

PREVISÃO DE ESCOLHA MODAL EM UM CAMPUS UNIVERSITÁRIO UTILIZANDO A REGRESSÃO LOGÍSTICA BINOMIAL

F. A. Ferreira, C. S. Pitombo e A. L. Cunha

RESUMO

Este trabalho apresenta um método para previsão de escolha modal de um *campus* universitário utilizando como ferramenta a regressão logística binomial. Os dados analisados referem-se a uma pesquisa realizada com os usuários do *campus* da Universidade de São Paulo, localizado em São Carlos, SP, Brasil. A partir de variáveis socioeconômicas e que caracterizam os principais modos de viagem, foram calibrados diferentes modelos para previsão de escolha modal. Em todos os casos, os valores de parâmetros estimados para cada variável independente foram considerados significativos, bem como os sinais calibrados foram coerentes. O poder preditivo (Pseudo R^2) mostrou-se alto e a validação apresentou taxas de acertos acima de 80% para todos os modelos. A aplicação desta ferramenta para previsão de escolha modal se mostrou adequada, possibilitando fazer uma análise dos principais fatores ou variáveis que influenciam nos modos escolhidos para acesso ao *campus*.

1 INTRODUÇÃO

As universidades são caracterizadas como grandes Polos Geradores de Viagens (PGVs) devido ao grande número de viagens gerado diariamente. Inseridos em meios urbanos ou rurais, muitas vezes os *campi* universitários possuem um sistema de mobilidade complexo, considerando as diversas atividades inseridas nestes espaços. Estes deslocamentos tendem a impactar, de maneira negativa, o sistema viário no qual estão inseridos (Oliveira, 2015).

Compreender os padrões de deslocamentos dos usuários é uma importante informação para o planejamento e gestão destas áreas. A fim de mitigar os efeitos negativos das viagens atraídas ou produzidas por PGVs, mais especificamente universidades, pode-se intervir de forma a promover a mobilidade sustentável e eficiente de estudantes, funcionários e visitantes.

Em planejamento de transportes, diversos estudos tratam de previsão de escolha modal. No entanto, a produção científica referente a escolhas de modos de transporte para universidades ainda é escassa (Jacques *et al.*, 2011). Assim, o presente trabalho tem por objetivo apresentar um método tradicional, utilizando a regressão logística binomial como ferramenta, para previsão de escolha modal de um *campus* universitário.

O *campus* escolhido refere-se à Universidade de São Paulo, localizada em São Carlos, São Paulo, Brasil. Neste trabalho, serão estimados modelos logísticos para prever a probabilidade de um usuário do *campus* utilizar determinado modo de transporte. Os modelos previstos consideraram os três principais modos de transporte utilizados para acessar o *campus*.

Este artigo é formado por cinco seções, além desta introdução. A seção 2 apresenta conceitos da técnica adotada para modelagem da escolha modal. A seção 3 descreve os materiais e procedimento metodológico, utilizados pelos autores. Os resultados e discussões são apresentados na seção 4. Finalmente, as principais conclusões são descritas na seção 5, anterior às referências bibliográficas.

2 REGRESSÃO LOGÍSTICA BINOMIAL

A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias.

Esta técnica permite estimar a probabilidade direta de ocorrer um determinado evento. No caso de um conjunto de n variáveis independentes ($X_1, X_2, \dots, X_n$) que caracterizam a ocorrência de um evento com dois resultados possíveis, o modelo de regressão logística pode ser escrito como apresentado na Equação 1.

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}} \quad (1)$$

Sendo:

P : probabilidade da variável dependente ser igual a 1;

$\beta_0, \beta_1, \dots, \beta_p$: coeficientes de calibração;

$X_1, X_2, \dots, X_n$: variáveis independentes;

n : número de variáveis independentes.

Os coeficientes estimados ($\beta_0, \beta_1, \dots, \beta_n$) indicam a variação da probabilidade de ocorrência de um evento. A interpretação destes parâmetros é obtida conhecendo o impacto que ele exerce sobre a razão da chance. Para isso, deve-se aplicar o inverso do logaritmo do próprio coeficiente. Assim, um coeficiente positivo indica aumento na probabilidade, ao modo que um valor negativo representa diminuição da probabilidade prevista. A relação entre as variáveis independentes e dependentes é representada por uma curva que tem comportamento probabilístico no formato de S, característica da regressão logística. Esse formato dá a regressão logística alto grau de generalidade (Hair Jr. *et al.*, 2009).

Para utilizar o modelo de regressão logística para discriminação de dois grupos, é necessário definir uma regra de classificação. De forma que, se o valor previsto estiver acima do valor definido na regra de classificação, aceita-se a hipótese atribuída ao número 1. Do contrário, aceita-se a atribuição dada ao valor 0.

Para determinar o ajuste geral do modelo, ou seja, medir seu poder de explicação, são utilizados, na regressão logística, métricas similares às usadas na análise de regressão tradicional, em que se calcula os chamados coeficientes de determinação (R^2). Na regressão logística, estes coeficientes recebem a denominação de Pseudo R^2 . Existem várias métricas para este Pseudo R^2 , tais como Cox & Snell e Nagelkerke (Hair Jr. *et al.*, 2009).

Os modelo de escolhas discretas de modo de transportes por regressão logística preveem a probabilidade direta de um evento ocorrer. Uma das vantagens da regressão logística é que se precisa saber apenas se um evento ocorreu para então utilizar o valor dicotômico da

Atualmente frequentam o *campus*, aproximadamente, 9.000 usuários regulares, constituídos de alunos de graduação, pós-graduação, funcionários e docentes (USP, 2016). Em média, 80% dos usuários frequentam a Área I e 20% a Área II (Oliveira, 2015). Para realizar a ligação entre as duas áreas, que distam cerca de 3,5 km, a universidade disponibiliza transporte gratuito por ônibus para os usuários do *campus*.

3.2 Banco de Dados

O banco de dados a ser estudado se trata de um questionário, parte da dissertação de mestrado de Stein (2013), aplicado, pela internet, a alunos de graduação, pós-graduação, funcionários técnico-administrativos e docentes no ano de 2012. Este estudo visava caracterizar e entender as condições de mobilidade do *campus*. O questionário forneceu 2.276 respostas válidas (25% da população total do *campus*) com uma taxa de resposta bem distribuída dos principais usuários. As informações coletadas referem-se a dados pessoais, caracterização dos deslocamentos, motivação para mudança de modo de transporte, políticas institucionais, entre outras.

Tendo em vista que as áreas possuem características diferentes, não foi possível realizar uma análise global do *campus*. Assim, dividiu-se o banco de dados considerando as variáveis presentes em cada uma das áreas. Após tratamento inicial e seleção das variáveis relevantes, o banco de dados para a Área I é formado por 757 registros e um total de 22 variáveis e o banco de dados para a Área II possui 550 registros e um total de 20 variáveis. As Tabelas 1 e 2 apresentam a lista completa das variáveis para cada área, a descrição de cada uma delas, bem como o seu tipo.

Tabela 1 Descrição das variáveis do banco de dados para Área I

VARIÁVEL	DESCRIÇÃO	TIPO
Gênero	Gênero dos usuários	Dicotômica
Idade até 18 anos	Usuários que possuem até 18 anos de idade	Dicotômica
Idade de 18 a 24 anos	Usuários que possuem idade entre 18 e 24 anos	Dicotômica
Idade de 25 a 30 anos	Usuários que possuem idade entre 25 e 30 anos	Dicotômica
Idade de 31 a 39 anos	Usuários que possuem idade entre 31 e 39 anos	Dicotômica
Idade de 40 a 49 anos	Usuários que possuem idade entre 40 e 49 anos	Dicotômica
Idade de 50 a 59 anos	Usuários que possuem idade entre 50 e 59 anos	Dicotômica
Idade maior que 60 anos	Usuários que possuem idade maior que 60 anos	Dicotômica
Acesso a carro e/ou moto	Usuários que possuem acesso regular a carro e/ou moto	Dicotômica
Acesso a bicicleta	Usuários que possuem acesso regular a bicicleta	Dicotômica
Vínculo - Graduação	Alunos de graduação	Dicotômica
Vínculo - Pós-graduação	Alunos de pós-graduação	Dicotômica
Vínculo - Funcionários	Funcionários técnico/administrativos	Dicotômica
Vínculo - Docentes	Docentes	Dicotômica
Área I - a pé	Principal modo de deslocamento a Área I é a pé	Dicotômica
Área I - bicicleta	Principal modo de deslocamento a Área I é a bicicleta	Dicotômica
Área I - ônibus municipal	Principal modo de deslocamento a Área I é o ônibus municipal	Dicotômica
Área I - motocicleta	Principal modo de deslocamento a Área I é a motocicleta	Dicotômica
Área I - carro	Principal modo de deslocamento a Área I é o carro	Dicotômica
Área I - carona	Principal modo de deslocamento a Área I é a carona	Dicotômica
Tempo até a área I	Tempo médio de deslocamento, em minutos, até o local de estudo/trabalho na Área I	Contínua
Mínima distância de deslocamento	Mínima distância para o usuário percorrer entre sua casa e o local de estudo/trabalho	Contínua

Tabela 2 Descrição das variáveis do banco de dados para Área II

VARIÁVEL	DESCRIÇÃO	TIPO
Gênero	Gênero dos usuários	Dicotômica
Idade até 18 anos	Usuários que possuem até 18 anos de idade	Dicotômica
Idade de 18 a 24 anos	Usuários que possuem idade entre 18 e 24 anos	Dicotômica
Idade de 25 a 30 anos	Usuários que possuem idade entre 25 e 30 anos	Dicotômica
Idade de 31 a 39 anos	Usuários que possuem idade entre 31 e 39 anos	Dicotômica
Idade de 40 a 49 anos	Usuários que possuem idade entre 40 e 49 anos	Dicotômica
Idade de 50 a 59 anos	Usuários que possuem idade entre 50 e 59 anos	Dicotômica
Idade maior que 60 anos	Usuários que possuem idade maior que 60 anos	Dicotômica
Acesso a carro e/ou moto	Usuários que possuem acesso regular a carro e/ou moto	Dicotômica
Acesso a bicicleta	Usuários que possuem acesso regular a bicicleta	Dicotômica
Vínculo - Graduação	Alunos de graduação	Dicotômica
Vínculo - Pós-graduação	Alunos de pós-graduação	Dicotômica
Vínculo - Funcionários	Funcionários técnico/administrativos	Dicotômica
Vínculo - Docentes	Docentes	Dicotômica
Área II - a pé	Principal modo de deslocamento a Área II é a pé	Dicotômica
Área II - bicicleta	Principal modo de deslocamento a Área II é a bicicleta	Dicotômica
Área II - ônibus interno	Principal modo de deslocamento a Área II é o ônibus interno	Dicotômica
Área II - motocicleta	Principal modo de deslocamento a Área II é a motocicleta	Dicotômica
Área II - carro	Principal modo de deslocamento a Área II é o carro	Dicotômica
Área II - carona	Principal modo de deslocamento a Área II é a carona	Dicotômica

A Tabela 3 e a Tabela 4, em seguida, caracterizam a ocorrência das categorias das variáveis nas duas amostras relativas à Área I e Área II.

Tabela 3 Frequência das variáveis para a Área I

VARIÁVEIS DICOTÔMICAS		FREQUÊNCIA	
		N	%
GÊNERO	MASCULINO	504	67%
	FEMININO	253	33%
IDADE	IDADE_<18	6	1%
	IDADE_18-24	321	42%
	IDADE_25-30	166	22%
	IDADE_31-39	86	11%
	IDADE_40-49	73	10%
	IDADE_50-59	79	10%
	IDADE_>60	26	3%
VÍNCULO	GRADUACAO	304	40%
	POS-GRADUACAO	214	28%
	FUNCIONARIOS	147	19%
	DOCENTE	78	10%
PRINCIPAL MODO DE DESLOCAMENTO ÁREA I	A PÉ	278	37%
	BICICLETA	40	5%
	ONIBUS MUNICIPAL	6	1%
	MOTOCICLETA	34	4%
	CARRO	375	50%
	CARONA	24	3%
ACESSO_CARRO-MOTO		520	69%
ACESSO_BICICLETA		171	23%
VARIÁVEIS QUANTITATIVAS		MÉDIA	DESVIO PADRÃO
TEMPO_CAMPUSI [min]		10,94	5,47
MINIMO_DISTANCIA [m]		1260,94	1259,29

Tabela 4 Frequência das variáveis para a Área II

VARIÁVEIS DICOTÔMICAS		FREQUÊNCIA	
		N	%
GÊNERO	MASCULINO	408	74%
	FEMININO	142	26%
IDADE	IDADE_<18	7	1%
	IDADE_18-24	278	51%
	IDADE_25-30	75	14%
	IDADE_31-39	61	11%
	IDADE_40-49	58	11%
	IDADE_50-59	55	10%
	IDADE_>60	16	3%
VÍNCULO	GRADUACAO	293	53%
	POS-GRADUACAO	80	15%
	FUNCIONARIOS	64	12%
	DOCENTE	106	19%
PRINCIPAL MODO DE DESLOCAMENTO ÁREA II	A PÉ	19	3%
	BICICLETA	13	2%
	ONIBUS INTERNO	184	33%
	MOTOCICLETA	16	3%
	CARRO	287	52%
	CARONA	26	5%
ACESSO_CARRO-MOTO		357	65%
ACESSO_BICICLETA		140	25%

3.3 Método

O método proposto para este trabalho é dividido em duas etapas: (1) Calibração e (2) Validação dos modelos estatísticos. Foram escolhidos os três principais modos de deslocamentos de cada área para aplicação do método. Para a Área I os modos analisados foram a pé, bicicleta e carro; e para Área II foram ônibus interno da universidade, bicicleta e carro. Após a definição das variáveis dependentes, iniciou-se a aplicação do modelo de regressão logística binomial, método *stepwise forward*, no IBM SPSS 22.0.

A calibração do modelo foi realizada com 70% dos dados, sorteados aleatoriamente, enquanto a validação foi realizada com os 30% restantes. Calculados os coeficientes da regressão, validou-se os modelos. No processo de validação, definiu-se como regra de classificação como sendo 0,5 a probabilidade de corte para atribuição do valor da variável dependente. No caso, se a probabilidade resultasse em um valor maior que 0,5, atribuía-se 1 para a variável dependente, se não, atribuía-se o valor 0. Após isso, calculou-se a porcentagem de acertos para cada modelo.

4 RESULTADOS E DISCUSSÕES

No IBM SPSS 22.0, foi aplicado o modelo de regressão logística binomial, método *stepwise forward*, para definir o modelo final que minimiza o número de variáveis e maximiza a precisão do modelo. Serão apresentados separadamente os resultados encontrados para cada uma das áreas.

4.1 Modelos para a Área I

Para a Área I foram gerados três modelos logísticos, utilizando a amostra de 70%, para previsão de escolha pelos usuários dos modos a pé, bicicleta e carro. Os valores de parâmetros calibrados para cada variável independente foram considerados significativos, considerando valores de estatística *Wald*. Além disso, a magnitude dos valores calibrados, bem como o sinal, em todos os casos, são coerentes. O poder preditivo dos modelos (Pseudo R²) mostrou-se alto. A Tabela 5 apresenta as variáveis incluídas, a porcentagem de acerto e valor do Pseudo R² para os modelos calibrados.

Tabela 5 Síntese dos resultados dos modelos logísticos para a Área I

ÁREA I					
Modelo	Variáveis incluídas no modelo	Coeficientes estimados	Pseudo R ²		Porcentagem global de acerto
			Cox & Snell	Nagelkerke	
1 Modo a pé	CONSTANTE	2,016	0,561	0,767	88,80%
	IDADE_31-39	-1,451			
	IDADE_41-49	-4,308			
	IDADE_50-59	-2,809			
	ACESSO_CARROMOTO	-2,898			
	ACESSO_BICICLETA	-1,147			
	TEMPO_AREA_I	0,268			
	MINIMA_DISTÂNCIA	-0,004			
2 Modo bicicleta	CONSTANTE	-20,443	0,205	0,562	94,80%
	ACESSO_CARROMOTO	-2,291			
	ACESSO_BICICLETA	20,023			
	MINIMA_DISTÂNCIA	0,0001			
3 Modo carro	CONSTANTE	-19,266	0,486	0,648	82,60%
	IDADE_<18	-20,304			
	IDADE_18-24	-1,709			
	IDADE_25-30	-1,331			
	ACESSO_CARROMOTO	21,454			
	TEMPO_AREA_I	-0,075			
	MINIMA_DISTÂNCIA	0,0001			

Com 30% da amostra restante, foi feita a validação dos três modelos calibrados anteriormente. Após a obtenção de probabilidades e categorias modais estimadas, foram calculadas as taxas de acertos para todos os modelos. A validação mostrou alta taxa de acertos para todos os modelos. A Tabela 6 indica os resultados de taxa de acertos para a validação.

Tabela 6 Taxa de acerto da validação dos modelos logísticos para a Área I

Modelo		Taxa de acerto
1	Modo a pé	90,10%
2	Modo bicicleta	96,40%
3	Modo carro	80,60%

A análise das variáveis presentes nos modelos para a Área I indicam que a idade é um fator determinante na escolha do modo de deslocamento. No caso do modo a pé, usuários com idade superior a 30 anos estão menos propensos a optar por esse meio de acesso ao *campus*. Para o modo carro, percebe-se que os usuários mais jovens, com idade de até 30 anos, são os que possuem menor tendência a utilização deste modo. Uma das características da Área I é que muitos de seus usuários, principalmente estudantes, moram próximos as entradas de acesso à universidade, o que influencia diretamente na escolha do modo de deslocamento. Outro aspecto importante observado é que a disponibilidade de bicicleta é o principal fator que influencia na sua adoção como forma de deslocamento. Além disso, é observado que o acesso a meios motorizados interfere de maneira direta na escolha do modo para se deslocar até a universidade, visto que este atributo está presente em todos os modelos.

4.2 Modelos para a Área II

Os modelos gerados para a Área II foram relativos à previsão de escolha modal do ônibus fornecido pela universidade que faz a ligação entre as áreas do *campus*, além do modo bicicleta e automóvel. No banco de dados é muito baixo o número de usuários que se deslocam a pé até esta área, pelo fato dela ser mais afastada da região central da cidade. Assim como para os modelos encontrados anteriormente, para a Área II todos os parâmetros calibrados para as variáveis independentes foram considerados significativos, bem como o sinal e magnitude dos valores se mostraram coerentes. Apesar da alta porcentagem global de acerto, o modelo para o modo bicicleta foi o único que apresentou baixos valores de Pseudo R². Os principais resultados dos modelos logísticos calibrados são apresentados na Tabela 7.

Tabela 7 Síntese dos resultados dos modelos logísticos para a Área II

ÁREA II					
Modelos	Variável	Coeficientes estimados	Pseudo R ²		Porcentagem global de acerto
			Cox & Snell	Nagelkerke	
1 Modo bicicleta	CONSTANTE	-5,488	0,069	0,284	96,90%
	IDADE_31-39	2,129			
	ACESSO_BICICLETA	2,673			
2 Modo ônibus interno	CONSTANTE	-0,914	0,467	0,648	89,00%
	ACESSO_CARROMOTO	-3,231			
	GRADUACAO	2,559			
	PÓS-GRADUACAO	1,705			
3 Modo carro	CONSTANTE	-4,633	0,525	0,700	86,40%
	ACESSO_CARROMOTO	6,074			
	ACESSO_BICICLETA	-1,121			
	DOCENTE	1,161			

Assim como no caso anterior, a validação mostrou alta taxa de acertos para todos os modelos (Tabela 8).

Tabela 8 Taxa de acerto da validação dos modelos logísticos para a Área II

Modelo		Taxa de acerto
1	Modo bicicleta	99,40%
2	Modo ônibus interno	88,10%
3	Modo carro	86,90%

O *campus* da USP em São Carlos possui como características deslocamentos entre as suas áreas. Esse aspecto é um dos principais fatores que influenciam no comportamento dos usuários pela escolha da melhor forma para acesso a Área II, visto que ela fica mais afastada da área central da cidade. No caso do modo carro, a disponibilidade do veículo é o atributo que mais influencia na escolha por esse modo. Observa-se também que usuários docentes estão mais propensos em utilizar o carro para acessar a Área II. O modelo para o modo bicicleta indica que os usuários com idade entre 30 e 39 anos e o acesso a bicicleta são as variáveis que mais interferem para optar por esse modo. O ônibus interno disponibilizado pela instituição tem predomínio de uso por estudantes de graduação e pós-graduação.

5 CONCLUSÕES

Este trabalho apresentou um método envolvendo regressão logística binomial para escolha por modo de transporte para um *campus* universitário. A partir de variáveis socioeconômicas e que caracterizam os modos de viagem, foram calibrados diferentes modelos para previsão de escolha modal. Os modelos calibrados tiveram bom poder preditivo, com alta taxa de acertos, além disso, os parâmetros obtidos foram significativos e coerentes. Observa-se que a escolha modal está fortemente relacionada com a posse (ou disponibilidade para uso) de veículos automotores, já que esta variável estava presente em grande parte dos modelos.

Na Área I percebe-se que o acesso à bicicleta tem forte correlação com a probabilidade de o usuário optar por este meio de transporte, pois esta variável apresenta alto valor de coeficiente associado. Também foi possível perceber que usuários mais jovens têm maior predisposição na escolha pelo modo a pé como forma de acesso ao campus.

Nos modelos da Área II, nota-se que os alunos de graduação e pós-graduação estão mais propensos a utilizar o ônibus interno da universidade para acessar esta área. O modelo de escolha do carro para acessar a esta área é influenciado pela disponibilidade a veículos motorizados e se usuários são docentes. O baixo valor de Pseudo R^2 para modelo bicicleta na Área II pode ser explicado pelo fato de haver um baixo número de usuários para esta área, que fazem uso deste modo de transporte.

Conclui-se que os modelos de regressão logística são ferramentas adequadas para estudos de previsão da escolha modal para *campus* universitários, visto que os modelos calibrados tiveram bom poder preditivo, com alta taxa de acertos e os parâmetros obtidos foram significativos e coerentes. Além disso, foi possível fazer uma análise dos principais fatores ou variáveis que influenciam na escolha modal para acesso ao *campus*. Conhecer as variáveis que interferem nas escolhas de acesso ao *campus* pode constituir uma informação valiosa para implementação de políticas públicas, sobretudo visando deslocamentos sustentáveis.

AGRADECIMENTOS

A CAPES pelo suporte financeiro fornecido e à Universidade de São Paulo pelo apoio para a realização da presente pesquisa.

6 REFERÊNCIAS

Hair Jr., J.F.; Anderson, R. E.; Tatham, R.L.; Black, W.C. (2009). **Análise multivariada de dados**. 6ª edição. Porto Alegre: Bookman.

IBGE (2010). **Censo Demográfico 2010**. Instituto Brasileiro de Geografia e Estatística. Acesso em 20 abril 2016. Disponível em: < <http://www.ibge.gov.br>>.

Jacques, M. A. P.; Bertazzo, A.; Galarraga, J.; Herz, M. (2011). **Polos Geradores de Viagens Orientados à qualidade de vida e ambiental**. Caderno: Estabelecimentos de Ensino. 1ª ed. Rio de Janeiro: Interciência.

Oliveira, A. M. (2015). **Um índice para o planejamento de mobilidade com foco em grandes Polos Geradores de Viagens - Desenvolvimento e aplicação em um *campus* universitário**. Dissertação de Mestrado, Universidade de São Paulo, São Carlos.

PNUD (2003). **Atlas de Desenvolvimento Humano no Brasil**. Programa das Nações Unidas para o Desenvolvimento. Acesso em 20 abril 2016. Disponível em: <http://www.pnud.org.br/atlas/ranking/IDH_Municipios_Brasil_2000.aspx?indiceAccordion=1&li=li_Ranking2003>.

Stein, P. P. (2013). **Barreiras, Motivações e Estratégias para Mobilidade Sustentável no *Campus* São Carlos da USP**. Dissertação de Mestrado, Universidade de São Paulo, São Carlos.

USP (2016). **Sobre o *Campus* da USP em São Carlos**. Acesso em: 20 abril 2016. Disponível em: <http://www.saocarlos.usp.br/index.php?option=com_content&task=view&id=61&Itemid=87>.