada pesquisa, sendo elas realizadas em 1977, 1987, 1997 e 2007. Porém, percebeu-se que o dinamismo da RMSP exigia uma frequência menor entre as pesquisas e, em 2002, o METRÔ realizou um levantamento com o intuito de aferição em relação ao levantamento feito em 1997. Utilizando-se de uma amostra menor que as de pesquisas de O/D realizadas a cada dez anos, a aferição foi realizada num zoneamento mais agregado e teve como objetivo verificar alterações no índice de mobilidade, ou seja, viagens por habitante, na RMSP e a distribuição das viagens entre os modos coletivo e individual. Em

2012 foi realizada a Pesquisa de Mobilidade da RMSP, com o intuito de atualizar esses mesmos indicadores, cotejando assim os dados dos levantamentos na metade do intervalo determinado na época em que foram iniciadas as investigações. (METRÔ, 2012)

No planejamento de atividades da última pesquisa feita (2007), participaram a Secretaria de Estado dos Transportes Metropolitanos – STM, Secretaria Municipal de Transportes – SMT, Companhia do Metropolitano de São Paulo – METRÔ, Companhia Paulista de Trens Metropolitanos – CPTM, Empresa Metropolitana de Transportes Urbanos – EMTU, São Paulo Transporte – SPTrans e a Companhia de Engenharia de Tráfego – CET.

### ***3.1.1 Planejamento***

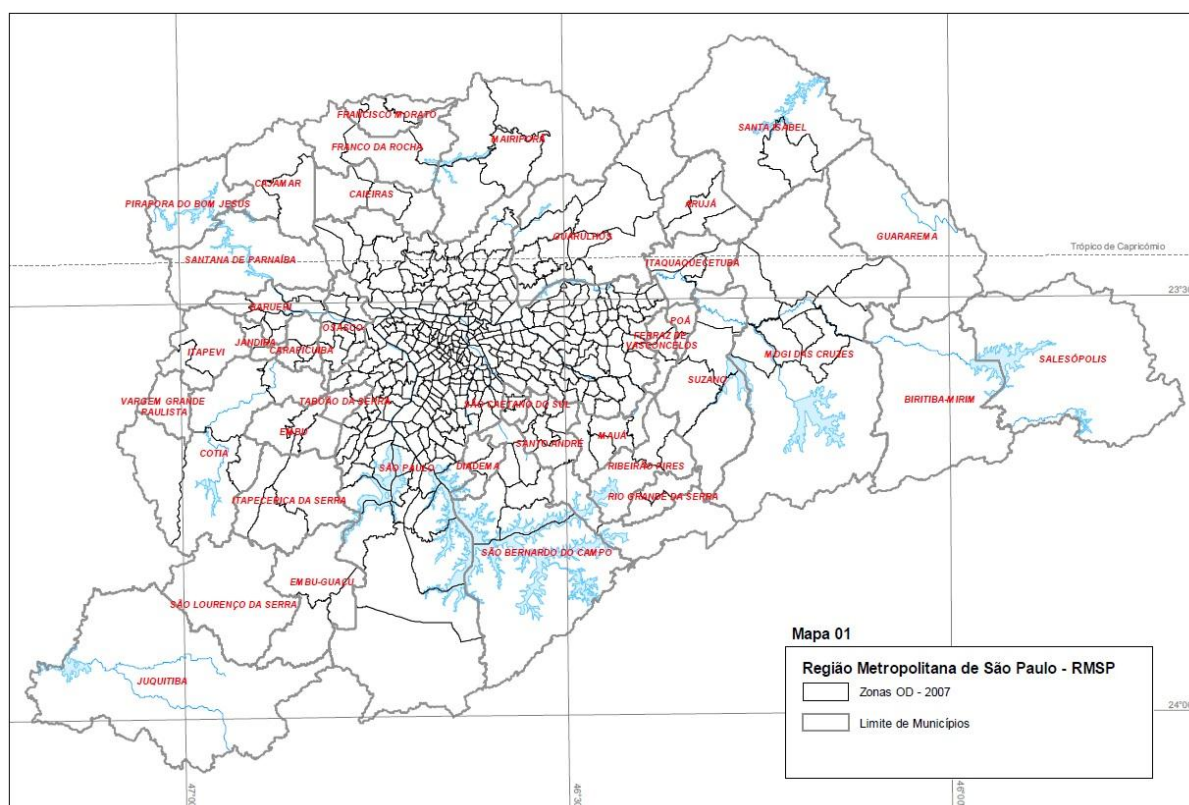
#### ***3.1.1.1 Zoneamento***

Uma característica importante dessa metodologia é a divisão da RMSP em zonas de pesquisa. Uma zona de pesquisa é a unidade básica para o levantamento da origem e destino das viagens, a menor unidade para a qual está garantida a validade estatística das informações. Essa divisão permite a localização espacial da população, assim como uma melhor interpretação visual dos trajetos feitos e regiões onde há maior quantidade de deslocamentos. A figura 3.1 mostra a divisão da RMSP na Pesquisa Origem e Destino 2007.

Os critérios usados para definir a divisão de zonas da Região Metropolitana de São Paulo, na Pesquisa Origem e Destino 2007, foram:

- Comparabilidade com o zoneamento da Pesquisa Origem e Destino 1997;
- Compatibilidade com os limites de municípios e de distritos do município de São Paulo;
- Consideração dos limites de setores censitários de 2000 do IBGE e;
- O sistema de transporte, equipamentos urbanos, barreiras físicas e as áreas vazias.

Figura 3.1 – Zoneamento – Pesquisa Origem e Destino 2007



Fonte: METRÔ (2007)

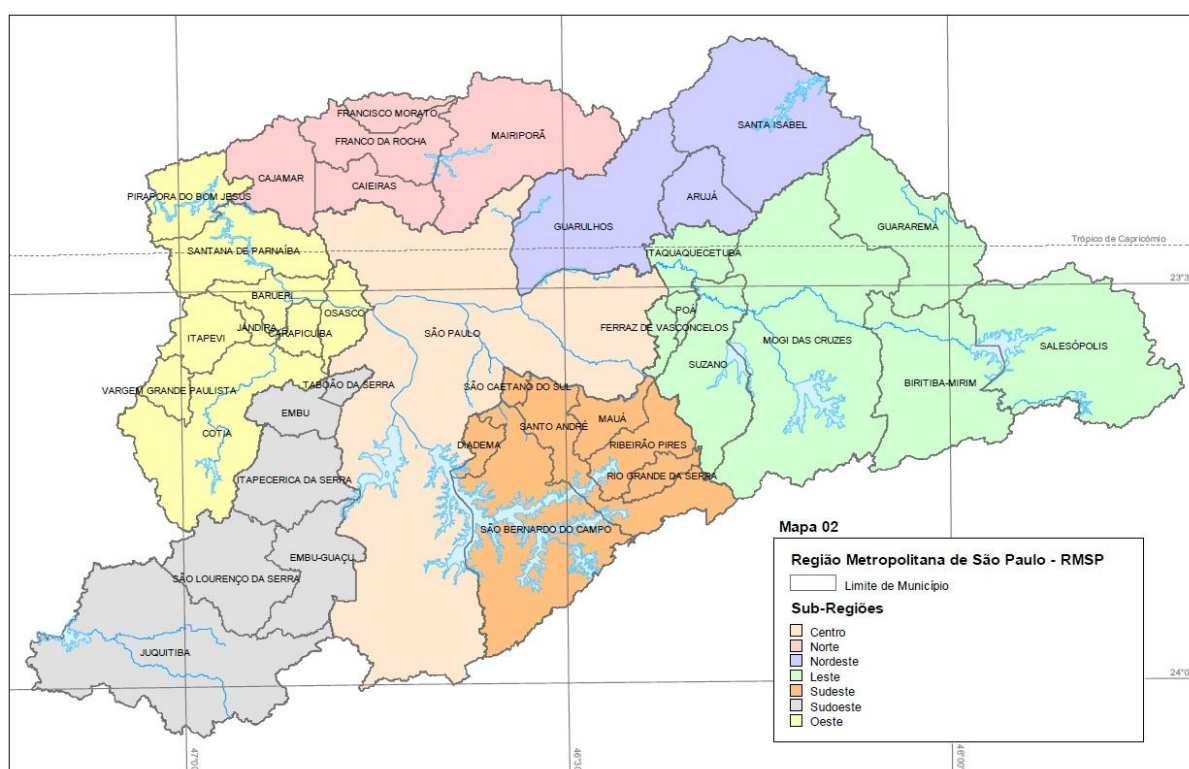
Na pesquisa de 2007, a Região Metropolitana de São Paulo, constituída por 39 municípios, foi dividida em 460 zonas de pesquisa, porquanto na pesquisa realizada em 1997 (também com 39 municípios) a mesma foi dividida em 389 zonas de pesquisa, demonstrando assim a dissemelhança entre os critérios adotados para a partição da mesma região. O município de São Paulo, composto por 96 distritos, foi dividido em 320 zonas de pesquisa.

No relatório de Síntese de Informações os dados foram organizados e apresentados por sub-regiões da Secretaria de Estado dos Transportes Metropolitanos (para melhor compreensão, ver figura 3.2), que englobavam essas zonas de pesquisa, para uma melhor e mais clara apresentação. As sub-regiões foram compostas pelos seguintes municípios:

- **Sub-região Centro:** município de São Paulo.
- **Sub-região Norte:** municípios de Cajamar, Caieiras, Franco da Rocha, Francisco Morato e Mairiporã.
- **Sub-região Nordeste:** municípios de Arujá, Guarulhos e Santa Isabel.
- **Sub-região Leste:** municípios de Biritiba-Mirim, Ferraz de Vasconcelos, Guararema, Itaquaquetuba, Mogi das Cruzes, Poá, Salesópolis e Suzano.

- **Sub-região Sudeste:** municípios de Santo André, São Bernardo do Campo, São Caetano do Sul, Diadema, Mauá, Ribeirão Pires e Rio Grande da Serra.
- **Sub-região Sudoeste:** municípios de Embu, Embu-Guaçu, Itapeçerica da Serra, Juquitiba, São Lourenço da Serra e Taboão da Serra.
- **Sub-região Oeste:** municípios de Barueri, Carapicuíba, Cotia, Itapevi, Jandira, Osasco, Pirapora do Bom Jesus, Santana de Parnaíba e Vargem Grande Paulista.

Figura 3.2 – Sub-Regiões da Secretaria de Estado dos Transportes Metropolitanos - 2007



Fonte: METRÔ (2007)

### 3.1.1.2 Amostra

O plano amostral da Pesquisa Origem/Destino 2007 contou com a colaboração das empresas concessionárias de energia elétrica AES Eletropaulo, Bandeirante Energia S.A. e Elektro Eletricidade e Serviços S.A., pois assim tornou-se possível adotar um sistema de amostragem estratificada a partir de faixas de consumo de energia elétrica, ou seja, as residências eram agrupadas em faixas de acordo com o seu consumo de energia. Foram definidas quatro faixas de consumo de energia elétrica para agrupar as residências: abaixo de 100 kwh/mês, 100 a 200 kwh/mês, 200 a 300 kwh/mês e acima de 300 kwh/mês. Isso

possibilitou que conjuntos homogêneos de residências fossem criados, onde esses conjuntos eram heterogêneos entre si, nos quais a variável característica de cada um era a faixa de consumo.

O motivo da amostragem estratificada ter usado como critério de divisão o consumo de energia é dado pelo fato do consumo de energia elétrica das unidades residências ter correlação com a renda familiar, que por sua vez tem correlação com o número de viagens da família (METRÔ, 2008).

Esse esquema permitiu a obtenção de totais de viagens produzidas em cada zona, com margens de erro inferiores a 5% e nível de confiança de 95%. Em poucas zonas onde não foi possível utilizar o sistema adotou-se a amostragem casual simples, que gerou erros pouco maiores que a estratificada, em torno de 7,5% (METRÔ, 2008).

O tamanho total da amostra foi de 30 mil residências. Feito isso foi definido então o tamanho da amostra necessária para cada zona e o passo seguinte foi o sorteio aleatório dos endereços das unidades consumidoras de energia (residências) para cada faixa de consumo. O georreferenciamento dos endereços adotado na pesquisa permitiu a distribuição da amostra da zona proporcionalmente às populações das subzonas que a compõem (METRÔ, 2008).

Após o sorteio dos domicílios a serem visitados, eles foram comunicados por meio de carta sobre a visita, o objetivo da pesquisa e fornecida uma senha a ser requisitada ao pesquisador no momento da visita, como critério de segurança (METRÔ, 2008).

Na Pesquisa de Origem e Destino de 1997 foi adotada a mesma estrutura para o plano amostral, sendo que as amostras finais foram: 26.278 domicílios, 26.845 famílias, 98.780 pessoas e 163.541 viagens.

### **3.1.2 Execução**

#### **3.1.2.1 Inquérito Residencial**

Após o sorteio das residências e a elaboração dos questionários ocorreu a aplicação deles nos domicílios, entrevistando todos os moradores, com a finalidade de obter dados socioeconômicos, endereços de trabalho e escola e dados sobre as viagens realizadas pelos residentes no dia anterior à entrevista (que sempre se refere a um dia útil da semana) (METRÔ, 2008).



### 3.1.2.1.1 Questionário

O questionário aplicado em cada uma das residências foi dividido em três blocos, numerados de 1 a 3. São eles:

- **Bloco 1: Obtenção dos dados da família.**

Essa parcela do questionário aplica-se apenas ao responsável pela família ou seu cônjuge. Busca conseguir informações sobre posse de bens de consumo durável, condições de acesso à moradia e informações socioeconômicas de todos os moradores.

- **Bloco 2: Caracterização dos dois principais motivos de viagens considerados pela pesquisa – Escola e trabalho.**

Aplicado a todos os moradores da residência que estudam e/ou trabalham, este bloco investiga o endereço da escola e do trabalho de cada morador que declarou anteriormente (bloco 1) que estuda regularmente ou trabalha.

- **Bloco 3: Caracterização das viagens realizadas.**

Essa parte do interrogatório é aplicada a todos os moradores do domicílio que realizaram pelo menos uma viagem no dia anterior ao da entrevista, independente do motivo ou do modo de transporte utilizado (METRÔ, 2007).

### 3.1.2.1.2 Abordagem

O pesquisador, sendo este qualificado e estando atualizado em relação a todos os dados necessários ao questionário, apresenta-se de forma adequada e expõe a Pesquisa e seus motivos, sua importância para o planejamento do transporte metropolitano, que as informações obtidas são confidenciais, não são individualizadas e que é garantido o sigilo dos dados que o entrevistado irá fornecer, buscando despertar o interesse das pessoas e a confiança, para que possam colaborar e fornecer o máximo de dados possível. Considerando que as pessoas entrevistadas são de níveis socioeconômicos e culturais muito diferentes e abstando-se de qualquer comentário em relação às respostas dos entrevistados, sendo eles de qualquer natureza. Qualificado, visita as residências escolhidas pessoalmente, provido das premissas expostas, com o intuito de obter o máximo de informações possíveis (METRÔ, 2007).

### 3.1.2.1.3 Checagem

Após a investigação feita e conferida pelo pesquisador, os questionários passam por um sistema de controle de qualidade que inclui quatro etapas, são elas: verificação pelo supervisor; checagem; crítica e consistência eletrônica. Essas quatro etapas tem como intuito

verificar se o questionário foi preenchido corretamente e analisar a coerência dos dados coletados. Em caso de dúvidas em aberto pode-se, se necessário, retornar ao domicílio pesquisado para completar informações e corrigir erros. Em caso de irregularidade na obtenção das informações o questionário é recusado, havendo necessidade de nova aplicação (METRÔ, 2007).

### 3.1.2.2 Tratamento dos Dados

Após a coleta de dados em campo, eles são digitados on-line, pelos institutos de pesquisa, em programa de digitação, consistência e georreferenciamento, desenvolvido por uma empresa de informática contratada para armazenar os dados na etapa de levantamento de campo. Finalizada a etapa de obtenção de informações e o banco de dados parcialmente consistido e georreferenciado, técnicos procedem à aplicação de fatores de expansão e à atribuição de renda familiar às famílias que não a declararam. Essa atribuição de renda às famílias que não declararam é feita por meio de análise estatística de regressão linear, com a variável renda familiar mensal como função de pontuação obtida pelo Critério Brasil de classificação econômica das famílias.

A aplicação de fatores de expansão aos dados da amostra permite obter estimativas de viagens realizadas na RMSP, balizadas pelos totais de passageiros transportados diariamente pelo METRÔ, CPTM, EMTU e SPTrans (METRÔ, 2008).

## 3.2 Metodologia de Levantamento de Usuários - CPTM

A Companhia Paulista de Trens Metropolitanos, com o intuito de verificar os fluxos de embarque de usuários nas estações em períodos de tempo determinados na rede, utiliza-se de uma metodologia a qual é baseada na geração de um banco de dados que sistematiza, a cada 60 minutos, a quantificação da utilização de Bilhete Único nos bloqueios das estações (CPTM, 2017).

Segundo a CPTM (2017) para o levantamento leva-se em consideração que a utilização do Bilhete Único é amplamente majoritária comparada ao bilhete do tipo Edmonson (que pela configuração tecnológica dos bloqueios, não permite a obtenção dos registros de forma informatizada), sendo assim, procede-se à expansão dos resultados obtidos, através das informações relativas ao Bilhete Único, para o total de embarcados em cada estação, para cada período de tempo.

Em estações onde há a ocorrência de transferências externas oriundas do metrô e pelo grande volume de embarcados, que inviabilizam matematicamente a expansão através do Bilhete Único, procede-se pela utilização das participações percentuais de cada hora, obtidas através da média percentual da transferência dos últimos trabalhos, aplicadas sobre os valores médios de embarcados dessas estações de transferência. Excetuam-se as estações de Itaquera e Tatuapé da Linha 11 – Coral, onde o trabalho é realizado com dados reais de transferências (CPTM, 2017).

Assim, são sistematizados os resultados de embarques hora a hora de todos os bloqueios das estações da CPTM, através da média dos dados obtidos (CPTM, 2017).

## 4 ESTATÍSTICA – DESCRIÇÃO E ANÁLISE

### 4.1 Séries Estatísticas

As séries estatísticas englobam a metodologia para se construir tabelas e gráficos provenientes de variáveis em que há interesse de serem estudadas. Segundo Martins e Domingues (2014, p. 10) “Séries estatísticas são tabelas construídas segundo determinados critérios que as diferenciam e as classificam, para sintetizar um conjunto de dados ou informações”.

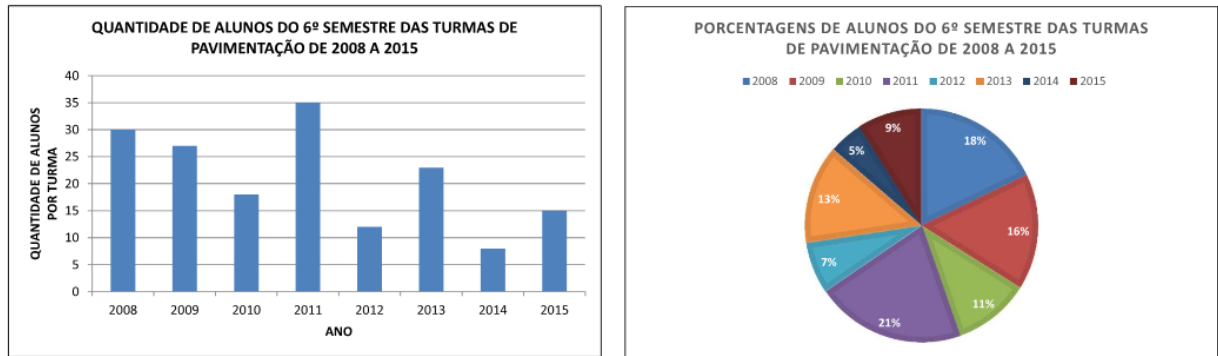
Existem muitas formas de obtenção de dados (publicações, experimentos, observações, dentre outros) e após decidida a forma de tal obtenção torna-se necessário identificar o Nível de Mensuração da variável em questão, pois isso elucidará as possíveis operações aritméticas e, consequentemente, técnica estatística permitidas para se ter a correta análise. (MARTINS; DOMINGUES, 2014)

Dentre os níveis de mensuração pode-se citar o Nível Nominal, Nível Ordinal, Nível Intervalar e Nível de Razão. Em definições simples e básicas de tais níveis pode-se dizer que o Nível Nominal é o ato de nomear uma variável por meio de números ou outros símbolos (por exemplo, sexo, onde masculino é classificado por M e feminino por F). O Nível Ordinal, dada uma variável já classificada (Nível de Mensuração Nominal), em que as relações do tipo maior que, mais alto que, mais bela que, preferível a, são aplicáveis ao conjunto. Ou seja, o Nível Ordinal define uma hierarquização. No Nível intervalar a variável pode assumir várias classes que mantêm uma relação de ordem e intervalos iguais de medição. O zero deste nível é arbitrário. Por exemplo, os calendários onde os anos são contados de forma progressiva a partir de um zero (o qual se sabe não ser o início, o nascimento do tempo), dado que antes dele também há uma escala, mas essa por sua vez é de uma forma gradativa-negativa, ou seja, tem uma contagem inversa à usada após o marco zero. E o Nível de Razão, que além de possuir todas as propriedades do Nível Intervalar, o seu zero é real e absoluto. Dentre esses níveis, as variáveis tratadas como Nominal ou Ordinal são classificadas como qualitativas, enquanto que as classificadas nos níveis Intervalar e/ou de Razão, quantitativas. (MARTINS; DOMINGUES, 2014)

Após a tomada de decisão em relação às premissas do estudo montam-se gráficos e tabelas para a organização e descrição da variável a qual deseja-se saber determinada característica. As descrições gráficas mais comuns usadas em variáveis qualitativas são os gráficos de barras (horizontais ou verticais) e o gráfico em forma de “pizza”. Os dados

oriundos de variáveis quantitativas tem melhor descrição gráfica através de histogramas, onde é apresentada a frequência de observações para cada valor ou conjunto de valores da variável em estudo. A figura 4.1 exemplifica esses gráficos.

Figura 4.1 – Exemplo de Gráficos de Barras Verticais e *Pizza*



Fonte: Elaborada pelo autor (2017)

## 4.2 Descrição Gráfica de Variáveis Quantitativas

Os histogramas são os gráficos mais adequados a uma apresentação baseada em dados provenientes de variáveis quantitativas. Os histogramas apresentam as frequências de observações para cada valor, ou classe de valores, da variável a qual se deseja estudar, objetivando a apresentação de uma forma mais clara e sucinta para que se permita extrair informações sobre o comportamento de tal variável. Como exposto, são apresentadas as frequências de observações, porém, isso não o isenta de retratar outras características que sejam relevantes ao escopo da pesquisa.

É necessária uma organização estruturada antes da apresentação gráfica. A priori organizam-se os dados em ordem crescente (ação denominada *Rol*) identificando, após organizados, a menor e a maior medida do conjunto de dados. A partir disso calcula-se o *Range* através da form. (1):

$$R = X_n - X_1 \quad (1)$$

Onde:

$X_n$  = Maior medida do *Rol*

$X_1$  = Menor medida do *Rol* (MARTINS; DOMINGUES, 2014).

Considerado casos de quantidades elevadas de elementos e para tal tratamento, os dados são agrupados em classes ( $K$ ) com tamanho do intervalo dessas classes ( $h$ ). Esses tamanhos de intervalos podem ser iguais ou desiguais, porém, para maior homogeneidade de apresentação e simples trabalhabilidade, é considerado esse tamanho de intervalo como igual à todas as classes.

Para a escolha do número de classes usa-se a fórmula de Sturges (conquanto há também outras formas para escolha desse número):

$$K \approx 1 + 3,32 \log(n) \quad (2)$$

Onde:

$K$  = Número de classes

$n$  = Número de elementos que se deseja representar

O número  $K$  de classes resultante da fórmula de Sturges, eq. (2), pode acabar resultando em um número racional na sua forma decimal. Como o número de classes deve ser um número natural (inteiro), adota-se o critério:  $0 \leq d \leq 4 \rightarrow$  arredonda-se para o natural anterior;  $5 \leq d \leq 9 \rightarrow$  arredonda-se para o natural posterior, sendo  $d$  o primeiro número após a vírgula (MARTINS; DOMINGUES, 2014).

Após determinado o número de classes calcula-se o tamanho de seus respectivos intervalos através da form. (3):

$$h = R/K \quad (3)$$

Onde:

$h$  = Tamanho do intervalo das classes

Usa-se o mesmo critério de arredondamento empregado na definição do número de classes, caso  $h$  seja um número racional na forma decimal.

São definidos os limites das classes no formato  $[a;b[$ , em que são incluídos todos os elemento maiores ou iguais à  $a$  e menores do que  $b$  ( $a \leq x < b$ ) (MARTINS; DOMINGUES, 2014).

Para uma apresentação satisfatória sobre características do comportamento de uma dada variável, além das frequências (absoluta, relativa, absolutas acumuladas e relativas

acumuladas) pode-se apresentar também suas porcentagens (de elementos e acumulada) e os pontos médios das classes. Quanto às frequências, porcentagens e médias são enunciadas a seguir as formas como são calculadas (MARTINS; DOMINGUES, 2014):

Frequência Absoluta ( $F_i$ ): quantidade de elementos que pertencem à classe  $i$ ;

Frequência Relativa ( $f_i$ ):

$$\frac{F_i}{n} \quad (4)$$

e

$$\frac{F_i}{N} \quad (5)$$

Onde:

$n$  = Número de elementos da amostra

$N$  = Número de elementos da população

A eq. (4) refere-se ao cálculo da Frequência Relativa para amostras, porquanto a eq. (5), para populações.

Porcentagem de elementos de uma dada classe  $i$  ( $f_i(\%)$ ):

$$f_i \times 100 \quad (6)$$

Frequência Absoluta acumulada ( $F_{ac}$ ): soma sequencial, do menor para o maior, dos valores da Frequência Absoluta de cada classe;

Frequência Relativa Acumulada ( $f_{ac}$ ): soma sequencial, do menor para o maior, dos valores da Frequência Relativa;

Porcentagem de frequências acumuladas ( $\%ac$ ):

$$f_{ac} \times 100 \quad (7)$$

Média aritmética da classe ( $x_i$ ):

$$\left(\frac{L_s - L_i}{2}\right) + L_i \quad (8)$$

Onde:

$L_i$  = Limite inferior da classe

$L_s$  = Limite superior da classe

Após os cálculos monta-se uma tabela de distribuição de frequências, alimentada pelos resultados das classes e suas frequências, como apresentado pela figura 4.2 (MARTINS; DOMINGUES, 2014).

Figura 4.2 – Exemplo de uma Tabela de Distribuição de Frequências

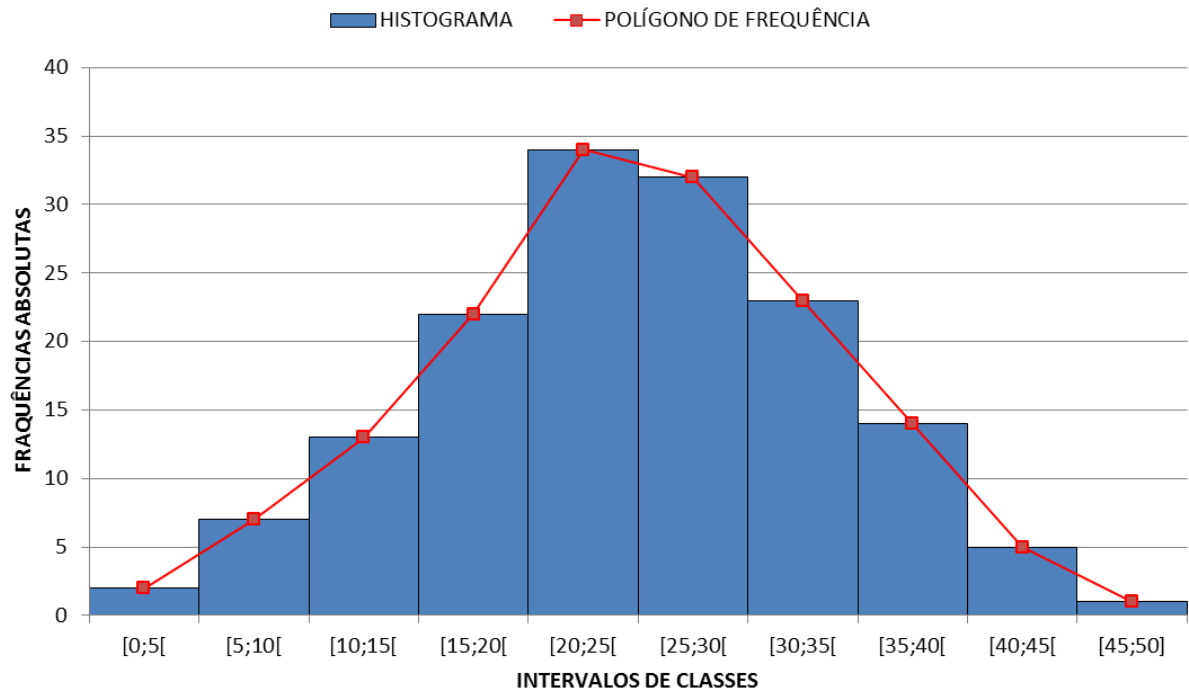
CLASSES	INTERVALOS DAS CLASSES	$Fi$	$fi$	$fi$ (%)	$Fac$	$fac$	$\%ac$	$x_i$
1	[0;5[	2	0,01	1	2	0,01	1	2,5
2	[5;10[	7	0,05	5	9	0,06	6	7,5
3	[10;15[	13	0,08	8	22	0,14	14	12,5
4	[15;20[	22	0,14	14	44	0,28	28	17,5
5	[20;25[	34	0,22	22	78	0,50	50	22,5
6	[25;30[	32	0,21	21	110	0,71	71	27,5
7	[30;35[	23	0,15	15	133	0,86	86	32,5
8	[35;40[	14	0,09	9	147	0,95	95	37,5
9	[40;45[	5	0,04	4	152	0,99	99	42,5
10	[45;50]	1	0,01	1	153	1,00	100	47,5
SOMAS		153	1,00	100				

Fonte: Elaborada pelo autor (2017)

Seguido à organização e concatenação dos dados oriundos de uma variável quantitativa constrói-se o histograma e o polígono de frequência. O polígono de frequência é um gráfico que une os pontos médios das classes. O gráfico 4.1 apresenta um histograma e o polígono de frequência baseados na figura 4.2 (PAULETTE, 2014?).



Figura 4.3 – Exemplo de um Histograma junto ao Polígono de Frequência



Fonte: Elaborada pelo autor (2017)

Logo ao término da organização e apresentação dos dados o pesquisador deduz as informações que tem como intento, sendo essas características da variável em estudo.

### 4.3 Medidas Descritivas

Integrantes da estatística descritiva, as medidas descritivas são métodos usados para descrever e analisar fenômenos coletivos. Elas podem ser divididas em medidas de:

Tendência Central: oferecem o posicionamento da distribuição dos valores;

Ordenação: usadas para secção ordenada em uma série;

Dispersão: avaliam o grau de variabilidade em relação à média;

Assimetria: medem o grau de afastamento em relação à unidade de simetria;

Curtose: Identificam o grau de achatamento de uma série. (MARTINS; DOMINGUES, 2014)

### 4.3.1 *Tendência Central*

A determinação e o cálculo de medidas que ofereçam o posicionamento da distribuição dos valores de uma variável que desejamos analisar. Dito de outra forma, são medidas utilizadas para representar fenômenos coletivos através de um único valor, fornecendo uma ideia geral a respeito do fato ou fenômeno analisado. (MARTINS; DOMINGUES, 2014, p. 30)

Tais medidas podem dividir-se em matemáticas ou não matemáticas.

#### 4.3.1.1 Média Aritmética

Medida matemática e possivelmente a mais comum e conhecida das medidas de posição, a média aritmética tem uso generalizado e aplica-se a uma vasta gama de situações. Um fato que se deve levar em consideração é que ela sofre grande influência de todos os valores da série, em função disso deve atentar-se ao seu uso. A média aritmética é representada pelo símbolo  $\bar{X}$  quando representa uma amostra e  $\mu$  quando uma população. A média aritmética de uma amostra é dada através do quociente da soma dos valores das observações feitas e o número de observações, onde o primeiro é o numerador e o segundo o denominador. Apresenta-se através da form. (9):

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n} \quad (9)$$

Onde:

$x_i$  = Valor individual da observação feita

$n$  = Número de observações

A média populacional é calculada de forma análoga, visto que a diferença é apenas conceitual. (MARTINS; DOMINGUES, 2014)

#### 4.3.1.2 Média Aritmética Ponderada

Há eventos onde além do levantamento de dados ter um amplo espaço amostral, boa parte dos elementos pertencentes a esse conjunto repetem-se. Em outras ocasiões decorre que é necessário apresentar a média de uma grande quantidade de dados já tabulados e

divididos em classes. Em ocasiões onde há uma grande quantidade de elementos e repetições usa-se a média aritmética ponderada (amostral ou populacional, divergentes apenas no conceito e símbolos), dada pelo quociente do somatório do produto dos valores das observações feitas e suas respectivas frequências pelo número de observações. Apresenta-se através da Form. (10):

$$\bar{X} = \frac{x_1F_1 + x_2F_2 + \dots + x_nF_n}{n} = \frac{\sum x_iF_i}{n} \quad (10)$$

Onde:

$F_i$  = Frequência do valor individual da observação feita (MARTINS; DOMINGUES, 2014).

#### 4.3.1.3 Mediana

A mediana destaca o elemento central de um conjunto de dados, é o elemento que divide a série em duas partes iguais. Considerada uma medida de posição importante, a mediana deixa 50% dos elementos de uma série abaixo do seu valor e 50% acima. Em suma, quando os elementos de uma série são colocados em ordem crescente, a mediana divide a amostra (ou população) em duas partes iguais.

Pelo fato da mediana localizar-se no centro da série, em termos de quantidade de elementos, ela não sofre interferência dos valores extremos, levando a considera-la uma medida útil e interessante para a análise e interpretação de fatos socioeconômicos. (MARTINS; DOMINGUES, 2014)

Para o cálculo é necessário dividir as possíveis situações de estudo em dois grupos, quando  $n$  (número de elementos) é ímpar e quando é par (PAULETTE, 2014?).

Quando o número de elementos é ímpar, para localizar posição da mediana usa-se a form. (11):

$$M_e = \frac{n + 1}{2} \quad (11)$$

O valor da mediana é igual ao valor do elemento que se encontra na posição obtida como resultado.

Quando o número de elementos é par, a mediana ocupa dois elementos centrais ( $M_{e_1}$  e  $M_{e_2}$ ). Para o cálculo das suas posições são usadas a form. (12) e form. (13):

$$M_{e_1} = \frac{n}{2} \quad (12)$$

e

$$M_{e_2} = \frac{n}{2} + 1 \quad (13)$$

Uma vez obtidas as posições, o valor da mediana ( $m(x)$ ) é definido como a média aritmética dos valores encontrados nas posições  $M_{e_1}$  e  $M_{e_2}$ :

$$m(x) = \frac{M_{e_1(valor)} + M_{e_2(valor)}}{2} \quad (14)$$

Onde:

$M_{e_n(valor)}$  = Valor encontrado na posição  $n$  (PAULETTE, 2014?).

#### 4.3.1.4 Moda

A moda ( $M_o$ ), em um conjunto de dados, é o elemento que apresenta a maior frequência. Por essa sua característica é facilmente identificada, estando em uma série simples ou agrupada (MARTINS; DOMINGUES, 2014).

### 4.3.2 Ordenação

Utilizadas para fazer cortes ordenados em uma série, as medidas de ordenação (ou separatrizes) visam identificar características relevantes dentro do conjunto. São elas a Mediana, Quartis, Decis e Percentis, que dividem a série, respectivamente, em duas partes (50% e 50%), quatro partes (25% em 25%), dez partes (10% em 10%) e cem partes (1% em 1%) (MARTINS; DOMINGUES, 2014).

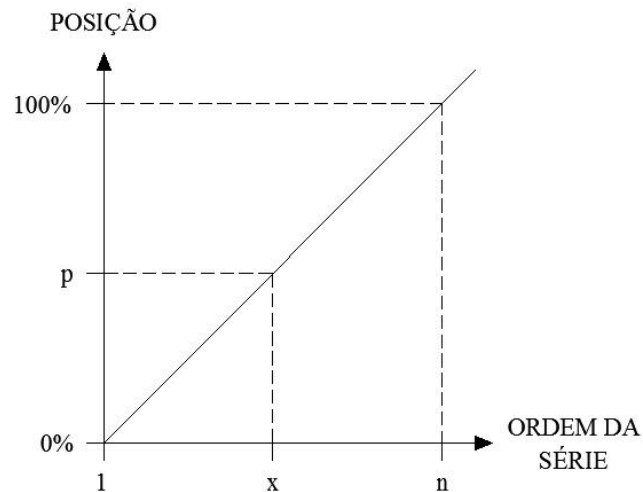
Os Percentis, por efetuarem cortes de 1% em 1%, podem substituir todas as outras separatrizes. O Percentil 50 (P50) equivale à mediana, os Percentis P25, P50 e P75 equivalem aos Quartis e os Percentis P10, P20, P30, P40, P50, P60, P70, P80 e P90 equivalem aos Decis (MARTINS; DOMINGUES, 2014).

Para o cálculo dos Percentis a série deve ser previamente ordenada de forma crescente, fazendo com que o menor valor da série assuma a posição 1 e o maior valor a

posição  $n$  e, após ordenada, enumerar as posições dos dados ordenados, associando os elementos da série a números naturais (MARTINS; DOMINGUES, 2014).

Como os Percentis dividem a série em 100 partes iguais de 1%, quando ordenada, 1 a  $n$  equivale 0% a 100% em termos de posicionamento. A figura 4.3 elucida a situação:

Figura 4.4 – Percentil – Equivalência em Termos de Posicionamento



Fonte: Elaborada pelo autor (2017)

A partir dessa equivalência é possível construir a relação:

$$\frac{n - 1}{100\% - 0\%} = \frac{x - 1}{p - 0\%} \quad (15)$$

Isolando o  $x$  com o intuito de saber a posição de uma determinada observação e a partir dela encontrar o respectivo valor do percentil, é obtida a form. (16):

$$x = \left[ (n - 1) \frac{p}{100} \right] + 1 \quad (16)$$

Onde:

$x$  = Ordem de uma determinada observação

$n$  = Número de observações

$p$  = Percentil desejado expresso em %

Caso a posição do elemento resultante seja um número racional na sua forma decimal, deve-se fazer a interpolação linear, através do quociente da diferença entre o valor da

posição imediatamente posterior e inteiro ao resultante e ele, sobre o valor dessa posição imediatamente posterior e  $x$  (que será a resultante real) (MARTINS; DOMINGUES, 2014).

### 4.3.3 Dispersão

As medidas de dispersão são utilizadas para avaliar o grau de dispersão, ou variabilidade, dos valores em torno da média. Elas demonstram a representatividade da média (MARTINS; DOMINGUES, 2014).

#### 4.3.3.1 Amplitude Total

A amplitude total ( $R$ ) é uma medida dada entre o maior e o menor valor da série ( $R = X_{max} - X_{min}$ , respectivamente). Esta é uma medida de dispersão limitada, pois, como utiliza apenas os valores extremos, não capta possíveis variações entre esses limites (MARTINS; DOMINGUES, 2014).

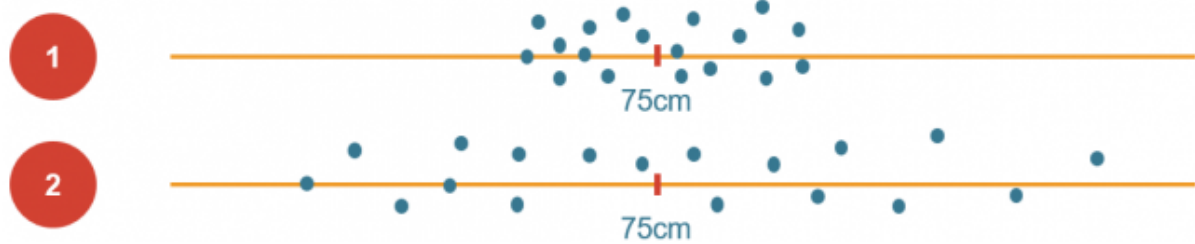
#### 4.3.3.2 Variância

No caso de medição da dispersão dos dados em relação à média, fica clara a necessidade de analisar os desvios ( $d_i$ ) de cada valor ( $x$ ) em relação a ela. Quando os desvios são altos, significa que há grande dispersão entre os dados e a média, assim como quando são baixos, há pouca dispersão. Essa informação denota um grau de confiabilidade na média (MARTINS; DOMINGUES, 2014).

Fica claro que o somatório dos desvios em relação à média é zero, portanto, para calcular a variância ( $S^2$ ) consideram-se os quadrados dos desvios ( $d_i^2$ ) (MARTINS; DOMINGUES, 2014).

Abaixo, a figura 4.4 mostra duas linhas com a mesma média (75 cm), porém, com grande diferença em seus desvios. A primeira com desvios baixos (pontos mais próximos da média) e a segunda com desvios mais altos (pontos mais distantes da média).

Figura 4.5 – Comparação de duas linhas com mesma média e desvios distintos



Fonte: Portal Action (2017)

A primeira linha, por ter desvios menores, mostra que os pontos terão medidas próximas à média (75 cm).

A variância amostral é calculada através da form. (17):

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1} \quad (17)$$

Onde:

$\bar{X}$  = Média amostral

#### 4.3.3.3 Desvio Padrão

A variância é obtida pela soma dos quadrados dos desvios em relação à média, portanto, a variância é expressa pelo quadrado da unidade de medida da variável que está sendo estudada. Para uma melhor interpretação da dispersão de uma variável, calcula-se o desvio padrão amostral ( $S$ ) (ou populacional, caso seja em relação a uma população), que é calculado através da raiz quadrada da variância, como mostra a form. (18) (MARTINS; DOMINGUES, 2014):

$$S = \sqrt{S^2} \quad (18)$$

O desvio padrão indica a dispersão média absoluta dos dados em torno da média aritmética. Quanto menor o desvio padrão de um processo produtivo, menor será a variabilidade apresentada no produto final e, portanto, maior qualidade terá o produto.

Segundo o **Teorema de Tchebycheff**, para qualquer distribuição com média amostral  $\bar{X}$  e desvio padrão amostral  $S$ , o intervalo  $\bar{X} \pm 2S$  contém, no mínimo, 75% de todas as observações, porquanto o intervalo  $\bar{X} \pm 3S$  contém, no mínimo, 89% de todas as observações (MARTINS; DOMINGUES, 2014).

#### 4.3.3.4 Coeficiente de Variação de Pearson

O Coeficiente de Variação de Pearson (CV) é uma **medida relativa de dispersão**. É obtido através da form. (19):

$$CV = \frac{S}{\bar{X}} \times 100 \quad (19)$$

Abaixo são demonstradas algumas regras empíricas para a interpretação do Coeficiente de Variação de Pearson:

- $CV < 15\% \rightarrow$  Há baixa dispersão. Boa representatividade para a média aritmética como medida de posição;
- $15\% \leq CV < 30\% \rightarrow$  Há média dispersão. A representatividade da média aritmética como medida de posição é regular;
- $CV \geq 30\% \rightarrow$  Há elevada dispersão. A representatividade da média aritmética como medida de posição é ruim (MARTINS; DOMINGUES, 2014).

#### 4.3.3.5 Escore Padronizado

O Escore Padronizado ( $Z_i$ ) é uma medida relativa de dispersão. Dado uma medida  $x_i$ , tem-se a form. (20):

$$Z_i = \frac{(x_i - \bar{X})}{S} \quad (20)$$

Um Escore Padronizado negativo mostra que a observação  $x_i$  está à esquerda da média, assim como um  $Z_i$  positivo indica que a observação está à direita da média (MARTINS; DOMINGUES, 2014).



#### 4.3.3.6 Outliers

Nos trabalhos de coleta de dados existe a chance de ocorrerem observações que fogem das dimensões esperadas. Esses valores atípicos são denominados *Outliers*.

A presença de dados que fogem dos extremos esperados provoca distorções na análise estatística dos resultados (MARTINS; DOMINGUES, 2014). Sendo assim, é pertinente detectar a presença de *Outliers* em um conjunto de dados antes mesmo de iniciar a análise.

Existem alguns métodos para a detecção desses dados atípicos. Aqui serão apresentados dois deles, cujos parâmetros usados já foram apresentados anteriormente neste trabalho.

É possível identificar os *Outliers* através do Escore Padronizado. Os dados cujos Escores, em módulo, sejam maiores do que três, são considerados *Outliers*.

Outra forma é através do *Boxplot*. Primeiro calcula-se o P25 e o P75 da série. A diferença entre P75 e P25 é chamada de Intervalo Interquartílico (form. (21)).

$$I = P75 - P25 \quad (21)$$

Nesse método é possível separar os *Outliers* em dois grupos: os *Outliers* Moderados e os *Outliers* Severos (MARTINS; DOMINGUES, 2014).

São considerados *Outliers* Moderados:

- No limite inferior ( $L_i$ ): entre  $(P25 - 3 \cdot I)$  e  $(P25 - 1,5 \cdot I)$ ;
- No limite superior ( $L_s$ ): entre  $(P75 + 1,5 \cdot I)$  e  $(P75 + 3 \cdot I)$ .

São considerados *Outliers* Severos:

- No limite inferior ( $L_i$ ): abaixo de  $(P25 - 3 \cdot I)$
- No limite superior ( $L_s$ ): acima de  $(P75 + 3 \cdot I)$

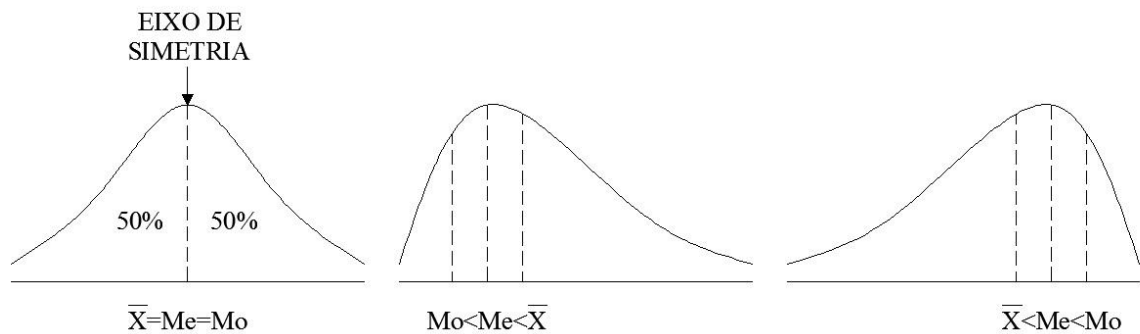
Este último método é mais conveniente que o primeiro apresentado, pois os Percentis não são afetados pela presença de valores extremos na série.

Apesar das formas de identificar os *Outliers*, nenhuma delas indica de forma absoluta a necessidade ou não de eliminar os dados da série em análise. Essa função cabe ao analista (MARTINS; DOMINGUES, 2014).

#### 4.3.4 Assimetria

Assimetria é o grau de afastamento, de uma distribuição, da unidade de simetria. Em uma distribuição simétrica os valores da média, mediana e moda são iguais ( $\bar{X} = M_e = M_o$ ). Em uma distribuição assimétrica positiva, o valor da moda é menor que o da mediana, que, por sua vez, é menor que o da média ( $M_o < M_e < \bar{X}$ ). Já na distribuição assimétrica negativa o valor da média é menor que o da mediana, que é menor que o da moda ( $\bar{X} < M_e < M_o$ ). A figura 4.5 demonstra tais situações (MARTINS; DOMINGUES, 2014).

Figura 4.6 – Distribuições simétrica, assimétrica positiva e assimétrica negativa



Fonte: Elaborada pelo autor (2017)

Existem várias fórmulas para o coeficiente de assimetria, aqui será apresentada a intitulada **Coeficiente de Pearson ( $A_s$ )**.

$$A_s = \frac{(\bar{X} - M_o)}{S} \quad (22)$$

A partir do resultado do Coeficiente de Pearson, pode-se obter duas interpretações, quanto ao sinal e quanto à intensidade.

Interpretação quanto ao sinal:

- $A_s = 0 \rightarrow$  Diz-se que a distribuição é simétrica;
- $A_s > 0 \rightarrow$  Diz-se que a distribuição é assimétrica positiva;
- $A_s < 0 \rightarrow$  Diz-se que a distribuição é assimétrica negativa.

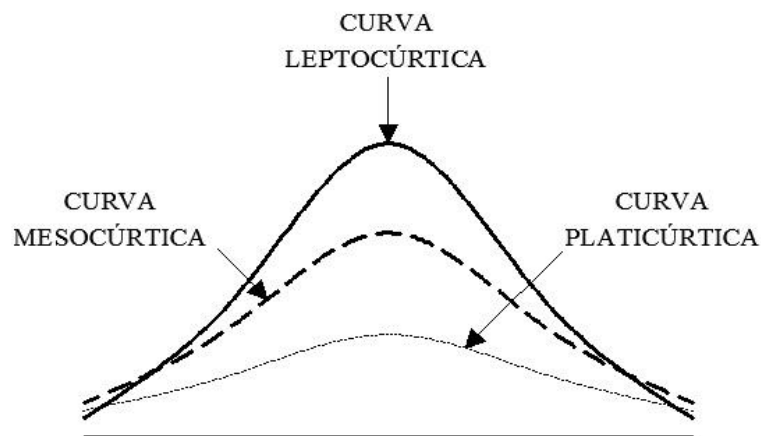
Interpretação quanto à intensidade (resultado em módulo):

- $0 < A_s \leq 1 \rightarrow$  Assimetria fraca;
- $A_s > 1 \rightarrow$  assimetria forte.

### 4.3.5 Curtose

A medida de Curtose visa identificar o grau de achatamento de uma série estatística. Neste caso existem três possibilidades, que a curva seja Platicúrtica, Mesocúrtica ou Leptocúrtica. A figura 4.6 elucida tais possibilidades (MARTINS; DOMINGUES, 2014).

Figura 4.7 – Curvas Platicúrtica, Mesocúrtica e Leptocúrtica



Fonte: Elaborada pelo autor (2017)

Para medir o grau de curtose, pode-se usar o **Coefficiente de Kelley (K)**, conforme form. (23):

$$K = \frac{P75 - P25}{2(P90 - P10)} \quad (23)$$

Quanto a sua interpretação:

- $K=0,263 \rightarrow$  Curva Mesocúrtica;
- $K<0,263 \rightarrow$  Curva Leptocúrtica;
- $K>0,263 \rightarrow$  Curva Platicúrtica.

## **5 METODOLOGIA**

A metodologia empregada dividiu-se em três partes. Na primeira os dados foram obtidos, separados e organizados, possibilitando e facilitando seu tratamento. Na segunda, cálculos das medidas descritivas da série foram engendrados para que os dados fossem analisados e validados. Na terceira e última etapa foram examinadas as demandas médias de usuários, a forma de suprimento que está sendo usada para atendê-las e a verificação se são ou não suficientes.

### **5.1 Levantamento e Organização dos Dados**

A metodologia usa como premissas dados fornecidos pela própria companhia responsável pelo transporte ferroviário de passageiros. A Figura 5.1 apresenta um exemplo de fornecimento desses dados.

As informações necessárias foram:

- As quantidades de passageiros que adentram as estações, por intervalos de horário, por dia;
- Histórico desses levantamentos diários ao longo do tempo, ou seja, os dados de semanas e meses também do ano anterior, em uma quantidade significativa para gerar um tratamento estatístico com uma boa margem de confiança e;
- Os tipos de trens adotados na época do estudo, suas capacidades e os tempos de ciclo entre eles, por intervalos de horário, em dia útil.

Figura 5.1 – Média de embarcados por intervalos de tempo na estação de Itaim Paulista na Linha 12 – Safira da CPTM em dias úteis em uma situação de normalidade



Fonte: CPTM (2017)

A Companhia Paulista de Trens Metropolitanos forneceu, para cada hora, as médias de passageiros que embarcaram nos dias úteis e finais de semana em cada uma das estações, durante todo o horário de funcionamento. Essas médias são relativas aos meses de agosto, setembro, outubro e novembro de 2016 e março, abril, maio e junho de 2017. São meses estes onde há simultaneidade entre os períodos letivos acadêmicos e de ofício, retratando assim uma situação de normalidade, em grande parte do ano, no modal.

Devido à falta de algumas informações e para suprir tais mínguas a pesquisa usou dados aproximados, balizados pelos informes da CPTM que, mesmo não sendo os dados exatos, conduziram para que não houvessem erros grosseiros e garantisse a demonstração da metodologia proposta.

Ao exposto nos dois últimos parágrafos, foi necessário simular a quantidade de passageiros que embarcaram no período de pico (apresentado a seguir) para cada um dos dias úteis dos meses apresentados. Para tal feito foram gerados números aleatórios em um intervalo fechado, os quais foram associados cada qual a um dia. Os limites desse intervalo foram definidos a partir da média das diferenças entre cada uma das médias mensais de passageiros fornecidas pela CPTM. Obtida a média das diferenças, esta foi somada à maior média, e subtraída à menor, fornecidas pela Companhia, obtendo assim os limites superior e

inferior do intervalo onde os números foram gerados. Procedeu-se dessa maneira visto que o necessário eram as quantidades de passageiros que adentraram as estações, por intervalos de horário, por dia, e o que foi fornecido foram as médias mensais para esses intervalos de horário.

No início dessa etapa foram analisadas as quantidades estratificadas das informações providas pela Companhia, com o intuito de identificar em qual intervalo ocorreu a maior quantidade de usuários que adentraram o sistema. Identificado, esse intervalo foi separado dos demais para ser tratado.

Para a apresentação da metodologia foi escolhido o período de pico da estação Itaim Paulista, da Linha 12 – Safira da CPTM. Período este compreendido entre as seis e sete horas da manhã.

Com as médias de passageiros que entraram na estação, entre as seis e sete horas da manhã, nos meses supracitados, foram geradas quantidades de utentes para cada dia útil, simulando um levantamento histórico desse intervalo de pico em um expressivo período de dias, assim reproduzindo as quantidades de passageiros que ingressavam no sistema neste intervalo. Obtidas essas quantidades diárias de usuários no intervalo de pico, deu-se início à organização dessas informações de forma a possibilitar e facilitar os procedimentos que foram realizados.

As quantidades foram reunidas e a partir desse conjunto de amostras simuladas foi gerado um *Rol* para organizá-las em ordem crescente. Ordenadas, as amostras foram associadas a números naturais e em seguida calculado o *Range* dessa amostragem.

Conhecido o valor do *Range* seguiram-se os cálculos do número de classes e o tamanho do intervalo dessas classes, as quais tiveram os dados organizados por intervalos de valores. Em seguida foi elaborada a tabela de distribuição de frequências e a partir desta gerado o histograma para uma melhor interpretação gráfica do comportamento das frequências absolutas dos valores da variável usuários que ingressaram no sistema.

Para a geração das tabelas e gráficos neste trabalho utilizou-se o software Excel, desenvolvido pela empresa Microsoft, comumente usado para elaboração de planilhas eletrônicas e atividades afins.

## 5.2 Análise e Validação das Informações

Na segunda etapa foram calculadas as medidas descritivas das informações tratadas no primeiro ciclo, verificadas suas precisões e analisadas se eram fidedignas ou não para a procedência do estudo, visando exatidão para uma simulação mais próxima da realidade da rotina do modal.

Após reunidas as quantidades de utentes que adentraram o sistema ao longo do tempo, foram calculadas a média aritmética, a mediana e identificada a moda dessa amostra, objetivando possuir o entendimento do comportamento médio das entradas, o valor central, ou seja, que divide a série em estudo em 50% dos valores abaixo dele e 50% acima, e o valor que apresentou-se com maior frequência ao longo desse histórico, respectivamente.

Logradas as medidas susoditas, procedeu-se com o cálculo da variância e a partir dela o desvio padrão, a fim de medir a dispersão dos valores individuais em relação à média, exprimindo um nível de confiança para a adoção dessa medida.

Com o objetivo de ter uma descrição mais minuciosa sobre os dados que são as proposições nas quais baseiam-se os resultados dos estudos, foi calculado o escore padronizado, para mostrar onde se encontra um dado em relação à média, e foi realizado um teste lógico para detecção de *outliers*, valores que fogem dos extremos e podem causar distorções na análise dos resultados. A figura 5.2 mostra uma tabela de verificação dos dados primários, baseada no supradito.

Figura 5.2 – Verificação dos dados primários

NÍVEL NOMINAL	NÍVEL ORDINAL	ESCORE PADRONIZADO	POSIÇÃO RELATIVA À MÉDIA	OUTLIER
1	5.620	-1,78	Esquerda	Não
2	5.625	-1,75	Esquerda	Não
3	5.629	-1,72	Esquerda	Não
4	5.634	-1,69	Esquerda	Não
5	5.634	-1,69	Esquerda	Não
6	5.636	-1,68	Esquerda	Não
7	5.637	-1,67	Esquerda	Não
8	5.639	-1,66	Esquerda	Não
...				
159	6.135	1,54	Direita	Não
160	6.143	1,59	Direita	Não
161	6.145	1,60	Direita	Não
162	6.146	1,61	Direita	Não
163	6.147	1,61	Direita	Não
164	6.149	1,63	Direita	Não
165	6.150	1,63	Direita	Não
166	6.151	1,64	Direita	Não
167	6.157	1,68	Direita	Não

Fonte: Elaborada pelo autor (2017)

A apresentação do *outlier* junto ao escore padronizado aponta se esse valor está excedendo os menores ou maiores valores do grupo.

Seguidamente foi calculado o Coeficiente de Variação de Pearson, que demonstrou o quanto a média é confiável.

Após, foram calculadas as posições dos percentis P10, P25, P75 e P90 no *Rol* ordenado e verificados os valores que se encontravam nessas posições, com o intuito de calcular a Curtose, a medida de assimetria que descreveu o grau de achatamento da série, auxiliando na interpretação dela, e o Coeficiente de Pearson em relação à assimetria que a curva gerada pela série possui, mostrando seu comportamento em relação a sua unidade de simetria.

As informações obtidas resultantes da ordenação e cálculos foram organizadas para que se pudesse ter uma noção ampla do comportamento de entradas de usuários na estação no intervalo de tempo estudado e validá-las para a próxima etapa (ver APÊNDICE A).

### **5.3 Exame da Demanda e Suprimento**

Terminadas as duas primeiras etapas da pesquisa, ambas foram executadas para as demais estações da linha férrea em estudo, no mesmo horário determinado para a primeira, excetuando-se a estação terminal, onde há a grande maioria dos desembarques.

Em posse das médias adotadas de passageiros que entraram no sistema no intervalo estipulado, por estação, foi feita uma tabela resumo de tais informações, contendo a identificação de cada estação, as quantidades de usuários que adentram nela, a soma parcial do sistema ao longo da linha e o total de passageiros.

Seguiu-se então à análise dos tipos de trens que estão sendo adotados para suprir as demandas das estações, suas capacidades e os tempos de ciclo que há entre cada um deles, no intervalo em estudo, objetivando o exame da demanda baseada nas análises feitas e a forma de suprimento empregada.

Feito o levantamento, foi projetada uma tabela resumo compreendendo o intervalo programado entre trens, quantidade adotada no intervalo, os modelos de composições usados, suas capacidades, respectivas quantidades empregadas e a oferta de suprimento dessa frota.

Após obtidas a quantidade de usuários que entram no sistema naquele intervalo de tempo e a oferta de suprimento oferecida pela frota adotada, verificou-se se esta última é compatível e supri a necessidade exigida pela quantidade de utentes ou não, se há a urgência de tomar alguma solução quanto a isso e, caso sim, onde ela deve operar.



## 6 RESULTADOS

A apresentação dos resultados neste capítulo visa, além de expor o corolário, melhor elucidar o apresentado na metodologia.

A Companhia Paulista de Trens Metropolitanos forneceu as médias mensais de usuários que entraram nas estações, para cada hora, durante todo o período de funcionamento destas, para todas as estações. Essas médias são relativas aos meses de agosto, setembro, outubro e novembro de 2016 e março, abril, maio e junho de 2017, como indicado no primeiro sub capítulo da metodologia.

Por ser uma grande gama de informações, será apresentado apenas um dos conjuntos de dados fornecidos, sendo cada conjunto relativo a uma estação, visto que os outros utilizam-se da mesma forma de exibição. Este pode ser visto através do ANEXO A, que mostra a forma como foram cedidos os dados.

Na metodologia foi usado o período de pico da estação de Itaim Paulista, assim como a linha a qual pertence. Este período é compreendido entre seis e sete horas da manhã.

Segundo as provisões da CPTM, as médias de usuários que adentravam o sistema através da estação Itaim Paulista no horário de pico foram as apresentadas na figura 6.1. Foram fornecidas quatro médias, cada qual associada a dois meses.

Figura 6.1 – Médias de embarcados por hora em dias úteis

HORÁRIO		EMBARCADOS - DIA ÚTIL							
		Ago_Set 2016	%	Out_Nov 2016	%	Mar_Abr 2017	%	Mai_Jun 2017	%
3h	4h	134	0,4	138	0,4	138	0,4	127	0,4
4h	5h	1.681	5,5	1.717	5,5	1.684	5,4	1.710	5,5
5h	6h	4.036	13,2	4.046	12,9	4.015	12,8	3.957	12,7
6h	7h	5.867	19,2	5.948	19,0	6.055	19,3	5.726	18,4
7h	8h	4.601	15,0	4.713	15,0	4.598	14,7	4.625	14,9
8h	9h	2.486	8,1	2.698	8,6	2.627	8,4	2.709	8,7
9h	10h	1.430	4,7	1.542	4,9	1.496	4,8	1.508	4,9
10h	11h	1.212	4,0	1.296	4,1	1.248	4,0	1.305	4,2
11h	12h	1.358	4,4	1.407	4,5	1.417	4,5	1.420	4,6
12h	13h	1.647	5,4	1.708	5,5	1.671	5,3	1.727	5,6
13h	14h	1.146	3,7	1.144	3,7	1.156	3,7	1.192	3,8
14h	15h	644	2,1	654	2,1	704	2,2	734	2,4

Fonte: CPTM (2017)

Como já exposto os motivos na primeira etapa da metodologia, foi então calculada a diferença entre cada uma dessas médias e em seguida a média das diferenças. Obtida esta última, foi subtraída à menor e somada à maior das médias mensais, definindo assim os limites inferior e superior do intervalo no qual foram gerados os números aleatórios que simularam as quantidades de passageiros no horário de pico para cada dia útil dos meses cujas médias de utentes foram fornecidas. A figura 6.2 apresenta os resultados desses cálculos.

Tabela 6.1 – Definição dos limites do intervalo onde foram gerados os números aleatórios

<b>MÊS</b>	<b>MÉDIA MENSAL DE PASSAGEIROS</b>	<b>DIFERENÇA</b>
Mai. - Jun.	5.726	-
Ago. - Set.	5.867	141
Out. - Nov.	5.948	81
Mar. - Abr.	6.055	107
<b>MÉDIA DAS DIFERENÇAS</b>		<b>110</b>
<b>LIMITE INFERIOR</b>		<b>5.616</b>
<b>LIMITE SUPERIOR</b>		<b>6.165</b>

Fonte: Elaborada pelo autor (2017)

Após, foi verificada a quantidade de dias úteis nesses meses, totalizando 167 dias. Sendo assim, foram gerados 167 números aleatórios dentro do intervalo já definido, sendo eles associados a cada dia, para assim representarem as quantidades de passageiros no horário de pico.

Gerados todos os números e o respectivo *Rol*, foi elaborada uma matriz de organização desses dados (ver figura 6.3), com a Quantidade de Passageiros que Entram no Sistema (Q.P.E.S.), associada ao seu Nível Ordinal (N.O.).

Figura 6.2 – Matriz de disposição dos dados ordenados

QUANTIDADES DE PASSAGEIROS EM DIAS ÚTEIS EM NÍVEL ORDINAL (CRESCENTE) NO PERÍODO ENTRE 6h00m e 7h00m													
N.O.	Q.P.E.S.	N.O.	Q.P.E.S.	N.O.	Q.P.E.S.	N.O.	Q.P.E.S.	N.O.	Q.P.E.S.	N.O.	Q.P.E.S.	N.O.	Q.P.E.S.
1	5.620	26	5.704	51	5.803	76	5.877	101	5.956	126	6.032	151	6.102
2	5.625	27	5.705	52	5.805	77	5.878	102	5.956	127	6.033	152	6.109
3	5.629	28	5.705	53	5.808	78	5.880	103	5.967	128	6.035	153	6.110
4	5.634	29	5.707	54	5.810	79	5.881	104	5.969	129	6.035	154	6.116
5	5.634	30	5.714	55	5.815	80	5.882	105	5.970	130	6.037	155	6.122
6	5.636	31	5.721	56	5.816	81	5.885	106	5.974	131	6.037	156	6.124
7	5.637	32	5.724	57	5.817	82	5.885	107	5.975	132	6.044	157	6.127
8	5.639	33	5.731	58	5.822	83	5.889	108	5.975	133	6.045	158	6.129
9	5.642	34	5.731	59	5.823	84	5.902	109	5.977	134	6.053	159	6.135
10	5.642	35	5.737	60	5.824	85	5.915	110	5.987	135	6.057	160	6.143
11	5.643	36	5.740	61	5.830	86	5.915	111	5.989	136	6.057	161	6.145
12	5.654	37	5.742	62	5.835	87	5.918	112	5.994	137	6.057	162	6.146
13	5.665	38	5.747	63	5.837	88	5.918	113	6.001	138	6.057	163	6.147
14	5.666	39	5.758	64	5.837	89	5.921	114	6.003	139	6.063	164	6.149
15	5.670	40	5.762	65	5.844	90	5.921	115	6.004	140	6.065	165	6.150
16	5.672	41	5.765	66	5.845	91	5.923	116	6.005	141	6.065	166	6.151
17	5.674	42	5.766	67	5.854	92	5.931	117	6.008	142	6.066	167	6.157
18	5.678	43	5.773	68	5.862	93	5.933	118	6.013	143	6.068	-	-
19	5.684	44	5.776	69	5.867	94	5.936	119	6.014	144	6.070	-	-
20	5.685	45	5.778	70	5.868	95	5.936	120	6.014	145	6.082	-	-
21	5.686	46	5.781	71	5.869	96	5.945	121	6.020	146	6.086	-	-
22	5.687	47	5.785	72	5.869	97	5.948	122	6.025	147	6.088	-	-
23	5.689	48	5.786	73	5.874	98	5.949	123	6.025	148	6.091	-	-
24	5.691	49	5.797	74	5.874	99	5.949	124	6.026	149	6.097	-	-
25	5.702	50	5.801	75	5.875	100	5.953	125	6.029	150	6.100	-	-

\* Levantamento referente aos meses de ago/16, set/16, out/16, nov/16, mar/17, abr/17, mai/17 e jun/17.

Fonte: Elaborada pelo autor (2017)

Projetada a matriz foi então calculado o *Range*, resultando em 537 como amplitude total.

Conhecido esse valor foram definidos o número de classes e o tamanho de seus intervalos, obtendo 8 classes e um tamanho de intervalo igual a 67 unidades.

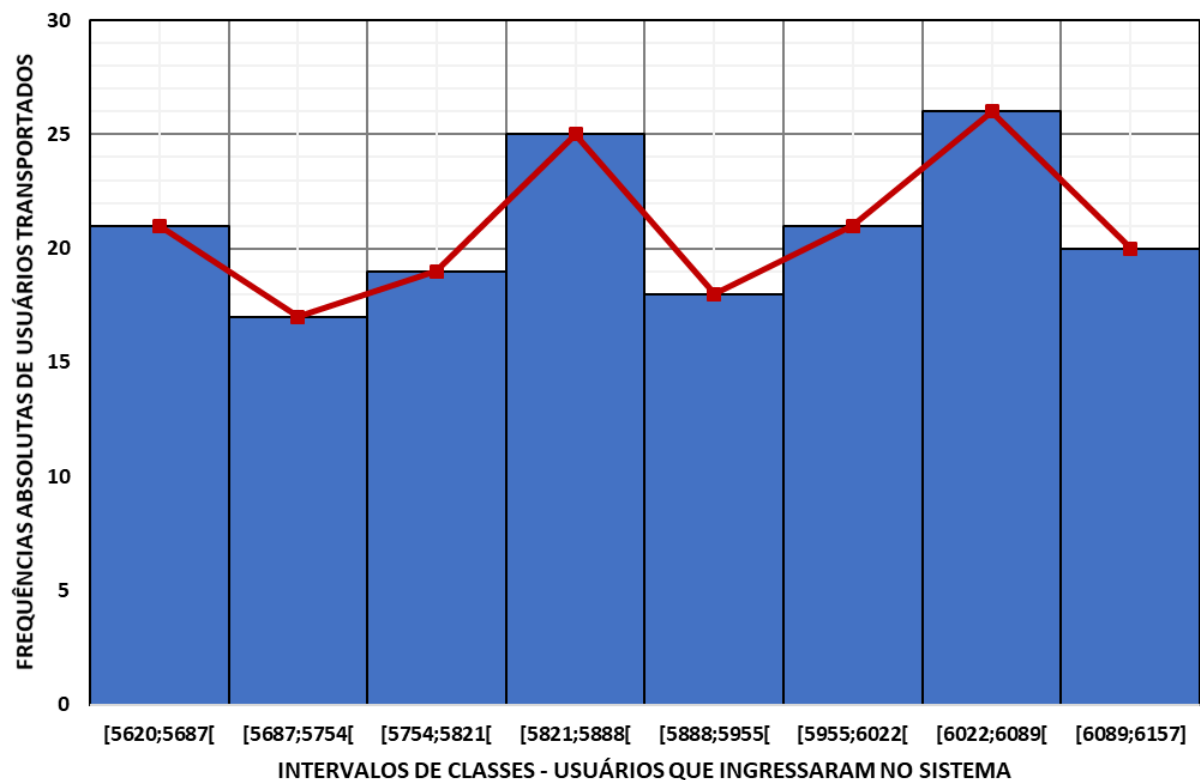
Em seguida foi elaborada a tabela de distribuição de frequências (figura 6.4) e a partir desta gerado o histograma (figura 6.5).

Tabela 6.2 – Distribuição de frequências

CLASSES	INTERVALOS DAS CLASSES	$F_i$	$f_i$	$f_i(\%)$	$F_{ac}$	$f_{ac}$	$\%ac$	$x_i$
1	[5620;5687[	21	0,1257	12,57	21	0,1257	12,57	5654
2	[5687;5754[	17	0,1018	10,18	38	0,2275	22,75	5721
3	[5754;5821[	19	0,1138	11,38	57	0,3413	34,13	5788
4	[5821;5888[	25	0,1497	14,97	82	0,4910	49,10	5855
5	[5888;5955[	18	0,1078	10,78	100	0,5988	59,88	5922
6	[5955;6022[	21	0,1257	12,57	121	0,7246	72,46	5989
7	[6022;6089[	26	0,1557	15,57	147	0,8802	88,02	6056
8	[6089;6157]	20	0,1198	11,98	167	1,0000	100,00	6123
SOMA		167	1,00	100				

Fonte: Elaborada pelo autor (2017)

Figura 6.3 – Histograma dos intervalos das classes e suas frequências absolutas



Fonte: Elaborada pelo autor (2017)

Na segunda etapa, onde foram calculadas as medidas descritivas das informações tratadas no primeiro ciclo e verificadas se são confiáveis para procedência do estudo, foi obtida uma média de 5.896,59 entradas, uma mediana igual a 5.902 e uma moda de 6.057.

Em seguida foram calculadas a variância e a partir dela o desvio padrão. No caso dos dados apresentados anteriormente, foram obtidas uma variância de 24.109 e um desvio padrão igual a 155,27.

Seguiu-se então o cálculo do escore padronizado e a realização de um teste lógico para detecção de *outliers*. Como não foram detectados, deu-se continuidade à análise sem alteração das premissas usadas, não havendo necessidade para reiteração dos cálculos das medidas descritivas.

Seguidamente foi calculado o Coeficiente de Variação de Pearson, que demonstrou a confiabilidade que a média apresentou. Nos dados usados de exemplo obteve-se um coeficiente de 2,63%, indicando uma baixa dispersão e uma boa representatividade para a média aritmética como medida de posição.

Após, foram calculadas as posições dos percentis P10, P25, P75 e P90 no *Rol* ordenado e verificados os valores que se encontravam nessas posições, com o intuito de calcular a Curtose e o Coeficiente de Pearson em relação à assimetria que a curva gerada pela série possui.

Os percentis resultaram nos valores apresentados na figura 6.6.

Tabela 6.3 – Valores encontrados no *Rol* de acordo com os percentis

PERCENTIL	ORDEM	VALOR
P10	18	5678
P25	43	5773
P75	126	6032
P90	150	6100

Fonte: Elaborada pelo autor (2017)

Com os dados de exemplo obteve-se uma Curtose de 0,31, apontando uma curva gerada pela distribuição das frequências mais achatada e um Coeficiente de Pearson de -1,03, mostrando que a curva comporta-se com uma distribuição assimétrica negativa, com assimetria quase forte, e uma média menor que a mediana que, por sua vez, tem o valor menor que o da moda.

As informações resultantes da ordenação e cálculos apresentados até aqui estão apresentadas no ANEXO C, dispostas com o intento de facilitar a compreensão.

Apesar da metodologia especificar que as etapas seguidas até o presente deveriam ser executadas para todas as outras estações da linha, excetuando-se a terminal, para o exemplo foram usadas as médias das médias mensais fornecidas do horário de pico em estudo

para representar as demais estações, visto que o essencial, a estrutura para chegar aos dados representativos de cada uma delas já foi demonstrada, sendo redundante repeti-las aqui.

Em posse das médias adotadas de passageiros que entram no sistema no intervalo estipulado, por estação, foi feita uma tabela resumo, que, além das médias, possui a soma parcial do sistema ao longo da linha e o total de passageiros. Esta tabela e seus resultados são apresentados na figura 6.7.

Tabela 6.4 – Resumo das médias adotadas de passageiros que entram no sistema

<b>MÉDIAS ADOTADAS DE PASSAGEIROS QUE ENTRAM NO SISTEMA</b>		
<b>ESTAÇÃO</b>	<b>Q.P.ES.</b>	<b>SOMA PARCIAL</b>
CALMON VIANA	203	203
ARACARÉ	1.410	1.613
ITAQUAQUECETUBA	2.486	4.099
ENGº MANOEL FEIO	2.332	6.431
JARDIM ROMANO	4.067	10.498
ITAIM PAULISTA	5.897	16.395
JARDIM HELENA - VILA MARA	2.587	18.982
SÃO MIGUEL PAULISTA	1.650	20.632
COMENDADOR ERMELINO	1.040	21.672
USP LESTE	775	22.447
ENGº GOULART	*	22.447
TATUAPÉ	367	22.814
BRÁS	-	-
<b>TOTAL</b>		<b>22.814</b>

\* Estação desativada no período de estudo.

Fonte: Elaborada pelo autor (2017)

Seguiu-se então à análise dos tipos de trens que estão sendo adotados para suprir as demandas das estações da linha em estudo. Como a adoção da frota é um sistema dinâmico, foram escolhidos alguns trens aleatoriamente, sendo eles baseados, tanto em modelos quanto em quantidade, através da provisão de dados da Linha 12 – Safira que foi fornecida. Essa provisão pode ser vista no ANEXO B.

Com o levantamento feito, foi projetada uma tabela resumo com os resultados. Esta está apresentada na figura 6.8.

Tabela 6.5 – Resumo da frota de material rodante para o intervalo em estudo

<b>RESUMO DA FROTA DE MATERIAL RODANTE PARA O INTERVALO EM ESTUDO</b>			
<b>INTERVALO PROGRAMADO ENTRE TRENS (min.)</b>			<b>6</b>
<b>QUANTIDADE TOTAL DE TRENS (un.)</b>			<b>10</b>
<b>MODELO DA COMPOSIÇÃO</b>	<b>CAPACIDADE (6 pass./m<sup>2</sup>)</b>	<b>QUANTIDADE EMPREGADA</b>	<b>OFERTA DE SUPRIMENTO</b>
SÉRIE 7000	2.008	4	8.032
SÉRIE 5550	1.852	2	3.704
SÉRIE 2070	1.924	2	3.848
SÉRIE 4400	1.616	1	1.616
SÉRIE 1400	1.816	1	1.816
<b>TOTAL</b>			<b>19.016</b>

Fonte: Elaborada pelo autor (2017)

Obtidas as quantidades de usuários que entram no sistema naquele intervalo de pico e a demanda de suprimento oferecida pela frota adotada, verificou-se que esta última não atende ao necessário, uma vez que a oferta é de 19.016 lugares e o total de utentes igual a 22.814, resultando em uma diferença de 3.798 passageiros sem oferta de lugares no transporte, segundo as proposições adotadas em relação às composições.

## 7 CONCLUSÃO

O desenvolvimento do estudo presente neste trabalho permitiu uma melhor interpretação do comportamento do fluxo de passageiros que entram no sistema de transporte ferroviário, assim como as estações que mais contribuem com o fluxo, e se há o provável atendimento dessa demanda com a forma de suprimento adotada. Para esse fim foram usadas como premissas os dados obtidos através de levantamentos de quantidades de passageiros que entram em cada estação em uma linha férrea por intervalos de tempo.

Ao decorrer da metodologia proposta foram apresentadas as fases junto às suas respectivas contribuições parciais para a formação integral do modelo, usando como exemplo os dados da estação de Itaim Paulista da Linha 12 – Safira da CPTM, assim como a linha em si, com o objetivo de melhor elucidar o modelo e levar a uma fácil compreensão do trabalho. No exemplo, o número de trens adotados no intervalo não atendia a quantidade de passageiros, mesmo consideradas suas capacidades máximas.

As fases supracitadas apresentaram como as frequências de quantidades de passageiros se portavam ao longo do período de estudo, suas médias baseadas em valores já organizados e tratados e se são ou não confiáveis para adoção, frente ao comparativo das provisões fornecidas pela frota adotada.

Dois óbices apresentaram-se ao decorrer do desenvolvimento deste trabalho. O primeiro refere-se o espaço de tempo para adquirir os dados a serem usados como premissas do estudo. As informações resultantes do modelo poderiam ter ainda mais precisão caso houvesse um número maior de amostras nas quais basear-se. Porém, em contrapartida, quanto maior o intervalo de tempo usado no estudo, mais ele sofreria a influência da flutuação populacional, limitando assim esse período de obtenção de amostras a um espaço que não sofresse ações graves.

O segundo é o fato de não saber-se qual sentido tomam os passageiros após entrarem nas estações, visto que as informações resultantes da metodologia usada pela Companhia Paulista de Trens Metropolitanos para o levantamento de usuários limitam-se à uma esfera tal qual não engloba esses dados, sendo causa deste a densidade de empregos, vagas acadêmicas, estabelecimentos e outros, no centro de São Paulo, adotados como proposição para a opção do sentido escolhido pelos utentes, além da convergência de todas as malhas ferroviárias para esse mesmo local, sendo para desembarque, transferência entre linhas ou mesmo entre modais.



A metodologia desenvolvida atendeu as expectativas apresentadas no início do trabalho, gerando resultados satisfatórios a permitindo que assim se possa ter uma interpretação adequada do trânsito de passageiros que estão utilizando o sistema, as estações que mais estão colaborando com fluxo e contribui com informações que possam ser usadas em estudos futuros e soluções para descongestionamentos.

## REFERÊNCIAS

BRASIL. **Estatísticas Históricas do Brasil: Séries Econômicas, Demografias e Sociais**, de 1550 a 1988, 2 ed. Rio de Janeiro: IBGE, 1990.

CPTM, **Nossa História**. Disponível em: <<http://www.cptm.sp.gov.br/a-companhia/Pages/Nossa-Historia.aspx>>. Acesso em: 5 de maio de 2016.

CPTM. **Passageiros Embarcados por Hora no Sistema CPTM – Dia Útil da Semana, Sábados e Domingos: Maio – Junho – 2017**. São Paulo, 2017.

CPTM. **Relatório da administração 2015**. São Paulo, 2015. 123 p.

CPTM. **Relatório da administração 2016**. São Paulo, 2016. 137 p.

MARTINS, Gilberto de Andrade; DOMINGUES, Osmar. **Estatística Geral e Aplicada**. 5. ed. São Paulo: Editora Atlas S.A., 2014.

METRÔ. **Pesquisa de Mobilidade da RMSP – 2012: Manual da Pesquisa Domiciliar**, 1 ed. São Paulo: METRÔ, 2012. 71 p.

METRÔ. **Pesquisa de Mobilidade da RMSP – 2012: Síntese das Informações - Pesquisa Domiciliar**, 1 ed. São Paulo: METRÔ, 2013. 69 p.

METRÔ. **Pesquisa Origem e Destino 2007: Manual da Pesquisa Domiciliar**, 1 ed. São Paulo: METRÔ, 2007. 76 p.

METRÔ. **Pesquisa Origem e Destino 2007: Síntese das Informações - Pesquisa Domiciliar**, 1 ed. São Paulo: METRÔ, 2008. 82 p.

MICALLI, Josie. **Modalidades de Transporte – Transporte Ferroviário**. 2016. São Paulo. 10 p. Notas de aula.

PAULETTE, Walter. **Estatística: Teoria e Aplicações**. São Paulo, [2014?]. (Apostila do Curso de Estatística da Faculdade de Tecnologia de São Paulo).

REIS FILHO, Nestor G. **São Paulo e Outras Cidades – Produção Social e Degradação dos Espaços Urbanos**. São Paulo: Hucitec, 1994.

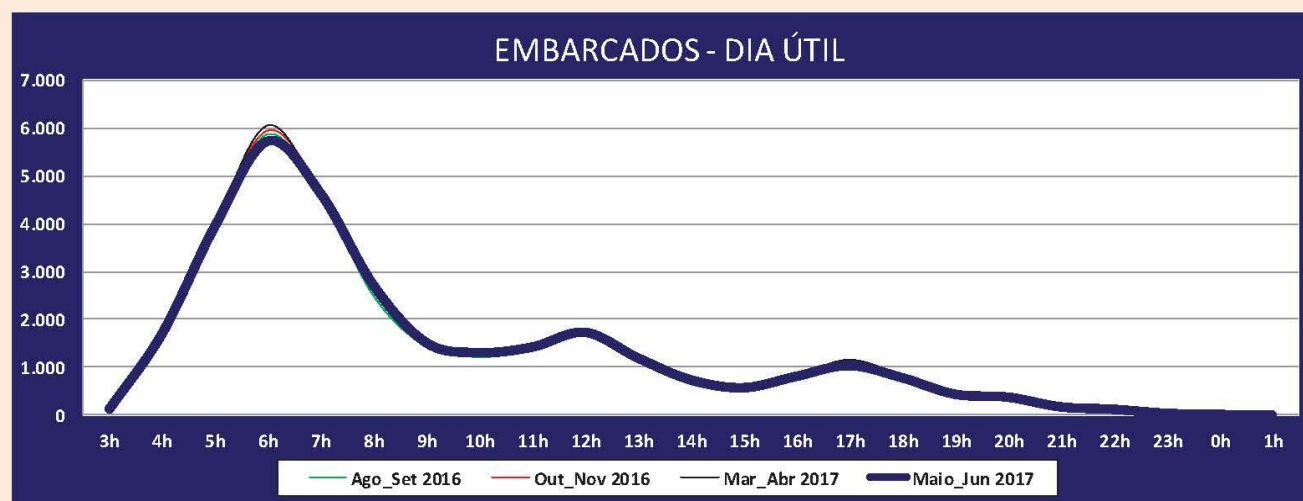
STEFANI, Celia R. B. **O sistema ferroviário paulista: um estudo sobre a evolução do transporte de passageiros sobre trilhos**. 2007. 304 f. Dissertação (Mestre em Ciências) - Universidade de São Paulo, São Paulo, 2007.

## ANEXO A



Passageiros Embarcados por Hora – Dia  
Útil e Final de Semana – Maio/Jun - 2017

## ITAIM PAULISTA



Linha 12 - Safira

HORÁRIO		EMBARCADOS - DIA ÚTIL							
		Ago_Set 2016	%	Out_Nov 2016	%	Mar_Abr 2017	%	Maio_Jun 2017	%
3h	4h	134	0,4	138	0,4	138	0,4	127	0,4
4h	5h	1.681	5,5	1.717	5,5	1.684	5,4	1.710	5,5
5h	6h	4.036	13,2	4.046	12,9	4.015	12,8	3.957	12,7
6h	7h	5.867	19,2	5.948	19,0	6.055	19,3	5.726	18,4
7h	8h	4.601	15,0	4.713	15,0	4.598	14,7	4.625	14,9
8h	9h	2.486	8,1	2.698	8,6	2.627	8,4	2.709	8,7
9h	10h	1.430	4,7	1.542	4,9	1.496	4,8	1.508	4,9
10h	11h	1.212	4,0	1.296	4,1	1.248	4,0	1.305	4,2
11h	12h	1.358	4,4	1.407	4,5	1.417	4,5	1.420	4,6
12h	13h	1.647	5,4	1.708	5,5	1.671	5,3	1.727	5,6
13h	14h	1.146	3,7	1.144	3,7	1.156	3,7	1.192	3,8
14h	15h	644	2,1	654	2,1	704	2,2	734	2,4

HORÁRIO		EMBARCADOS - DIA ÚTIL							
		Ago_Set 2016	%	Out_Nov 2016	%	Mar_Abr 2017	%	Maio_Jun 2017	%
15h	16h	516	1,7	525	1,7	566	1,8	571	1,8
16h	17h	789	2,6	771	2,5	820	2,6	809	2,6
17h	18h	1.118	3,6	1.038	3,3	1.154	3,7	1.034	3,3
18h	19h	785	2,6	744	2,4	827	2,6	779	2,5
19h	20h	441	1,4	458	1,5	450	1,4	435	1,4
20h	21h	394	1,3	409	1,3	366	1,2	373	1,2
21h	22h	182	0,6	197	0,6	169	0,5	170	0,5
22h	23h	119	0,4	123	0,4	117	0,4	120	0,4
23h	0h	37	0,1	43	0,1	36	0,1	35	0,1
0h	1h	6	0,0	8	0,0	11	0,0	11	0,0
1h	2h	0	0,0	0	0,0	0	0,0	0	0,0
TOTAL		30.631	100,0	31.327	100,0	31.325	100,0	31.076	100,0

## ANEXO B

CPTM					DADOS BÁSICOS OPERACIONAIS							
					LINHA 12 SAFIRA							
					Brás - Calmon Viana		Brás - Engº Manoel Feio					
Extensão (km)					38,8		30,7					
Municípios atendidos pela CPTM					SP, IQC, POA		SP, IQC					
Passagens em Nível (Rodoviário)					1		----					
Quantidade de Estações por Loop					13		10					
Horário Comercial (Partida do Primeiro/Último Trem das Estações Terminais)					04h00 às 24h00							
Tempo de Percurso (minutos)					63		50					
Velocidade Comercial (km/h)					37		37					
Quantidade de Viagens Programadas					Dias Úteis		249		44			
					Sabados		268		----			
					Dom. / Fer.		222		----			
Quantidade de Trens Programados Operacional) (Reserva					Dias Úteis		Pico M/T		20 + ( 0 )		----	
							Vale		16 + ( 0 )		----	
							Pós-vale		13 + ( 0 )		----	
					Sabados		Pico Manhã		15 + ( 0 )		----	
							Vale		15 + ( 0 )		----	
							Pós-vale		6 + ( 0 )		----	
					Dom. / Fer.		Vale		12 + ( 0 )		----	
Faixas Horárias Intervalo Programado entre Trens (minutos) Ref. Est. Engº Goulart e					Dias Úteis		Pré-pico		04:00 às 04:12 = 12			
									04:12 às 06:19 = 06			
							Pico Manhã		06:19 às 07:19 = 06			
									07:19 às 08:24 = 06			
							Vale		08:24 às 15:49 = 08			
							Pico Tarde		15:49 às 17:42 = 06			
									17:42 às 18:42 = 06			
									18:42 às 20:31 = 06			
					Pós-pico		20:31 às 23:38 = 10					
							23:38 às 24:00 = 10					
					Sabados		Pré-pico		04:00 às 19:04 = 08			
							Pico Manhã					
							Vale					
							Pós-vale		19:04 às 24:00 = 20			
					Dom. / Fer.		Pré-vale		04:00 às 20h00 = 10			
							Vale					
Pós-vale		20:00 às 24:00 = 20										
Média de Passageiros Transportados (Média: DEZEMBRO/2016)					Dias Úteis		260.699					
					Sabados		180.281					
					Dom. / Fer.		90.505					
Oferta de Lugares / hora / sentido (Média Ponderada 8/7000, 1/5550, 4/2070, 5/4400 e 2/1400)					Manhã (06h00 às 07h00) (EGO)		18.630		18.630			
Carregamento / hora / sentido Considerando 2.664.804 pass/transportados					Tarde (17h15 às 18h15) (EGO)		17.166		19.126			
Capacidade da Composição Oferta de Lugares - 6 pass./m2 (Carros por Composição)					Série 7000 (08)		2.008					
					Série 5550 (08)		1.852					
					Série 2070 (08)		1.924					
					Série 4400 (06)		1.616					
					Série 1400 (06)		1.816					
Frota Patrimonial Alocada de Material Rodante (Trens Unidades)					Série 7000		18					
					Série 5550		6					
					Série 2070		10					
					Série 4400		12					
					Série 1400		4					
Sistema de Sinalização					ATS							
Subestações de Tração (Potência MW)					ISP (12), SGU (9), ERM (8), EMF (8), CVN (9)							
Cabinas Seccionadoras					BAS / EGO							

## APÊNDICE A

### TABELA RESUMO - LEVANTAMENTO ESTATÍSTICO DE UTENTES QUE ENTRARAM NO SISTEMA

**ESTAÇÃO:** Itaim Paulista

**LINHA:** 12 – Safira

**INTERVALO (HORÁRIO):** 6h00m – 7h00m

**AMOSTRA (DIAS ÚTEIS):** 167

**PERÍODO (ANO/MÊS):**

<b>2016</b>	<b>ago; set; out; nov.</b>
<b>2017</b>	<b>mar; abr; mai; jun.</b>

DADOS	
OUTLIERS	NÃO
MÉDIA	5.896,59
MEDIANA	5.902
MODA	6.057
VARIÂNCIA	24.108,95
DES. PAD.	155,27
C.V. PEARSON	2,63
C. PEARSON	-1,03
CURTOSE	0,31

#### COMPORTAMENTO DA SÉRIE

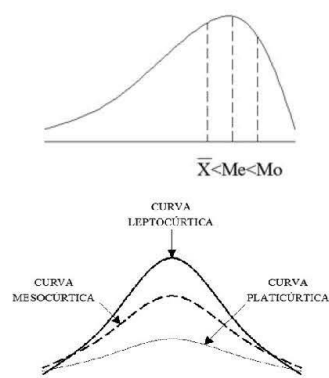
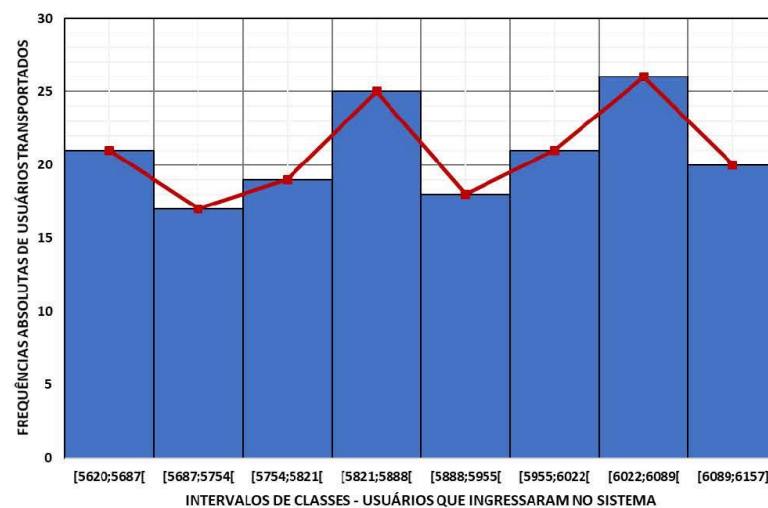


TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

CLASSES	INTERVALOS DAS CLASSES	$F_i$	$f_i$	$f_i(\%)$	$F_{ac}$	$f_{ac}$	$\%ac$	$x_i$
1	[5620;5687[	21	0,1257	12,57	21	0,1257	12,57	5654
2	[5687;5754[	17	0,1018	10,18	38	0,2275	22,75	5721
3	[5754;5821[	19	0,1138	11,38	57	0,3413	34,13	5788
4	[5821;5888[	25	0,1497	14,97	82	0,4910	49,10	5855
5	[5888;5955[	18	0,1078	10,78	100	0,5988	59,88	5922
6	[5955;6022[	21	0,1257	12,57	121	0,7246	72,46	5989
7	[6022;6089[	26	0,1557	15,57	147	0,8802	88,02	6056
8	[6089;6157]	20	0,1198	11,98	167	1,0000	100,00	6123
SOMA		167	1,00	100				

HISTOGRAMA



VALOR ADOTADO ( $\bar{X}$ ) =

5.897