



## **CRIAÇÃO DE UM MODELO PROBABILÍSTICO DE PREVISÃO DO TEMPO DE TRÁFEGO NAS VIAS DE SÃO PAULO**

**Deivis Coutinho Attuy** – tmasson@mackenzie.br

**Fabio Jesus Moreira de Almeida** – fabio.almeida@mackenzie.br

**Terezinha Jocelen Masson** – tmasson@mackenzie.br

Universidade Presbiteriana Mackenzie – Escola de Engenharia

Rua da Consolação, 896, prédio 06.

CEP 01302-907–São Paulo – São Paulo

*Resumo: As recentes abordagens diárias à respeito dos congestionamentos na cidade de São Paulo tem levado a muitos questionamentos sobre como agir para ajudar a minimizar os altos índices de lentidão e tráfego na cidade e também como otimizar o tempo gasto em uma trajetória. O presente trabalho apresenta um desenvolvimento de uma nova aplicação que contempla vários campos da estatística e analisa dados preexistentes e novos dados com intuito da criação de um modelo probabilístico que auxilie aos envolvidos o entendimento dos fluxos da amostra recolhida. Para tanto, foram relacionados os métodos estatísticos de organização de dados, as análises qualitativas e quantitativas, os testes de aderência, as estatísticas de tempo real, o desenvolvimento polinomial de distribuições que representem as amostras, dentre outras aplicadas no sistema atual de mobilidade urbana paulistana. Como resultado foi desenvolvida uma ferramenta que analise estatisticamente variações da amostra. Este trabalho tem como objetivo modelar e construir um modelo probabilístico, por intermédio de uma base de dados estatísticos extraídos durante 43 dias entre 01 de setembro até 29 de outubro de 2011, por acesso público e gratuito ao veículo de comunicação Porto Vias®, que refletem os tempos de tráfego de amostras do trânsito de São Paulo, viabilizando uma análise futura em ambiente controlado dos sistemas descritos.*

*Palavras Chaves: Trânsito; Organização de Dados, Testes de Aderência, Qui-Quadrado, Normalização Diferencial de Distribuições.*

### **1. INTRODUÇÃO**

O deslocamento das pessoas é um dos elementos mais presentes na história, desde a movimentação dos povos em busca das terras mais férteis, quanto dos próprios animais em suas migrações sazonais. Essas movimentações foram incorporadas no instinto dos povos, e os benefícios que eram obtidos sustentavam esses trâmites (LARRAZÁBAL, 2002).

Realização:

 **ABENGE**

Organização:



**O ENGENHEIRO  
PROFESSOR E O  
DESAFIO DE EDUCAR**



Esses benefícios obtidos no decorrer do tempo passaram a virar necessidades, e o transporte passou a ser um elemento crescente no meio urbanizado. Contudo, o aumento da quantidade de veículos de transportes urbano acabou sendo maior que o crescimento das vias. Esse crescimento acarretou muitos focos de congestionamentos (VASCONCELLOS, 2000).

A cidade de São Paulo é a maior cidade brasileira possuindo 25% da frota nacional dos veículos. Os seus 14.000 km de vias urbanas não suportam os grandes fluxos nos horários de picos, pois a maior parte dos deslocamentos ocorrem em uma extensão de 2.500 km apenas. (SCARINGELLA, 2001).

Essa abordagem dos congestionamentos diários e o tráfego cada vez mais lento na cidade de São Paulo geraram a intenção de desenvolvimento de um estudo apurado e técnico sobre o assunto. A literatura apresenta inteligentes mecanismos de otimização de tempo, sendo que muitos destes escopos já são conhecidos e vigentes, tais como semáforos inteligentes (que visam apenas avenidas grandes e desconsideram seus afluentes), rodízio, (SCARINGELLA, 2001) operação rótula, rota alternativa, entre outros, ou seja, ferramentas que fazem uso de conceitos estatísticos e matemáticos para modelar os movimentos constantes de fluxo.

## 2.1 SISTEMAS EM COLAPSO

Para Bertalanffy (1975) sistema é o conjunto de elementos que se inter-relacionam com pelo menos um objetivo em comum. A união desses elementos para formar um sistema maior faz com que os elementos desempenhem capacidades que não teriam em estado de isolamento. O sistema depende do funcionamento de seus elementos e para que um sistema tenha total desempenho no seu objetivo é necessário que todos os seus elementos desempenhem suas funções de maneira saudável.

Entretanto os sistemas podem ser sobrecarregados, e trabalharem na sua capacidade total. Quando isso acontece, qualquer um dos elementos que não esteja com seu desempenho perfeito pode fazer com que o sistema pare de funcionar. Senge (1998) chamou esses de sistemas prestes ao colapso ou sistemas colapsais. Segundo ele, um sistema que esteja preste ao colapso está vulnerável a qualquer desconformidade de seus elementos. O momento em que esse sistema entra em colapso é chamado de ponto de ruptura, pois se relaciona com o momento de desligamento do funcionamento do sistema, o que pode se acontecer no desempenho de um dos seus elementos ou na informação trocada entre eles.

Um grande exemplo de sistema preste a colapso é o sistema de trânsito da cidade de São Paulo, que por estar em sua capacidade máxima de seu ambiente viário, acaba vulnerável a qualquer acontecimento. (SCARINGELLA, 2001)

## 2.2 TRÂNSITO E CONGESTIONAMENTO

Para Rozestraten (2005) o trânsito é composto primordialmente por três elementos básicos que são: a via (ou ambiente viário) que corresponde a todos os meios estáticos que compõem o trânsito, o veículo que são os meios móveis controlados pelos usuários e os usuários da via e/ou do veículo.

Ao se construir uma interação desses três elementos, pode-se conferir que existe a possibilidade de que haja discrepância em algum deles o que pode fazer com que o sistema não venha a fluir normalmente.

Hoje em São Paulo é identificado uma deficiência no fator da via ou ambiente viário (SCARINGELLA, 2001). Isto é, as condições de fluxo de veículos estão acima do que as vias



suportam, fazendo com que o sistema, do trânsito de São Paulo, trabalhe a beira de uma ruptura e um proveniente colapso. Isso ocorre, pois o crescimento dos deslocamentos, ao longo dos últimos anos, serem maiores que o crescimento dos ambientes viários, fato esse visto que num período de cinco anos (1992 a 1997), em dados medidos pela CET (Companhia de Engenharia de Tráfego), órgão responsável pelo controle e monitoramento do trânsito de São Paulo, afirmam que a média de quilômetros congestionados passou de 40 km para 120 km nos horários de pico das tardes paulistanas.

### 2.3 CLASSIFICAÇÃO DE VARIÁVEIS

Todos os dados analisados estatisticamente são relacionados a alguns elementos que determinam parâmetros a esses dados. Esses fatores são denominados de variáveis. As variáveis são divididas em dois tipos as qualitativas e as quantitativas.

As variáveis do tipo quantitativas apresentam como possível realização números resultantes de uma contagem ou mensuração como número de filhos, salário, tempo de trajeto, tamanho de peças, entre outros. As variáveis quantitativas podem sofrer uma classificação dicotômica tais como:

a) discretas, cujo valores possíveis formam um número finito ou numerável de valores, e que resultam de uma contagem, como por exemplo o número de filhos.

b) contínuas, cujo valores possíveis pertencem a um intervalo de números reais e que resultam de uma mensuração como por exemplo estatura e peso. (BUSSAB, 2007).

Quando se estuda amostras não é possível julgar qual a magnitude do erro em que pode se estar cometendo ao analisar os valores da mesma. Nesse sentido surge a ideia de criar um intervalo de confiança onde pode ser ter essa magnitude baseando-se na distribuição amostral do estimador pontual.

O intervalo de confiança da média da população a ser estudada ( $\mu$ ) representa uma porcentagem  $\gamma$  (coeficiente de confiança) da população. Isto é, ao serem colhidas  $n$  amostras da população e aplicadas aos intervalos de confiança nas amostras em pelo menos uma porcentagem  $\gamma$  das amostras conterá a média ( $\mu$ ) da população (BUSSAB, 2007).

## 3 MÉTODOS E PROCEDIMENTOS

A determinação do número de dados coletados foi feita por inferência estatística (BUSSAB, 2007), e por teorema de Sturges (HYNDMAN, 1995) que embasou um número confiável de casos para o problema vigente. Para tanto foram colhidas amostras dos tempos de trajeto de tres grandes vias da cidade (Marginal do Rio Pinheiros, Marginal do Rio Tietê e Avenida Paulista).

A escolha das três avenidas que foram estudadas nesse trabalho partiu de uma pesquisa básica no site da Companhia de Engenharia de Tráfego (CET) responsável por todo o trânsito da cidade. O site contempla uma página onde atualiza quais são as ocorrências do trânsito nas principais vias de São Paulo, nas quais as três rotas escolhidas fazem parte.

Durante a definição da metodologia do presente trabalho, estabeleceu-se a necessidade obter dados dos tempos de trajeto das rotas onde seriam estudados estatisticamente. Para tanto foi constatado que por meio da ferramenta “online” Porto Vias, disponibilizada pela seguradora de veículos Porto Seguro, poderia fornecer os tempos dos trajetos.

A ferramenta Porto Vias calcula os tempos de trajeto por meio dos rastreadores instalados nos carros que utilizam seu serviço. Com os dados da localização dos carros é calculado qual



o tempo previsto do trajeto no exato momento da consulta. Esse serviço é de domínio público e acessado por um perfil virtual gratuito no *site* da seguradora.

Para obtenção dos dados do *site* foi desenvolvido um arquivo de repetição na ferramenta Microsoft Excel. Foi programado o acesso ao *site* e cópia do tempo do trajeto e foi construída uma matriz com os tempos dos trajetos de cada rota, contendo data e horário da consulta. Os dados se atualizam em tempo real. O acesso as consultas foi programado para cada 5 minutos, com a compilação dos dados. As consultas utilizadas nas análises do trabalho foram realizadas durante 43 dias, entre 01 de setembro à 29 de outubro de 2011 e consolidadas em um banco de dados com mais de 7000 linhas conforme consta no apêndice A.

Os dados foram consolidados e divididos conforme as seis rotas provenientes das consultas: a) Av. Paulista sentido Paraíso-Consolação; b) Av. Paulista sentido Consolação-Paraíso; c) Marginal do Rio Pinheiros sentido Lapa-Interlagos; d) Marginal do Rio Pinheiros sentido Interlagos-Lapa; e) Marginal do Tietê sentido oeste-leste; f) Marginal do Tietê sentido leste-oeste.

No tratamento dos dados obtidos, foram consideradas algumas premissas para criar parâmetros de análises. Os dados que foram coletados no mesmo dia da semana para a criação de espaços amostrais que possibilitou analisar por dia da semana. Pois em São Paulo existem fatores que podem influenciar o comportamento durante os dias da semana como, o rodízio municipal de veículos (SCARINGELLA, 2001).

Os dados foram consolidados contendo os horários que permitiu que as análises selecionassem os intervalos de horários para analisar permitindo a visualização de eventuais mudanças entre horários diferentes facilitando a análise do horário específico bem como daqueles que compartilham do rodízio municipal de veículos.

Para o desenvolvimento da metodologia de análise foram estabelecidos quatro critérios de consultas das informações do banco de dados que são: a) Horário inicial da consulta; b) Horário final da consulta; c) Dia da semana; d) Rota e sentido desejados.

Na ferramenta de análise dos dados, foram destinados campos representados na Figura 1, específicos para a seleção do que foi denominada premissas de análises. Esses campos que devem ser escolhidos pelos usuários embasam as análises estatísticas que a ferramenta realiza.

Hora inicial	07:00:00	▼	Rodar
Hora final	10:00:00	▼	
Dia da Semana	segunda-feira	▼	
Rota	Paulista sentido Consolação	▼	

Figura 1 – Campos de premissas de análise.

Fonte: Arquivo de Análises, o autor.

Por meio das consultas do banco de dados foram extraídos os elementos típicos de posição, que expressam em números as tendências que representam a distribuição da amostra (CRESPO, 2009) apresentando os dados de posição organizados, a fim de aperfeiçoar a consulta e auxiliar os cálculos posteriores bem como as análises subjetivas. Os dados apresentados foram: a) Média amostral; b) Erro padrão amostral; c) Mediana amostral; d) da



amostral; e) Desvio padrão amostral; f) Variância amostral; g) Curtose; h) Assimetria; i) Intervalo entre picos de máximo e mínimo.

Para a descrição da distribuição que representa a amostra foi inserido um intervalo de confiança da mesma, onde os valores fora do intervalo fossem descartados, de tal forma a assegurar que a distribuição que representa a amostra fosse significativa, estabelecendo a amplitude dos dados a serem estudados (BUSSAB, 2007).

A amostra consultada foi testada por aderência com as distribuições Gaussiana, Gama, Weibull, Poisson e exponencial por meio do teste de aderência de Qui-Quadrado. Esse teste baseou suas classes utilizando a regra de Sturges, que calcula a quantidade de classes que indica uma melhor representatividade da amostra.

Além de testar a amostra com as distribuições existentes foi desenvolvido por intermédio de um polinômio de interpolação de Newton-Gregory o cálculo de uma função de grau seis para demonstrar a distribuição pelas suas frequências e seus tempos de trajeto. O polinômio resultante foi normalizado por cálculo diferencial para que o intervalo da amostra representasse a probabilidade de ocorrência pelo tempo de trajeto.

#### 4 RESULTADOS E DISCUSSÕES

Com os dados a serem analisados tratados e selecionados nos campos de premissas de análises foram obtidas das amostras os parâmetros estatísticos que auxiliam nas análises e que descreve parâmetros da curva da amostra.

O programa Microsoft Excel tem uma biblioteca suplementar que analisa e retorna os valores dos parâmetros de organização dos dados. Essa biblioteca suplementar é utilizada como um recurso estatístico. Para que fosse atualizado a cada consulta das premissas de análise foi desenvolvido um programa para ativar o suplemento por meio do acionamento do botão “rodar” conforme a Figura 1.

Através de uma consulta às informações do banco de dados foi verificado qual é o tempo mínimo da amostra e qual o seu respectivo horário ideal de maneira a orientar o usuário em relação à amostra em qual a média dos tempos que foram obtidas a menores coletas de tempo conforme a Tabela 1.

Tabela 1 – Organização de Dados.

<i>Marginal Pinheiros sentido Lapa</i>	
Média	27,7872
Erro padrão	0,488
Mediana	24
Moda	22
Desvio padrão	11,7323
Variância da amostra	137,648
Curtose	22,4992
Assimetria	4,21758
Intervalo	92
Mínimo	20
Máximo	112
Soma	16061
Contagem	578
Nível de confiança(95,0%)	0,95848
Tempo médio Mínimo	21,40 min
Horário Ideal	15:20:00





Os parâmetros da organização de dados são base para análises da distribuição da amostra. A média, mediana a moda demonstram a posição da distribuição da amostra em relação ao eixo das abscissas.

Pelos coeficientes de curtose e assimetria e ainda considerando a variância e o desvio padrão é possível verificar a dispersão da curva e a variabilidade.

Os dados obtidos na organização de dados serão ainda abordados no presente trabalho pelas informações necessárias para o desenvolvimento de algumas das distribuições estudadas.

#### 4.2 APLICAÇÃO DO INTERVALO DE CONFIANÇA NOS DADOS

Para que a seleção das amostras das distribuições realizou-se uma aplicação do intervalo de confiança nas mesmas para separar os dados que continham um nível de significância acima de 1%.

Com a aplicação do intervalo de confiança nas amostras foi verificado que quanto maior era o intervalo de tempo analisado menores eram os dados que faziam parte do intervalo de confiança. Conforme demonstra os dados da Tabela 2.

Tabela 2 – Quantidade da amostra para significância de 1%.

Hora inicial	Hora final	Dia da semana	Rota	Quantidade da amostra	Quantidade da amostra para significância de 1%	Porcentagem
07:00:00	08:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	64	22	34%
07:00:00	09:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	133	37	28%
07:00:00	10:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	200	54	27%
07:00:00	11:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	267	71	27%
07:00:00	12:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	336	81	24%
07:00:00	13:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	401	88	22%
07:00:00	14:00:00	Segunda-Feira	Marginal Pinheiros sentido Lapa	468	94	20%

Observando-se os tempos de trajetos dos picos de horários, conforme o Gráfico 2, pode-se identificar picos durante os chamados horários de síncope, por os quais quando se aplica a significância reflete os resultados da Tabela 2.

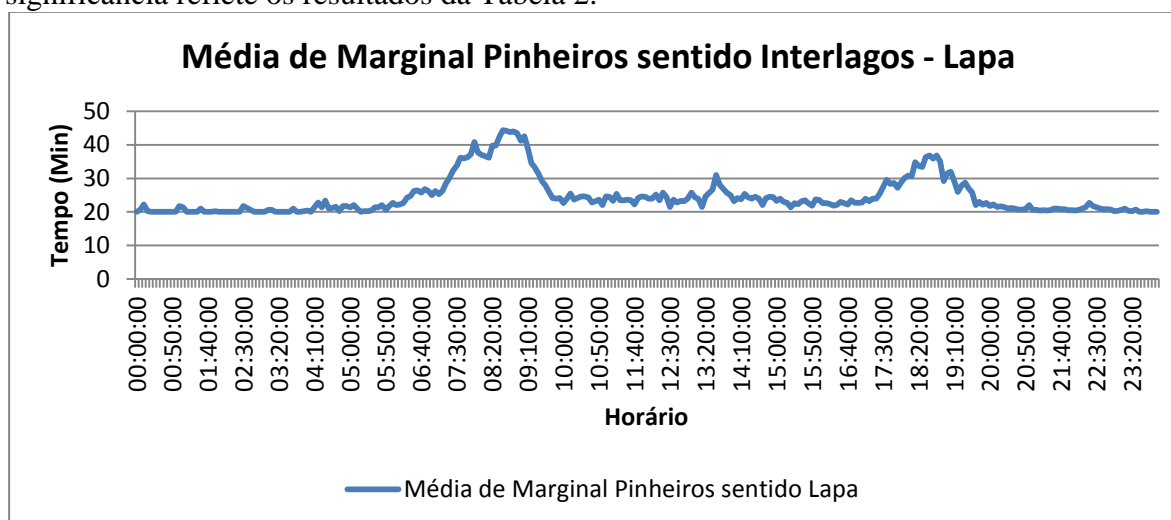


Gráfico 2 – Média de Marginal Pinheiros sentido Interlagos - Lapa.  
 Fonte: o autor.



#### 4.3 TESTE DE ADERÊNCIA NAS DISTRIBUIÇÕES TESTADAS

O teste de aderência escolhido para ser aplicado nas amostras foi o teste de qui-quadrado a fim de testar a amostra versus as distribuições que foram estudadas. Assim sendo o intervalo de confiança aplicado nos dados amostrais selecionados pelas premissas de análises, considerou os limites amostrais a serem testados à aderência da amostra com as distribuições estudadas, ou seja, os valores que estavam dentro dos limites do intervalo de confiança foram considerados para a realização do teste.

Para a definição dos intervalos do teste de aderência foi pesquisado uma metodologia para calcular quantas classes a amostra deveria conter para um teste confiável. A regra de Sturges foi a metodologia utilizada para definir qual número de classes do teste.

Tabela 3 – Regra de Sturges aplicada ao teste de aderência.

AMOSTRA				
	Intervalo Min.	Intervalo Max	Xi	Fi
1	8	9	8,50	7
2	9	10	9,50	11
3	10	11	10,50	4
4	11	12	11,50	7
5	12	13	12,50	7

Os tempos que foram coletados da amostra são em minutos e em números inteiros, e a medida de não haver inconsistência nas distribuições, quando o passo calculado pela regra de Sturges, resultou em um número menor que 1 minuto. Assim, o passo utilizado para realizar o teste passou a ser 1 minuto, mitigando a hipótese de aparecerem falsas frequências zeradas na distribuição.

A fim de estabelecer um padrão simples para as análises foram definidos passos iguais para todos os intervalos do teste e a média aritmética do intervalo como o valor padrão das distribuições, conforme os valores da Tabela 3.

##### 4.3.1 Distribuição normal

A primeira distribuição que foi introduzida no estudo do presente trabalho foi a distribuição normal ou gaussiana conforme o Gráfico 3. Como parâmetros para calcular a probabilidade da distribuição, considerando-se a média aritmética e o desvio padrão do intervalo de confiança da amostra, a fim de reproduzir a curva da distribuição.

Os dados da distribuição foram testados quanta a aderência à amostra, por meio do teste qui-quadrado. Os valores de qui-quadrado calculados das premissas de análises foram confrontados com o valor crítico onde se testa a hipótese de  $H_0 = \text{Amostra} = \text{Distribuição Normal}$  quando  $X^2$  da amostra  $< X^2$  da distribuição normal.

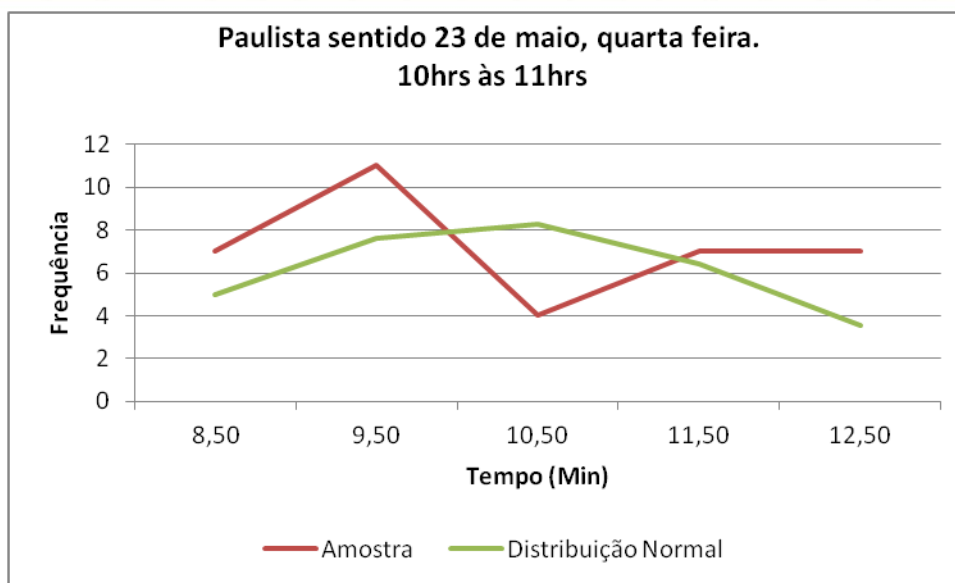


Gráfico 3 – Dados da amostra versus distribuição Normal.  
 Fonte: o autor.

O sucesso na hipótese  $H_0$  não foi observado com frequência nos testes realizados em relação a aderência de qui-quadrado. Em sua maior parte a distribuição normal não aderiu a amostra. Como observa-se na Tabela 4.

Tabela 4 – Teste de aderência de qui-quadrado amostra versus distribuição normal.

AMOSTRA					DISTRIBUIÇÃO NORMAL		
	Intervalo Min.	Intervalo Max	$X_i$	$F_i$	Dist.Normal	$F_i$ .Normal	Teste
1	8	9	8,50	7	0,138229708	5	1,854
2	9	10	9,50	11	0,210817909	8	0,343
3	10	11	10,50	4	0,229386959	8	8,171
4	11	12	11,50	7	0,178067662	6	2,627
5	12	13	12,50	7	0,09861807	4	0,099
					X <sup>2</sup> Dist.Normal		13,095
Passo:		1			X <sup>2</sup> Critico		5,991

#### 4.3.2 Distribuição gama

Para a definição da distribuição gama que seria utilizada para testar a aderência com a amostra foi utilizado um método de definição dos parâmetros alfa e beta, para minimizar o valor de  $X^2$  calculado em relação à distribuição gama.

No desenvolvimento do método de definição dos parâmetros alfa e beta foi utilizado a ferramenta Solver®, na qual por meio de técnicas lineares e não lineares, ocorre a aproximação de valores que correspondem as restrições pré-definidas e minimiza a função





objetivo descrita. Na realização do presente trabalho foi inserido como função objetivo o cálculo de  $X^2$  e variáveis os parâmetros alfa e beta apenas para valores positivos.

Os resultados foram testados a hipótese de  $H_0 =$  Amostra = Distribuição Gama quando  $X^2$  da amostra  $< X^2$  da Distribuição Gama e mesmo com a metodologia de minimização de  $X^2$  da amostra, não possível obter sucesso e muitos casos, conforme Tabela 5.

Tabela 5 – Teste de aderência de qui-quadrado amostra versus distribuição gama.

AMOSTRA					DISTRIBUIÇÃO GAMA		
	Intervalo Min.	Intervalo Max	Xi	Fi	Dist.Gama	Fi Gama	Teste
1	8	9	8,50	7	0,11555	4	1,939
2	9	10	9,50	11	0,12605	5	9,203
3	10	11	10,50	4	0,12331	4	0,043
4	11	12	11,50	7	0,11035	4	2,307
5	12	13	12,50	7	0,09170	3	4,144
					<b><math>X^2</math> Distribuição Gama</b>		<b>17,64</b>
Passo:		1				<b><math>X^2</math> Critico</b>	<b>5,991</b>

### 4.3.3 Distribuição Weibull

Gráfico 4 – Dados da amostra versus distribuição Weibull.

Fonte: o autor.

Assim como a distribuição gama apresentado anteriormente, a distribuição de Weibull é composta por dois parâmetros que representam a curva de probabilidade, sendo eles Alfa e Beta. Para o calculo dos parâmetros foi utilizado a ferramenta Solver® para calcular os valores de Alfa e Beta que minimizam o valor de  $X^2$  calculado para distribuição Weibull a fim de testar a hipótese de  $H_0 =$  Amostra = Distribuição Weibull quando  $X^2$  da amostra  $< X^2$  da Distribuição Weibull. (Gráfico 4)

A distribuição de Weibull com os parâmetros calculados pelo solver foi a distribuição que teve maiores resultados com em relação a hipótese  $H_0$ , conforme a Tabela 6.



Tabela 6 – Teste de aderência de qui-quadrado amostra versus distribuição Weibull.

AMOSTRA					DISTRIBUIÇÃO WEIBULL		
	Intervalo Min.	Intervalo Max	Xi	Fi	Dist. Weibull	Fi Weibull	Teste
1	6	7	6,50	2	0,10150777	5	1,526
2	7	8	7,50	15	0,198684276	9	3,758
3	8	9	8,50	11	0,28171452	13	0,296
4	9	10	9,50	12	0,2468901	11	0,036
5	10	11	10,50	6	0,101824154	5	0,370
				46	X <sup>2</sup> Dist. Weibull		5,986270595
Passo:		1			X <sup>2</sup> Calculado		5,991464547

#### 4.3.4 Distribuição de Poisson

A fim de abranger mais distribuições no estudo também foi testado a aderência das premissas de análise com distribuição de Poisson, conforme o Gráfico 5.

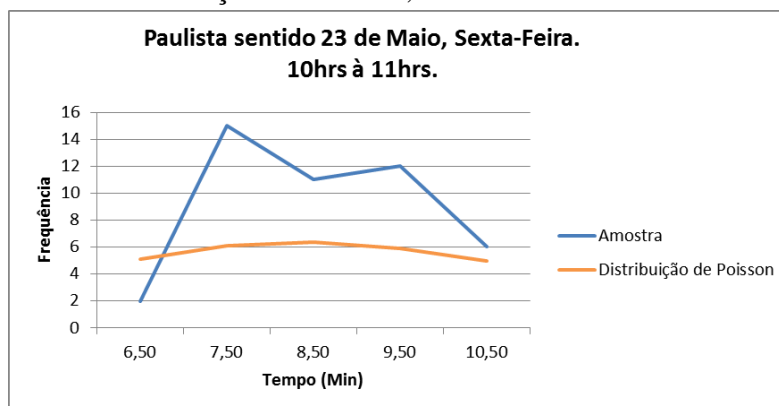


Gráfico 5 – Dados da amostra versus distribuição de Poisson.

Para calcular a função da curva de probabilidade foi utilizado o parâmetro da média do intervalo de confiança da amostra com o valor de lambda e os valores das médias dos passos como valores de X conforme exemplo do Gráfico 5.

A aderência da distribuição de Poisson não obteve sucesso nos teste de hipótese  $H_0 = \text{Amostra} = \text{Distribuição de Poisson}$  quando  $X^2$  da amostra <  $X^2$  da Distribuição de Poisson. Isso ocorreu, pois o intervalo de confiança acabou sugerindo um intervalo na maior parte das vezes menor que a abrangência da distribuição de Poisson. Como a distribuição de Poisson utiliza apenas a média e não utiliza o desvio padrão assim sua abrangência acaba seguindo um padrão que a própria distribuição segue. É possível verificar um exemplo do teste de aderência na Tabela 7.



Tabela 7 – Teste de aderência de qui-quadrado amostra versus distribuição de Poisson.

Média	8,38
-------	------

AMOSTRA					DISTRIBUIÇÃO POISSON		
	Intervalo Min.	Intervalo Max	Xi	Fi	Dist.Poisson	Fi.Poisson	Teste
1	6	7	6,50	2	0,110342764	5	1,864
2	7	8	7,50	15	0,132096051	6	13,105
3	8	9	8,50	11	0,138370614	6	3,375
4	9	10	9,50	12	0,128838416	6	6,224
5	10	11	10,50	6	0,107966593	5	0,215
				46	X <sup>2</sup> Dist.Poisson		24,78277216
Passo:		1			X <sup>2</sup> Calculado		5,991464547

Fonte: o autor.

#### 4.3.5 Distribuição exponencial

Para se calcular os valores da distribuição exponencial foi utilizado como parâmetro a média do intervalo de confiança das premissas de análises e a média entre os passos como os valores de X, conforme Figura 9, assim como no teste da distribuição de Poisson, conforme o Gráfico 6.

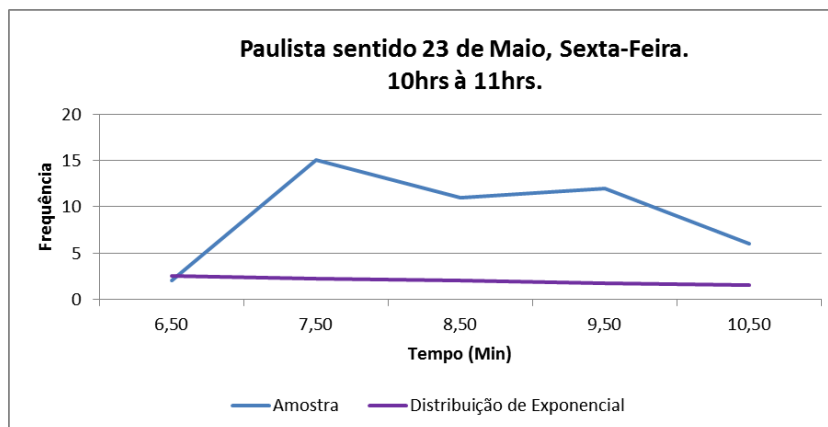


Gráfico 6 – Dados da amostra versus distribuição exponencial.

Fonte: o autor.

Como a curva da função exponencial tende sempre, quando visualizada em um limite curto, representar uma quase reta, sua aderência a hipótese  $H_0 = \text{Amostra} = \text{Distribuição Exponencial}$  quando  $X^2$  da amostra <  $X^2$  da Distribuição Exponencial, conforme a Tabela 8.



Tabela 8 – Teste de aderência de qui-quadrado amostra versus distribuição exponencial.

Média	8,38
-------	------

AMOSTRA					DISTRIBUIÇÃO EXPONENCIAL		
	Intervalo Min.	Intervalo Max	Xi	Fi	Dist.Expo	Fi.Expo	Teste
1	6	7	6,50	2	0,054940514	3	0,110
2	7	8	7,50	15	0,048760439	2	72,556
3	8	9	8,50	11	0,043275539	2	40,774
4	9	10	9,50	12	0,038407618	2	59,272
5	10	11	10,50	6	0,034087272	2	12,527
46					X <sup>2</sup> Dist.Exponencial		24,78277216
Passo:		1		X <sup>2</sup> Calculado			5,991464547

Fonte: o autor.

#### 4.4 INTERPOLAÇÃO POLINOMIAL DE NEWTON-GREGORY

Nas análises das distribuições estatísticas não foi obtido nenhum consenso de qual é a distribuição que melhor adere às amostras selecionadas pelas premissas de análise. Contudo foi analisada a possibilidade de se modelar uma distribuição que desenvolvesse uma aderência com as amostras. Para realizar essa modelagem recorreu-se a utilização de métodos de cálculo numérico para interpolações dos pontos da curva.

O método de interpolação que foi escolhido para representar a distribuição da amostra foi o método de Newton – Gregory. A escolha baseou-se no fato da interpolação polinomial de Newton – Gregory, calcular um polinômio qualquer de infinitos graus que passe pelos pontos da amostra.

Para a realização do cálculo da interpolação foi utilizado a técnica de uma pirâmide para auxiliar o cálculo dos parâmetros, para obtenção dos valores das variáveis auxiliares b, para a determinação numérica dos valores polinomiais conforme a exemplo da Tabela 9.

Tabela 9 – Utilização da técnica da pirâmide auxiliar.

NEWTON GREGORY								
		1	2	3	4	5	6	7
xi	fi	b1	b2	b3	b4	b5	b6	b7
12,50	0,00	0	0	14	-13,33	0	0	0
		0						
13,50	0,00	0	0	84				
		0						
14,50	0,00	84	84	-320				
		84		-236				
15,50	84,00	-68	-152					
		-68						
16,50	16,00							

Fonte: o autor



Por meio do cálculo dos parâmetros de Newton – Gregory pode se interpolar os valores e determinar qual o polinômio que descreve a amostra. Isso é, os valores polinomiais descrevem a curva da função que representa as frequências dos pontos das classes dos intervalos de confiança da amostra, de acordo com a Tabela 10.

Tabela 10 – Valores polinomiais resultantes da interpolação de Newton-Gregory.

Valores Polinomiais	
$x^6$	0
$x^5$	0
$x^4$	-1.41667
$x^3$	60,16667
$x^2$	-951,708
$x$	6642,958
1	-17251

Foi escolhida a determinação de polinômios de 6 graus, pois através dos testes das amostras foi constatado que com os polinômios de sexto grau se obtém números muito próximos de zero. Assim sendo, a partir do sexto grau os valores não seriam necessários por não influenciarem na construção da curva conforme pode ser verificado na Tabela 10.

#### 4.5 NORMALIZAÇÃO DO POLINÔMIO POR CÁLCULO DIFERENCIAL

O cálculo do polinômio representa as frequências da amostra escolhida pelas premissas de análises. Essa curva que representa a amostras não representa uma distribuição, mas apenas os dados, pois tanto é necessário que a curva represente as probabilidades de ocorrência dos tempos.

Com esse conceito foi pesquisado uma metodologia onde pudesse se realizar a normalização da curva a fim de criar uma distribuição que represente a probabilidade de ocorrência dos tempos das amostras.

A metodologia que se aplicou para a normalização da curva se baseou em tornar a área da curva do polinômio igual a 1 no intervalo representado pelo intervalo de confiança. Para isso realizou-se o cálculo da área  $A(P(x))$  da curva que representava a amostra através da integral da curva polinomial  $P(x)$  nos limites do intervalo de confiança conforme a equação 20.

$$A(P(x)) = \int_{Lim.Inf.IC}^{Lim.Sup.IC} P(x)dx \quad (20)$$

Sendo:  $Lim.Inf.IC$  = Limite inferior do intervalo de confiança;  $Lim.Sup.IC$  = Limite superior do intervalo de confiança.

Com a área da curva calculada realizou-se uma normalização para o valor 1, através do cálculo de um coeficiente de normalização (CN) representado pela equação 21.

$$CN = \frac{1}{A(P(x))}. \quad (21)$$

Ao multiplicar o polinômio pelo coeficiente de normalização foi obtida a distribuição resultante do polinômio esquematizado no Gráfico 7.



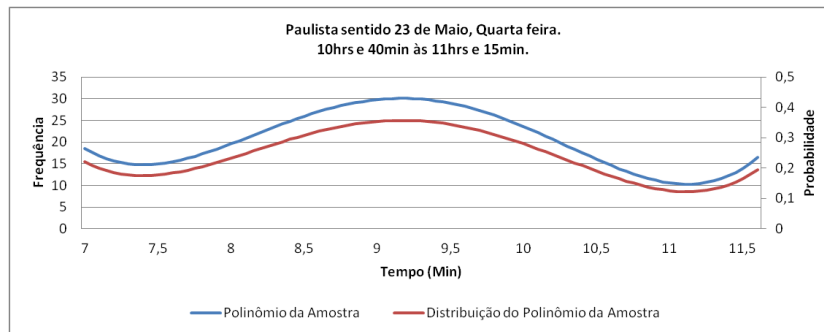


Gráfico 7 –Valores do polinômio da amostra versus distribuição normalizada do polinômio da amostra.

## 5 ARQUIVO DE ANÁLISES

Junto ao desenvolvimento dos estudos do presente trabalho foi desenvolvida uma ferramenta experimental para auxiliar os cálculos utilizados e também para o desenvolvimento das análises das informações provenientes dos testes das amostras.

O desenvolvimento da ferramenta resultou em um arquivo onde qualquer usuário possa consultar as informações necessárias através das escolhas das premissas de análises conforme a **Figura 11** e apertando o botão rodar.

Quando acionado o botão rodar a ferramenta começa a analisar a amostra proveniente das premissas de análises da seguinte forma:

- coletando as informações da organização de dados;
- testando a hipótese  $H_0$  para a aderência de qui-quadrado em relação as distribuições estudadas;
- constrói um gráfico das frequências da amostra versus as distribuições testadas;
- calcula o polinômio interpolador da amostra bem como sua respectiva distribuição;
- realiza um relatório com um resumo das informações para que o usuário da ferramenta possa consultar.

Conforme esquematizado na Figura 2. (não entendi, porque Figura 2?)

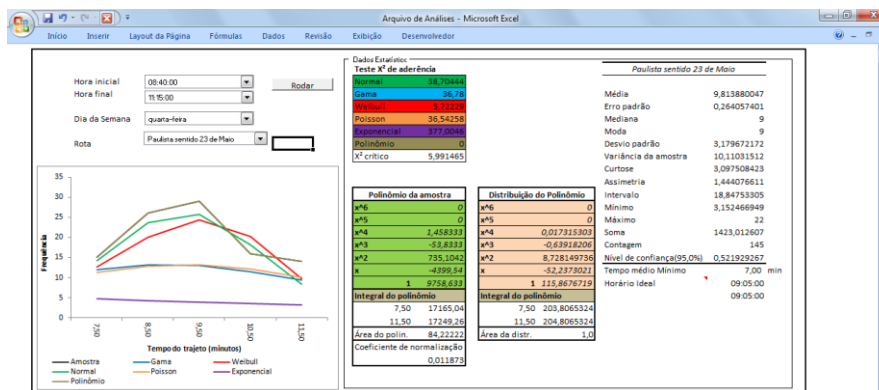


Figura 2 - Relatório da ferramenta arquivo de análise.  
 Fonte: Arquivo de Análises, o autor



## 2 CONSIDERAÇÕES FINAIS

A metodologia aplicada mostrou-se satisfatória, pois possibilitou o desenvolvimento de ferramentas e procedimentos para se estudar e entender melhor os fatores que influenciam as modelagens do trânsito e estabelecendo caminhos para analisar.

Os parâmetros da organização de dados que foram calculados nos testes demonstraram-se satisfatórios as medidas que os estudos foram se aprofundando. Os teste estatístico do intervalo de confiança de cada amostra correspondente às premissas de análises foram considerados aceitáveis quanto aos cálculos das distribuições que foram testadas.

Ao analisar as amostras dos tempos de trajeto das rotas observou-se que o intervalo de confiança, apresenta um espaço amostral onde que possibilitou a aplicação das distribuições estudadas e assim descartar os dados que puderam influenciar alguma tendência que não refletiam a amostra.

Com a utilização da ferramenta Solver® auxiliar do Excel®, foram calculados os parâmetros alfa e beta das distribuições gama e Weibull por meio dos métodos de minimização dos valores de  $X^2$  para testar a hipótese  $H_0$  à aderência com a amostras. A hipótese  $H_0$  também foi testada em relação às outras distribuições listadas para observar a aderência com a amostra e demonstrar qual a melhor distribuição para representar a amostra das premissas de análises.

Dentre as distribuições estudadas a que apresentou um maior índice de aderência em relação à amostra foi a de Weibull. Entretanto não foi categorizado como alternativa concreta para definir como padrão representativo das amostras definidas pelas premissas de análises.

As análises do presente trabalho demonstraram que é viável a utilização da normalização da interpolação polinomial para a representação da distribuição da probabilidade de ocorrência dos tempos da amostra das premissas de análises conforme foi apresentado na Figura 18.

O desenvolvimento de uma ferramenta que auxilie as análises estatísticas para as amostras das premissas de análises onde sejam demonstradas as aderências das distribuições bem como as informações da organização de dado e a distribuição polinomial é um aspecto que representa as informações ao usuário de maneira a consolidar os cálculos e assim auxiliar nas suas tomadas de decisões.

Essas informações podem auxiliar nas tomadas de decisões dos usuários civis do trânsito, como empresas que utilizam da logística das rotas para seus trajetos e até mesmo para as autoridades que necessitam tomar decisões que envolvem o trânsito.

O polinômio Newton Gregory se mostrou eficiente e foi possível descrever o tempo de trajeto utilizando sua função

O modelo foi construído, é funcional e respeita o cálculo de tempo médio com 99% de acerto. Este modelo é de excelente previsão, pois analisa o tráfego do dia da semana baseado num histórico recente Markoviano e descreve o tempo por comportamento lidando com estatística de tempo real.

A aderência mostra que o modelo criado é extremamente útil e pode ser utilizado, pois descreve o tráfego com excelente precisão.

Os erros do programa e os desvios padrões são pequenos o que mostra que a descrição é muito confiável.



## CREATION OF A probabilistic model for WEATHER FORECAST TRAFFIC ROUTES IN SAINT PAULO

**Abstract:** *The recent daily approaches about of congestion in the city of Sao Paulo has led to many questions about how to act to minimize the high levels of traffic in the city and also how to optimize the time which is spend on a trajectory. This work presents a development of a new application that includes several fields of statistics and analyzes data and new data with the aim of creating a probabilistic model to assist those involved to understand the pattern of the sample collected. To do so, will be related statistical methods of data organization, analysis, qualitative and quantitative adhesion test, statistical real-time and development polynomial distributions representing the samples, among others applied in the current system of urban mobility in Sao Paulo. As a result of a statistical analysis tool those variations of the sample.*

**Key-words:** *Transit, Data Organization, Adhesion Test, Chi-Square, Newton-Gregory, Rule Sturges, Standardization differential distributions.*