
APLICAÇÃO CONJUNTA DE TÉCNICAS DE ANÁLISE MULTIVARIADA DE DADOS E ANÁLISE ESPACIAL EXPLORATÓRIA PARA AVALIAÇÃO DO DESEMPENHO DE TRANSPORTE PÚBLICO

VIVIAN DE OLIVEIRA FERNANDES
CIRA SOUZA PITOMBO
MAURO JOSÉ ALIXANDRINI JUNIOR
DENISE VAZ DE CARVALHO SANTOS

Departamento de Transportes/ Escola Politécnica/ Universidade Federal da Bahia
Programa de Pós-Graduação em Eng. Ambiental Urbana
{vivian.fernandes, cira.pitombo, mauro.alixandrini, denise.vaz} @ufba.br

RESUMO - O principal objetivo do presente trabalho foi fazer uma análise do desempenho do sistema de transporte público por ônibus da cidade de Salvador sob a ótica do usuário. Foram utilizadas ferramentas espaciais e não espaciais para apoio à análise. As técnicas não espaciais utilizadas foram (1) Análise de Cluster, a qual identificou grupos de usuários completamente insatisfeitos e que consideram o serviço regular; (2) Árvore de decisão, que distinguiu as principais variáveis independentes que influenciaram a satisfação do usuário; e (3) Regressão Logística, com a finalidade de corroborar resultados obtidos pelas técnicas exploratórias. A análise espacial permitiu a visualização dos bairros críticos por região principal. Resultados apontaram como variáveis críticas a confiabilidade do sistema, a segurança, tarifa e locais de parada. A análise por região identificou a população residente nos locais mais pobres (A- Subúrbio e B – Miolo) como aquela mais descontente.

ABSTRACT - The main aim of the present work was analyzing the performance of public transportation system for bus of the Salvador city concerning the user view. Spatial and not spatial support tools had been used. The not spatial techniques used were: (1) Cluster Analysis, which identified groups of completely unsatisfied users and regular satisfaction regarding the service; (2) Decision Tree, that distinguished the main independent variables that influenced the satisfaction of the user; and (3) Logistic Regression, with the purpose to corroborate gotten results from the exploratory techniques. The spatial analysis allowed the visualization of the critical quarters for main regions. Results had pointed as critical variables the constancy of the system, the security, tariff and stop places. The analysis for region identified that the residents in poorer places (A -Subúrbio and B - Miolo) are the more unsatisfied.

1. INTRODUÇÃO

Atualmente, grande parte das cidades do mundo promove diferentes políticas urbanas de gestão e planejamento da mobilidade, priorizando o gerenciamento da demanda das viagens, em detrimento da criação de novas infra-estruturas para o transporte. A nova tendência da mobilidade urbana vai de encontro a políticas utilizadas no passado que favoreciam a construção ou ampliação permanente da infra-estrutura viária, incentivando o uso do automóvel, aumentando, conseqüentemente os congestionamentos. Desta forma, a atual dificuldade de circulação prejudicou inclusive o transporte público que é realizado junto com o transito em geral. A nova abordagem mundial está voltada ao incentivo ao uso de modos de transporte não motorizados e público.

Vale ressaltar que nos países desenvolvidos a maioria da população tem recursos financeiros para efetuar viagens por automóvel, como também dispõe de sistemas de transportes coletivos de boa qualidade, ao se contemplar atributos tais como regularidade, segurança e confiabilidade, por exemplo.

Em países em desenvolvimento, no entanto, especificamente no Brasil, deve-se atentar para fatores como falta de segurança nos transportes coletivos (especialmente por ônibus e micro-ônibus), bem como má qualidade no sistema de transporte público como o todo. Para reverter esta situação é necessário que haja o incentivo ao uso do transporte público, não somente do usuário cativo (em geral população de baixa renda), como especialmente também do atual usuário de transporte individual motorizado. Uma das principais dificuldades então seria tornar o transporte público atraente, já que, em média, o trajeto feito pelo ônibus demora 2,3 vezes mais que o automóvel, além da sua falta de flexibilidade, tempo necessário de espera devido a baixas frequências ou atrasos causados por congestionamentos, necessidade de transbordos, conduções lotadas, etc. Levando-se em conta tais considerações, torna-se necessário avaliar a qualidade do transporte público através de indicadores como tempo de viagem, frequência, conectividade, acessibilidade, lotação e etc., para que propostas de melhorias possam ser seguidas e políticas implementadas.

O objetivo do presente trabalho é avaliar o desempenho do transporte público por ônibus na cidade de Salvador através de ferramentas como técnicas de Análise Multivariada de Dados (AM) e Análise Espacial Exploratória (AEE). A análise não espacial utilizou três técnicas de forma conjunta: (1) Análise de *Cluster* – a fim de discriminar grupos de indivíduos totalmente insatisfeitos daqueles que avaliam o atual sistema como regular; (2) Árvore de Decisão – com objetivo de identificar regiões da cidade que apresentam piores desempenhos por indicador considerado bem como as variáveis mais importantes para o desempenho global do transporte público; e (3) Regressão Logística – com a finalidade de corroborar os resultados obtidos com a técnica de Árvore de Decisão. Para a representação dos dados espacialmente foi utilizada a técnica de mapeamento coroplético a fim de identificar de forma desagregada o desempenho dos indicadores.

2. ESTUDO DE CASO

Nos últimos anos, Salvador se deparou com um processo de transformação nos planos físico, econômico e social. Tais mudanças acarretaram significativas variações no padrão de mobilidade da cidade. Observou-se o aparecimento de múltiplos centros, maior diversificação das atividades, aumento na economia informal, assim como uma falta de controle do uso do solo. Além disso, o aumento de 94% da frota de automóveis em dez anos. Conseqüentemente, houve redução das viagens pendulares e aumento das viagens encadeadas, maior dispersão espacial e temporal dos fluxos de viagens, diluição dos horários de pico, maior exigência de rapidez e flexibilidade e fim do usuário cativo.

Salvador possui uma frota de 2445 ônibus, com idade média de 4,97 anos, dezoito empresas de ônibus, 445 linhas e uma rede viária de 1223 km. Em termos de transporte público, o Sistema de Transporte Público da cidade é operado basicamente por ônibus (467,9 milhões de passageiros por ano), micro-ônibus e uma linha de trem. Ainda possui quatro ascensores com 11,6 milhões de passageiros por ano.

Para efeito de organização do sistema integrado de linhas, a cidade foi dividida em quatro grandes áreas pelo Sistema de Transporte Coletivo por Ônibus (SCTO). A cidade foi separada em quatro regiões (Subúrbio, Miolo, Orla e Centro), as quais serão utilizadas no presente trabalho. A Figura 1 ilustra as citadas regiões assim como a Tabela 1 apresenta as principais características de cada uma delas.

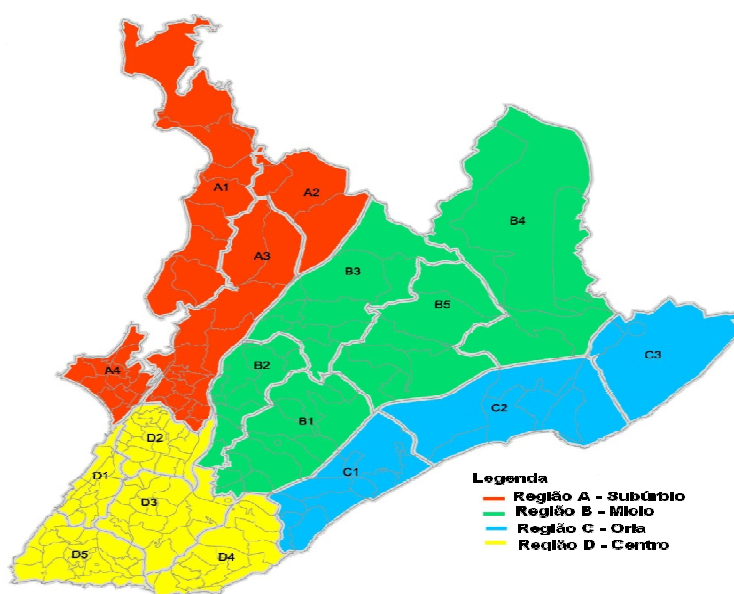


Figura 1 - Divisão do município SCTO

Tabela 1 - Quadro descritivo das regiões

A. Subúrbio: A área correspondente ao Subúrbio Ferroviário e compõe um dos maiores territórios de pobreza de Salvador. Teve sua ocupação iniciada pela construção da linha férrea. O subúrbio tem aproximadamente 500 mil habitantes, em sua maioria pobres e com baixa escolaridade. Há predominância de habitações precárias e deficientes. A região A é localizada na periferia da cidade, formada por trinta bairros. A população encontra-se segregada espacialmente, geralmente com longos percursos de viagem, com necessidade de transbordos. É atendida por 108 linhas e onze empresas.

B. Miolo: A região situa-se, em termos geográficos, na parte central do município de Salvador, possuindo cerca de 11.500 ha. O miolo urbano de Salvador começou a ser ocupado pela implantação de conjuntos residenciais para a “classe média baixa”, tendo a sua expansão continuada por loteamentos populares e sucessivas invasões coletivas. A área do miolo é formada por cerca de 41 bairros que ocupam aproximadamente 36% da superfície da cidade, em uma

densidade demográfica no intervalo de 15.000 – 25.000 hab/km². A área considerada do miolo urbano teve uma ocupação mista, com predominância de áreas residenciais, apresentando importantes pólos de atração de viagens. A região é atendida por 182 linhas e dezessete empresas.

C. Orla: A orla, especialmente a parte norte (orla marítima norte) constitui a “área nobre” da cidade, local de moradia, serviços e lazer, onde se concentram a riqueza, os investimentos públicos, os equipamentos urbanos e os interesses de produção imobiliária. A região também apresenta bairros populares com alta densidade demográfica como Boca do Rio. A região é composta de treze bairros, sendo operada por 8 empresas e 30 linhas regulares.

D. Centro: A área do centro, composta em geral por bairros bem atendidos em serviços e equipamentos urbanos, difere das áreas do subúrbio ferroviário e do miolo urbano que podem ser considerados como territórios de pobreza e violência. Nesta região também são encontrados vários pólos geradores, como: centro da cidade, *shoppings*, comércio, estações, centro histórico, hospitais, faculdades, escolas, praias e vários pontos turísticos e repartições públicas. Tendo a maior densidade demográfica e de atividades econômicas, é atendida por 18 empresas de ônibus, 124 linhas e contém 4 terminais de ônibus.

3. DADOS

Para obtenção da amostra, foram seguidas algumas etapas para obtenção do banco de dados: (1) elaboração dos questionários; (2) entrevistas em campo; (3) tabulação de dados e; (4) tratamento da amostra. A elaboração do questionário foi feita utilizando alguns parâmetros propostos por RODRIGUES (2006). Baseando-se ainda na metodologia do mesmo autor, o usuário deveria responder o seu grau de satisfação para cada parâmetro/indicador, bem como a importância de cada parâmetro. O questionário é composto por 25 questões relativas aos seguintes parâmetros: Acessibilidade (2); Segurança (3); Freqüência (1); Tempo de viagem (3); Lotação (1); Confiabilidade (1); Características dos veículos (3); Locais de parada (3); Sistema de Informação (1); Conectividade (3); Operadores (1); Vias (1); Tarifa (2).

O banco de dados final foi composto por 430 usuários de transporte público por ônibus residentes nas quatro regiões (Miolo, Subúrbio, Centro e Orla), 25 variáveis numéricas discretas (escala *Likert* com cinco posições – escore numérico atribuído pelo usuário quanto ao desempenho e importância de cada parâmetro) e uma variável contínua que representava o desempenho global do transporte público, variando de 1-25. Vale ressaltar que a amostra pode não ser suficiente representativa, considerando que 52% dos deslocamentos totais da cidade de Salvador são realizados por ônibus (Pesquisa Origem/ Destino Domiciliar realizada em 1995), no entanto, trata-se de uma avaliação preliminar de um trabalho realizado inicialmente com fins didáticos.

A Figura 2 traz um quadro com as médias das notas para cada parâmetro por região. Pode-se verificar nesta análise preliminar da amostra que as regiões Orla e Centro possuem usuários um pouco menos insatisfeitos quando comparados aos usuários residentes nas regiões do Miolo e Subúrbio.

Figura 2 - Dados regiões

<i>Fatores de Qualidade</i>	<i>(Suburbio)</i> A	<i>(Miolo)</i> B	<i>(Orla)</i> C	<i>(Centro)</i> D
Acessibilidade	2,60	2,90	3,20	2,90
Segurança	2,40	2,20	2,40	2,80
Freqüência	2,50	2,20	2,80	2,30
Tempo de viagem	2,20	2,10	3,80	2,60
Lotação	2,00	1,80	2,20	2,30
Confiabilidade	2,60	2,40	2,90	2,70
Veículos	2,70	2,40	2,70	2,60
Locais de Parada	2,40	2,00	2,30	2,30
Sistema de Info.	2,20	2,00	2,10	2,20
Conectividade	2,60	2,60	2,50	2,50
Operadores	2,80	2,60	2,90	2,90
Vias	2,80	2,60	3,00	2,90
Tarifas	2,50	2,70	2,80	2,80
Global	9,9	9,6	10,5	10,0

4. APLICAÇÕES DAS TÉCNICAS DE AM

As técnicas de Análise Multivariada de dados (AM) podem ser definidas como ferramentas analíticas que auxiliam na investigação acerca de fenômenos complexos envolvendo múltiplas dimensões. Identificam padrões que emergem de uma profusão de variáveis em interação. Hair *et al.* (1998) conceituam AM como um conjunto de técnicas estatísticas utilizadas com o objetivo de explicar e prever o grau de relações entre diversas variáveis independentes (inclusive entre si) e a variável dependente.

Existem diferentes técnicas multivariadas que podem ser utilizadas para diversas finalidades específicas. Neste trabalho, foram utilizadas três técnicas em conjunto: (1) Análise de *Cluster* – a fim de discriminar grupos de indivíduos totalmente insatisfeitos daqueles que avaliam o atual sistema como regular; (2) Árvore de Decisão – com objetivo de identificar regiões da cidade que apresentam piores desempenhos por indicador considerado bem como as variáveis mais importantes para o desempenho global do transporte público; e (3) Regressão Logística – com a finalidade de corroborar os resultados obtidos com a técnica de Árvore de Decisão.

5.1 Análise de *Cluster* (AC)

A Análise de *Cluster* (AC) é um conjunto de técnicas estatísticas cujo objetivo é agrupar objetos segundo suas características, formando grupos homogêneos. Os objetos em cada grupo tendem a ser semelhantes entre si, porém diferentes dos demais objetos dos outros conglomerados. Os conglomerados obtidos devem apresentar tanto uma homogeneidade interna (dentro de cada conglomerado), como uma grande heterogeneidade externa (entre conglomerados). Portanto, se a aglomeração for bem sucedida, quando representados em um gráfico, os objetos dentro dos conglomerados estarão muito próximos, e os conglomerados distintos estarão afastados (Hair *et al.*, 1998).

Foram obtidos dois grupos de usuários, homogêneos segundo a variável “Satisfação global” indicados na Tabela 3. A mesma tabela mostra as médias não somente da variável “Satisfação global”, como também dos indicadores considerados. As médias do grau de importância atribuído pelos usuários por indicador também estão representadas. Além disso, verifica-se a distribuição das regiões (A, B, C e D) em cada grupo bem como uma nomenclatura adotada pelos autores para caracterizar os dois grupos.

Tabela 3- Características gerais dos 2 aglomerados obtidos

Grupos	Satisfação Global				Região				Acessibilidade		Segurança	
	Média	Desvpad	Mínimo	Maximo	A(%)	B(%)	C(%)	D(%)	Satisf	Import	Satisf	Import
0	8,26	1,34	4,20	10,20	27,35	31,84	20,82	20,00	2,60	3,90	2,20	4,30
1	12,32	1,73	10,60	20,00	23,91	19,57	31,52	25,00	3,20	4,20	2,90	4,70
Grupos	Frequência		Tempo de Viag		Lotação		Caonfiabilidade		Veículos		Loc. de Parada	
	Satisf	Import	Satisf	Import	Satisf	Import	Satisf	Import	Satisf	Import	Satisf	Import
0	2,10	3,90	2,20	3,90	1,80	3,90	2,30	4,10	2,30	3,90	2,00	3,90
1	2,80	4,50	2,60	4,30	2,50	4,30	3,40	4,50	2,90	4,40	2,50	4,40
Grupos	SI		Conectividade		Operador		Vias		Tarifa		Nomeclatura	
	Satisf	Import	Satisf	Import	Satisf	Import	Satisf	Import	Satisf	Import		
0	1,90	4,00	2,40	3,90	2,60	4,10	2,60	4,20	2,40	4,10	Insatisfeitos	
1	2,30	4,40	2,80	4,30	3,00	4,50	3,10	4,60	2,90	4,40	Satisfeitos	

A nomenclatura foi adotada considerando a média, desvio padrão, valor mínimo e máximo da “Satisfação global” por grupo. Assim, o primeiro grupo (grupo 0) caracteriza os usuários “**Completamente insatisfeitos**”, com valores de “satisfação global” muito baixos. Este grupo é composto predominantemente por usuários residentes nas regiões A (subúrbio) e B (Miolo). Tais regiões são as mais pobres e carentes da cidade. Usuários residentes no subúrbio, por exemplo, enfrentam altos tempos de viagem, já que os mesmos encontram-se espacialmente segregados e precisam percorrer altas distâncias, além da necessidade de transbordo. Considerando valores médios de satisfação e importância por indicador, o grupo dos usuários “**Completamente insatisfeitos**” apresenta valores baixos (abaixo ou igual a 2,50) de desempenho e altos de importância para a maioria dos indicadores, exceto “Acessibilidade” e “Comportamento dos operadores”.

O grupo 1, dos usuários que julgam o serviço de transporte público por ônibus de Salvador “regular”, denominados “**Usuários com satisfação regular**”, é composto por usuários que na sua maioria, possuem valores médios de “Satisfação global”. Mais frequentemente são residentes das regiões C (Orla) e D (Centro). Conforme descrição sucinta da Tabela 1, ambas as regiões são mais desenvolvidas e bem atendidas em relação a serviços e equipamentos urbanos. No entanto, vale ressaltar que, de uma maneira geral, o usuário está descontente com o serviço de ônibus, independente da região que reside. Fatores com “Lotação”, “Locais de parada” e “Sistemas de Informação” merecem uma atenção especial para esse grupo de usuários.

5.2 Árvore de Decisão

Outra técnica utilizada neste trabalho, em conjunto com a AC, é Árvore de Decisão (AD), considerada uma forma simples de representação de relações existentes em um conjunto de dados. Os dados são divididos em subgrupos, com base nos valores das variáveis. O resultado é uma hierarquia de declarações do tipo "Se ... então ..." que são utilizadas, principalmente, para classificar dados.

Uma árvore de decisão pode ser definida como um gráfico acíclico e direto que satisfaz as seguintes propriedades: (1) A hierarquia é denominada árvore e cada segmento é denominado nó; (2) Há um nó, chamado raiz, que contém todo o banco de dados; (3) Este nó contém dados que podem ser subdivididos dentro de outros sub-nós,

chamados de nós filhos; (4) Existe um único caminho entre o nó raiz e cada nó; (5) Quando os dados do nó não podem ser mais subdivididos dentro de um outro subconjunto ele é considerado um nó terminal ou folha.

Para processamento da árvore foi utilizado o *Software* SPSS 15.0. Utilizou-se o algoritmo do CART, que estabelece uma relação entre variáveis independentes e variável dependente.

A árvore foi gerada a partir da amostra final com 430 usuários, variável dependente (*clusters* – variável com duas categorias – grupos descritos na sub-seção anterior), adotando-se o mínimo de 25 observações por nós filhos. As variáveis independentes foram aquelas que representavam a satisfação nos indicadores considerados no trabalho, bem como a região que reside (variável com quatro categorias: A, B, C e D).

A variável de maior importância (que melhor explica a variabilidade dos dados) é “Confiabilidade”. A partir da raiz, a árvore se ramifica em dois grupos principais: (1) “Confiabilidade” menor ou igual a 2.5 (nó 1 – 77.9% completamente insatisfeitos) e (2) “Confiabilidade” superior a 2.5 (nó 2 – 59.2% satisfação regular). Tal variável relaciona-se ao grau de certeza dos usuários de que o veículo de transporte público vai passar na origem e chegar ao destino no horário previsto (Ferraz e Torres, 2004). Diante deste conceito, pode-se afirmar que a “Confiabilidade” engloba critérios como “Tempo de viagem”, “Frequência” e “Conectividade”. Este fato pode explicar a sua importância quanto à variabilidade em relação à variável dependente (indivíduos completamente insatisfeitos e indivíduos que acham o transporte público regular).

Posteriormente ocorrem novas segmentações do conjunto de dados considerando variáveis como “Segurança”, “Tarifa”, “Locais de parada” e “Região”. Ao final da segregação dos dados foi encontrado um total de 6 folhas. A Figura 3 representa a árvore gerada. Nas folhas encontram-se ilustrados as categorias da variável dependente (Grupo 0 e Grupo 1) e a frequência de cada grupo em cada nó.

Observando-se os nós terminais, pode-se afirmar que “Tarifa” pode interferir na satisfação do transporte público por ônibus. As folhas 3 e 4 apresentam grau de satisfação quanto à confiabilidade baixo (menor ou igual a 2.5) e são diferenciadas pelo grau de satisfação da tarifa (nó 3 – “Tarifa” ≤ 3.2; nó 4 – “Tarifa” > 3.2). O grupo de indivíduos que compõem o nó 3 predominantemente é formado por usuários completamente insatisfeito em relação ao serviço de transporte público por ônibus (Grupo 0 – 85.1%) enquanto que o a maioria dos indivíduos classificados no nó 4 julga o serviço de transporte público de Salvador regular (Grupo 1 – 62.1%).

“Segurança” foi uma variável que se mostrou de extrema importância para discriminar os grupos de indivíduos quanto à sua satisfação global. A segmentação dos dados a partir do nó 2 mostra que quando o valor de satisfação atribuído para o parâmetro “Segurança” é inferior ou igual a 2.8, a maioria dos indivíduos permanece extremamente insatisfeita, mesmo que o grau de “Confiabilidade” seja relativamente alto (nó 5 – Grupo 0 – 55.4%). No entanto, aliando-se alto grau de satisfação relativo à “Confiabilidade” e alto grau de satisfação relacionado à “Segurança”, verifica-se que a maior parte dos usuários considera o transporte regular (nó 6 – Grupo 1 – 80.2%).

Finalmente, observa-se a influência da variável “Locais de parada” (nó 7 e 8) na satisfação do usuário, bem como da variável “Região” (nó 9 e 10). Verifica-se novamente que indivíduos residentes nas regiões A e B (Subúrbio e Miolo, respectivamente) estão menos satisfeitos do que aqueles residentes nas regiões melhor atendidas C e D (Orla e Centro, respectivamente). Vale ressaltar ainda que, apesar do alto grau de satisfação atribuído a “Confiabilidade” (superior a 2.5) e “Segurança” (superior a 2.8), a região que usuário reside, e conseqüentemente, inicia e finaliza pelo menos a primeira e última viagem do dia pode influenciar seu grau de satisfação.

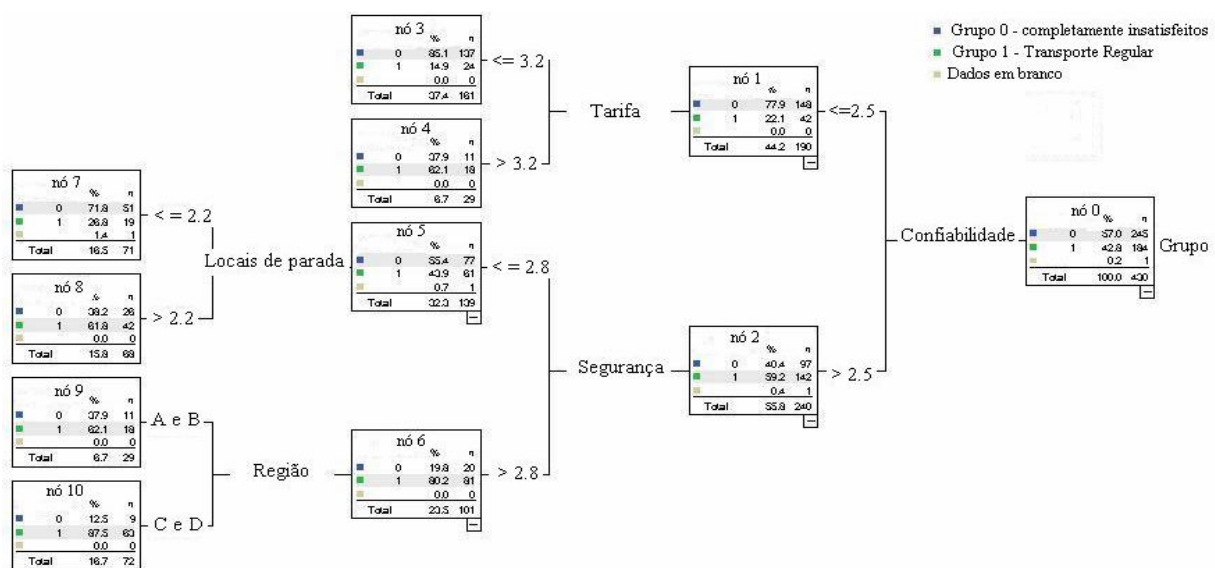


Figura 3 - Árvore de Decisão – Satisfação do Transporte Público

5.3 Regressão Logística

A regressão logística é uma forma especializada de regressão que é formulada para prever e explicar uma variável categórica binária, e não uma medida dependente numérica.

No presente trabalho, o uso de regressão logística deu-se a fim de corroborar resultados inicialmente obtidos com a Árvore de Decisão. Assim, através dos coeficientes das variáveis, bem como parâmetros que mensurem o ajuste do modelo geral ($-2LL$, R^2 de *Cox e Snell*, R^2 de *Nagelkerke*), é possível avaliar a qualidade de ajuste do modelo, verificar a importância das variáveis independentes no cálculo da probabilidade dos usuários estarem no grupo dos usuários que consideram o transporte público regular (Grupo 0 (completamente insatisfeitos) = 0; Grupo 1 (satisfação regular) = 1), bem como identificar se o modelo é bem explicado pelas variáveis independentes. Assim, para estimação do modelo logístico foram consideradas as variáveis descritas na Tabela 4.

Tabela 4 - Variáveis consideradas no modelo logístico

<i>Dependente</i>		<i>Descrição</i>	<i>Independentes</i>		<i>Descrição</i>	
Satisfação	0	Completamente insatisfeitos	Acessibilidade - Satisfação Segurança - Satisfação Frequência - Satisfação Tempo de Viagem - Satisfação Lotação - Satisfação Confiabilidade - Satisfação Veículo - Satisfação Locais de Parada - Satisfação Sistema de Informação - Satisfação Conectividade - Satisfação Operadores - Satisfação Vias - Satisfação Tarifa - Satisfação	1- péssimo 2- ruim 3- regular 4 - bom 5- ótimo		
	1	Satisfação regular				
<i>Independentes</i>		<i>Descrição</i>				
Região A	0	não reside no subúrbio				
	1	reside no subúrbio				
Região B	0	não reside no miolo				
	1	reside no miolo				
Região C	0	não reside na orla				
	1	reside na orla				
Região D	0	não reside no centro				
	1	reside no centro				

O modelo de regressão logístico foi estimado utilizando o método *Forward Stepwise: Wald*. Neste método cada variável é incluída separadamente em cada passo do modelo e analisado seu desempenho em estimar a “satisfação” do usuário. O processo termina até as variáveis que melhor explicam conjuntamente a variável binária “satisfação” forem selecionadas, chegando ao modelo final. Observa-se que o processo foi formado por 7 passos. O sétimo modelo apresentou valores menores de $-2LL$, bem como valores maiores de R^2 de *Cox e Snell*, R^2 de *Nagelkerke*. Além disso, o modelo apresentou uma frequência de acerto de 81,6% para o grupo dos “completamente insatisfeitos” (Grupo 0) e 73,2% para o grupo daqueles que consideram o transporte regular (“satisfação regular – Grupo 1). A Tabela 5 traz o sumário dos principais parâmetros dos modelos obtidos no método *stepwise*. O modelo considerado no trabalho foi sétimo, mais bem ajustado.

Tabela 5 - Sumário dos modelos propostos na *stepwise*

<i>Modelos</i>		<i>% Acertos</i>	<i>-2 Log likelihood</i>	<i>Cox&Snell R Square</i>	<i>Nagelkerke R Square</i>
1	<i>Grupo 1</i>	82,91	486,090	0,186	0,249
	<i>Grupo 0</i>	53,55			
2	<i>Grupo 1</i>	79,49	448,730	0,256	0,343
	<i>Grupo 0</i>	65,03			
3	<i>Grupo 1</i>	78,63	420,375	0,305	0,408
	<i>Grupo 0</i>	71,58			
4	<i>Grupo 1</i>	81,20	403,328	0,332	0,445
	<i>Grupo 0</i>	70,49			
5	<i>Grupo 1</i>	82,91	373,841	0,378	0,507
	<i>Grupo 0</i>	72,13			
6	<i>Grupo 1</i>	81,62	368,303	0,386	0,518
	<i>Grupo 0</i>	71,04			
7	<i>Grupo 1</i>	81,62	363,698	0,393	0,527

A Tabela 6 mostra as variáveis que entraram no modelo (modelo 7), gerada pela ferramenta de regressão logística bem como os seus coeficientes e os resultados do teste de significância das variáveis independentes. Na coluna “B (coeficientes)”, são apresentados, os coeficientes do modelo para cada variável explicativa. Já na coluna Erro Padrão, é apresentada uma medida de variabilidade destes coeficientes. A estatística *Wald* é utilizada para testar a hipótese de cada coeficiente do modelo de regressão. Tal teste irá identificar o quanto à variável explicativa participa individualmente da explicação da variável dependente (“satisfação”). A estatística de *Wald*, testa para cada coeficiente

das variáveis do modelo a seguinte hipótese de que o coeficiente da variável explicativa é igual a zero contra uma hipótese alternativa de que não são iguais a zero. A coluna “Exp(B)” é o exponencial dos coeficientes estimados de cada categoria variável do modelo.

Tabela 6 - Variáveis explicativas do modelo

Variáveis do modelo		B(coef)	Erro padrão	Wald	G. J.	Sig	Exp(B)
Modelo 7	Acessibilidade	0,36	0,15	6,27	1	0,01	1,49
	Segurança	0,74	0,19	14,24	1	0,00	2,10
	Frequência	0,36	0,13	7,09	1	0,01	1,44
	Conforto	0,65	0,15	18,53	1	0,00	1,92
	Loc. de Parada	0,62	0,18	11,37	1	0,00	1,88
	Operador	0,29	0,14	4,52	1	0,03	1,35
	Tarifa	0,65	0,17	14,27	1	0,00	1,93
	Canstant	-10,07	1,0	101,32	1	0,00	0,00

Considerando os parâmetros apresentados na tabela, verifica-se que as variáveis com maiores valores de coeficientes são, em ordem decrescente “Segurança”, “Tarifa”, “Confiabilidade” e “Locais de Parada”. Todas estas variáveis também foram selecionadas pela Árvore de Decisão e têm um poder discriminatório de grupos de satisfação de usuários de transporte público. Valores altos de grau de satisfação especialmente destas quatro variáveis aumentam a probabilidade de determinado usuário considerar o sistema de transporte público regular. A variável “Região”, por sua vez, não entrou no modelo logístico. Pode-se afirmar que possivelmente “satisfação” não está apenas relacionada a quesitos estritamente relacionados ao transporte público como frequência, tempo de viagem, etc, mas, sobretudo a problemas de segurança pública (como segurança, por exemplo) e sociais (valor da tarifa). A próxima seção traz uma análise espacial exploratória das principais variáveis selecionadas tanto pela Árvore de Decisão como pelo modelo logístico. É possível visualizar, dentro de cada uma das quatro regiões, padrões diferentes relacionados ao grau de satisfação.

6. ANÁLISE ESPACIAL EXPLORATÓRIA

O planejamento das atividades que acontecem no espaço urbano tem se aperfeiçoado muito com a adoção de técnicas exploratórias que permitem a análise de uma quantidade cada vez maior de dados que são registrados para os mais diversos fins. Nesse campo de conhecimento destaca-se a análise de dados espaciais georreferenciados que devido aos avanços da computação, referentes à capacidade de processamento dos equipamentos e a evolução dos algoritmos de análise de dados vetoriais, aumentaram as possibilidades do emprego de técnicas de análise espacial. Em paralelo, uma série de técnicas estatísticas vem sendo definidas e adaptadas aos dados espaciais. Estas técnicas, combinadas com funções de visualização, formam, um conjunto de ferramentas que também são conhecidas como análise exploratória de dados espaciais (ANSELIN & BAO, 1997).

Nas análises tabulares, por exemplo, apenas correlações podem ser deduzidas dos conjuntos de dados, enquanto a associação da variável espacial dos elementos pode revelar diferentes graus de autocorrelação que pode ser descrita como influência da vizinhança na variável estudada.

Com os dados tabulares foram gerados vários mapas de diferentes temas os quais foram analisados neste trabalho. Foi utilizada uma base de dados vetoriais com os limites de bairros de Salvador. Verifica-se que o levantamento de dados não foi realizada em todos os bairros da cidade, uma vez que o município foi dividido em regiões, porém a análise espacial foi realizada levando em consideração os bairros onde moravam dos entrevistados, ocasionando algumas regiões em branco. Recomenda-se que em futuros estudos, uma amostra maior de dados seja levantada para maior efetividade da representação.

Para a representação dos dados espacialmente foi utilizada a técnica de mapeamento coroplético. Os mapas coropléticos são elaborados com dados quantitativos e apresentam sua legenda ordenada em classes conforme as regras próprias de utilização da variável visual valor por meio de tonalidades de cores. Levando-se em consideração que a variável investigada estima os valores de satisfação podem ser positivos, com conceito bom e excelente e negativo, com conceito ruim e péssimo. Para uma analogia ao semáforo, utilizaram-se as cores vermelho para avaliações negativas, verde para as positivas e amarela para as avaliações intermediárias.

Os mapas no modo de implantação zonal são os mais adequados para representar distribuições espaciais de dados que se refiram as áreas, que no caso deste trabalho a análise foi realizada levando-se em consideração os bairros onde os usuários entrevistados residiam.

Foram confeccionados mapas temáticos para todos os critérios avaliados na pesquisa. Para exemplificação desta análise são apresentados nas figuras 4 e 5 os mapas referentes aos critérios: segurança e a média global ponderada de todos os critérios observados.

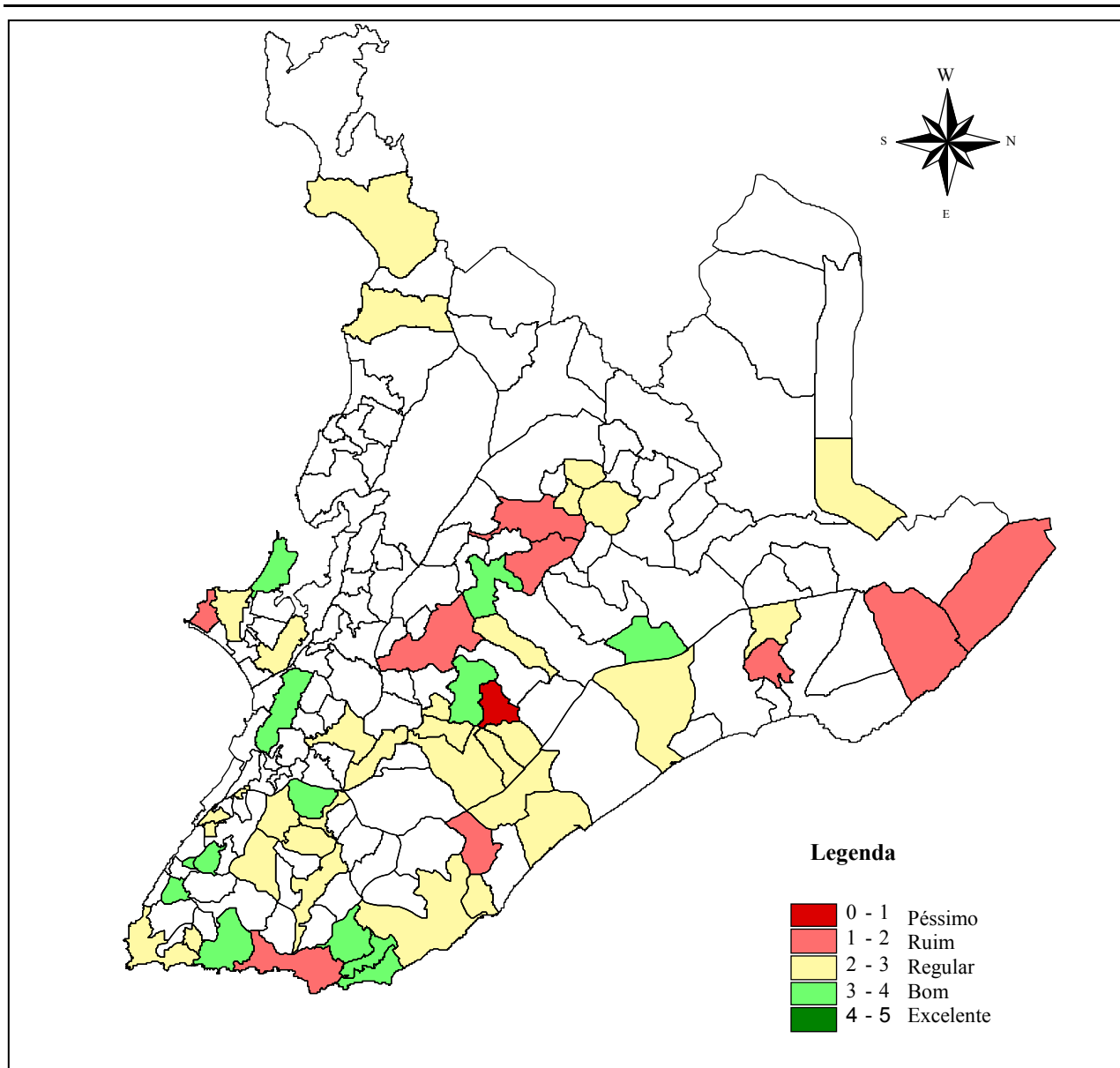


Figura 4 - Mapa de satisfação do usuário quanto à segurança no transporte público no Município de Salvador/BA.

Na figura 4, é apresentado o mapa de satisfação do usuário quanto à segurança no transporte público no município de Salvador/BA. A legenda mostra a correspondência entre as notas atribuídas na segurança do transporte público e os conceitos. Aos bairros onde houveram residentes entrevistados são atribuídas as tonalidades graduadas em vermelho, amarelo e verde, uma analogia ao semáforo e os bairros em branco, são aqueles que não tiveram entrevistados com residência nos mesmos.

Ao analisar as tendências apresentada no mapa da figura 4, verifica-se que a região do miolo apresenta alternância entre avaliações positivas e negativas, sem uma tendência homogênea, revelando a necessidade de uma investigação mais aprofundada e uma amostra mais representativa de dados para esta região. Ao contrário da região da orla que apresenta uma tendência predominantemente negativa. No subúrbio verifica-se uma avaliação regular da segurança no transporte público e no centro predominantemente positiva, com exceção do bairro Rio Vermelho.

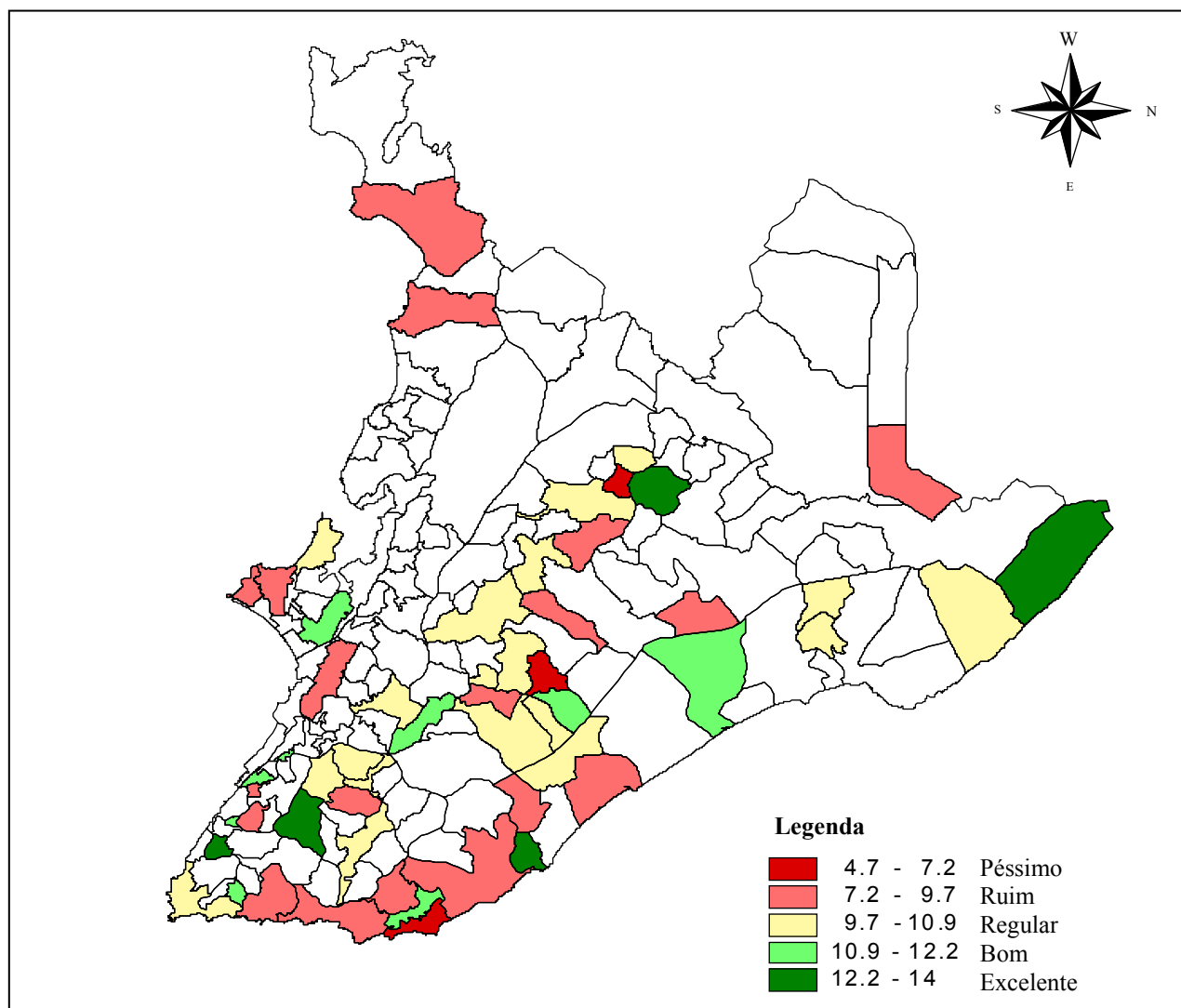


Figura 5 - Mapa Global de satisfação do usuário no Município de Salvador/BA.

Através da análise da figura 5, o mapa apresenta o comportamento das avaliações por bairro ponderadas pela importância dada pelo usuário a cada um dos critérios. Neste mapa a região da orla apresenta avaliações regulares e positivas, isto reflete que para os critérios de maior importância dos usuários naquela região, a avaliação é predominantemente positiva. Ao contrário do que se percebe nas regiões do centro e do subúrbio e onde a avaliação é mais negativa, porque nos critérios mais importantes aos usuários, o transporte obteve uma avaliação desfavorável. Percebe-se que a avaliação dos usuários, levando-se em consideração da importância dada aos critérios, reflete um resultado mais negativo para estas duas regiões, do que no mapa apresentado na figuras 4. Na região do miolo ainda se percebe que na periferia do miolo ainda percebe-se alternância entre as avaliações positivas e negativas. Mas com esta ponderação é possível definir uma tendência, de que as avaliações positivas estão na região periférica do miolo e na região central, avaliações negativas e regulares.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anselin, L. & Bao, S. **Exploratory Spatial Data Analysis Linking SpaceStat and ArcView**. In: Fischer, M. M. & Getis, A. Recent developments in spatial analysis. New York: Springer, 1997, p. 35-59.
- Hair, J.F.; R.E Anderson.; R.L Tatham.; W.C. Black (1998). **Multivariate Data Analysis**. 5ª ed. Prentice-Hall. Upper Saddle River, New Jersey, 730p.
- Rodrigues, M. O. (2006). **Avaliação da qualidade do transporte coletivo da cidade de São Carlos**. Dissertação Escola de Engenharia de São Carlos, Universidade de São Paulo.