Análise de regressão linear: teoria e aplicações

Maria Clara Rizental Raicoski (PIBIC/Fundação Araucária/UEPG),
Airton Kist (Orientador), e-mail: kist@uepg.br

Universidade Estadual de Ponta Grossa/Departamento de Matemática e Estatística

Probabilidade e Estatística, **Estatística**

Palavras-chave: Correlação, Regressão linear simples, Regressão linear múltipla.

Resumo

A proposta deste trabalho é estudar a regressão simples e múltipla que objetivam a construção de modelos que descrevem de maneira razoável relações entre uma ou várias variáveis explicativas de um determinado processo e uma variável dependente. Foram feitos estudos teóricos desses métodos de regressão e replicações de aplicações.

Introdução

A pesquisa estatística se volta para solucionar problemas envolvendo variados aspectos da inferência, cada uma tendo a sua aplicação a situações específicas. Estimulada pelos seus campos de aplicação, ao lado das facilidades de processamento introduzidas pela informática, a Estatística tem enfatizado, o desenvolvimento dos procedimentos bi e multivariados.

A regressão linear simples estuda a relação entre uma variável resposta (Y) e uma variável explicativa ou regressora (X). Ao serem incorporadas várias variáveis independentes, o modelo denomina-se de Regressão Linear Múltipla. Assim como na Regressão Simples, a relação entre as variáveis é representada por um modelo matemático, que associa a variável dependente com as variáveis independentes.

O exemplo utilizado neste trabalho foi o Freitas e Ribeiro (2012) que diz respeito ao transporte rodoviário. Um dos procedimentos mais empregados para avaliar e classificar a qualidade dos serviços de transporte é o uso da regressão linear múltipla para identificar a importância relativa de alguns critérios. Esses critérios levam em consideração que cada viagem é um serviço específico e varia de acordo com as condições climáticas, condições de tráfego, condições do veículo, tipos de passageiros, entre outras. Os usuários nem sempre apresentam o mesmo perfil ou compartilham a mesma opinião, o que trás uma heterogeneidade de respostas ao estudo.

Para tanto, considerou-se para efeito de análise, os dados e resultados da abordagem metodológica proposta por Freitas et al. (2011) para avaliar a qualidade do transporte rodoviário intermunicipal de passageiros, realizado por ônibus. A abordagem em questão concentra-se na avaliação da qualidade do serviço de transporte rodoviário intermunicipal de passageiros realizado por ônibus, segundo a percepção dos usuários. Essa abordagem fundamenta-se na captação do grau de importância dos critérios e do grau de satisfação dos passageiros em relação aos critérios considerados.

Material e Métodos

A análise de regressão será realizada a partir de um estudo sobre qualidade do transporte rodoviário intermunicipal por ônibus previamente desenvolvido por Freitas e Ribeiro (2012). E todo o embasamento teórico utilizado no estudo foi baseado em Bussad e Morettin (2010) e MONTGOMERY et al. (2012).

O estudo foi realizado com 209 pessoas do município de Campos dos Goytacazes, localizado na região norte do Estado do Rio de Janeiro.

Foram identificados por Freitas e Ribeiro (2012), seis dimensões de qualidade e 17 critérios relevantes à avaliação da qualidade do transporte público por ônibus urbano, são eles:

- Atendimento: grau de cortesia dos funcionários (I_1), filas rápidas e organizadas (I_2), apresentação dos funcionários (I_3), e facilidade de compra (I_{11}),
- Veículo: estado de conservação dos veículos (I_4), existência de banheiro (I_5), existência de ar-condicionado (I_6), limpeza do veículo (I_7), e veículos acessíveis a portadores de necessidades especiais (I_8);
- Segurança: dispositivos de segurança adequados ao uso (I_9) e respeito às leis de trânsito (I_{10});

Deslocamento horário de saída conforme previsto (I_{12}), diversidade de horários de partida (I_{13}), tempo de viagem adequado ao proposto (I_{14}) e quantidade de paradas durante o percurso (I_{15});

- Serviços Diferenciais: existência de serviços diferenciais, tais como água, jornal e cobertor (I_{16}), e;
- Preço: relação custo-benefício, ou seja, o valor pago em relação ao serviço oferecido (I_{17}).

Um questionário foi desenvolvido com o intuito de captar o perfil dos respondentes e a percepção destes acerca do grau de importância dos critérios e do grau de satisfação em relação a tais itens. Para captar o grau de importância dos critérios e o grau de satisfação dos respondentes em relação a cada critério considerado foram utilizadas escalas não-comparativas contínuas de 0 a 10 pontos.

Resultados e discussões

A Tabela 1 apresenta o Grau de Satisfação Médio dos passageiros em relação a cada item (IQS_i). É possível notar que I_{11} (facilidades de compra) e I_{13} (diversidade de horários de partida) são os itens que promovem maior satisfação dos passageiros. Por outro lado I_2 (filas rápidas) e I_8 (veículos acessíveis a portadores de necessidades especiais) proporcionam menor satisfação dos passageiros.

De acordo com Bussad e Morettin (2010) a variável com o maior coeficiente de Pearson é a que melhor se relaciona com a variável dependente, que neste caso é o índice de qualidade de serviço. Na quarta coluna é apresentado o coeficiente de Pearson de cada uma das variáveis. Quando não controlado ou ajustado os efeitos das outras variáveis, o item I_{12} (horário de saída do ônibus conforme o previsto) possui o maior coeficiente de correlação, 0,437. O grau de cortesia dos funcionários (I_1) é a com o menor coeficiente de correlação de Pearson (0,111), e portanto, a variável que menos se relaciona com o índice de qualidade de serviço.

O coeficiente de correlação parcial tende a ser mais significativo que o coeficiente de Pearson, visto que as variáveis independentes também possuem correlações entre si, o que interfere nos resultados levando muitas vezes a

equívocos. O item I_2(filas rápidas e organizadas) é a variável que possui o maior coeficiente de correlação parcial em relação à variável dependente (0,737). Conclui-se que as filas rápidas possuem um alto valor de influência sobre o índice de qualidade de serviço (IQS). O segundo maior coeficiente corresponde ao item I_{10} (respeito às leis de trânsito). Por outro lado, o item I_{16} (existência de serviços diferenciais) possui menor coeficiente de correlação parcial com a variável dependente, cujo valor é 0,311.

Tabela 1 – Resultado das análises da qualidade do transporte rodoviário

Var.	Média (IQS_i)	Desvio Padrão	Corr. Pearson	Corr. Parcial	VIF	b_i	$b_{i(inf)}$	$b_{i(sup)}$	Erro Padrão	t_{cal}
IQS	6,219	0,570	---	---	---	---			---	---
Cons.	---	---	---	---	---	0,215	-0,039	0,468	0,129	1,667
I_1	6,049	1,584	0,111	0,544	1,207	0,062	0,049	0,076	0,007	8,971
I_2	3,227	1,955	0,338	0,737	1,208	0,085	0,074	0,096	0,006	15,065
I_3	7,565	1,864	0,232	0,372	1,103	0,031	0,020	0,042	0,006	5,537
I_4	7,259	1,419	0,404	0,601	1,267	0,083	0,067	0,099	0,008	10,398
I_5	7,662	1,647	0,304	0,593	1,187	0,068	0,054	0,081	0,007	10,171
I_6	6,779	2,219	0,335	0,539	1,219	0,044	0,034	0,054	0,005	8,834
I_7	7,751	1,272	0,422	0,371	1,369	0,051	0,033	0,069	0,009	5,529
I_8	3,497	1,541	0,270	0,604	1,155	0,073	0,059	0,087	0,007	10,460
I_9	7,447	1,697	0,221	0,490	1,259	0,051	0,038	0,065	0,007	7,760
I_{10}	4,407	1,812	0,360	0,725	1,156	0,086	0,075	0,098	0,006	14,552
I_{11}	8,554	1,468	0,397	0,324	1,383	0,038	0,022	0,054	0,008	4,738
I_{12}	5,112	1,830	0,437	0,672	1,267	0,077	0,065	0,089	0,006	12,515
I_{13}	8,216	1,366	0,287	0,368	1,227	0,047	0,031	0,063	0,008	5,786
I_{14}	6,812	1,476	0,289	0,472	1,158	0,054	0,040	0,069	0,007	7,392
I_{15}	6,136	2,368	0,140	0,627	1,194	0,052	0,042	0,061	0,005	11,121
I_{16}	6,277	2,872	0,180	0,311	1,270	0,018	0,010	0,026	0,004	4,515
I_{17}	5,512	2,230	0,368	0,714	1,158	0,068	0,059	0,078	0,005	14,102

R =	0,970
R^2 =	0,941
R^2 ajustado =	0,936
Erro padrão =	0,14472

	Soma dos Quadrados	g.l.	Quadrado Médio	F	Signif. de F
Regressão	63,562	17	3,739	178,523	0,0000
Resíduo	4,000	191	0,021		

Fonte: Ribeiro,A.;Freitas,A. (2012)

Com os dados da Tabela 1, Freitas (2012) concluiu que o modelo que melhor exprime a relação entre as variáveis e o índice de qualidade de serviço,

IQS = 0,062(I_1) + 0,085(I_2) + 0,031(I_3) + 0,083(I_4) + 0,068(I_5) + 0,044(I_6) + 0,051(I_7) + 0,073(I_8) + 0,051(I_9) + 0,086(I_{10}) + 0,038(I_{11}) + 0,077(I_{12}) + 0,047 (I_{13}) + 0,054(I_{14}) + 0,052(I_{15}) + 0,018(I_{16}) + 0,068(I_{17})

Os valores encontrados para o coeficiente de correlação múltipla, 0,970 e o coeficiente de determinação múltipla, 0,941 indicam de acordo com Bussad e Morettin (2010) que existe um alto poder de explicação do modelo de regressão e assim uma boa previsão da variável dependente (a qualidade do transporte intermunicipal de passageiros é bem estimada pela percepção da satisfação dos passageiros).

Foram realizados os testes t, de acordo com a seguinte hipótese: H_0: b_i = 0; 0 ≤ i ≤ k, quando i igual a zero equivale a constante da equação, e com i variando

entre 1 e k, são os coeficientes de regressão parcial. Considerando os valores contidos na Tabela 1, obteve-se: $t_{cal} = \frac{0,215}{0,129} = 1,667$. De acordo com a tabela de distribuição de t com n – k – 1 graus de liberdade e 5% de significância bicaudal, tem-se que t_{tab} é 1,96. Como t_{cal} é menor que t_{tab}, aceitou-se a hipótese nula (H_0). Desta forma, conclui-se que a constante encontrada não era significativa para o modelo. Aplicou-se o teste t aos dezessete itens, objetivando identificar os itens que influenciam na variação do índice de qualidade de serviço. Os valores de t_{cal} encontrados para cada item encontram-se na Tabela 1. O valor de t_{tab} será o mesmo encontrado para a constante e valerá para todas as variáveis. Analisando a Tabela 1 observa-se que para todos os itens os t_{cal} foram maiores que t_{tab}.

Por fim, obtém-se que a hipótese nula foi rejeitada para todos os itens, ou seja, tais itens são significativos e influenciam na variação do índice de qualidade de serviço.

Conclusões

A técnica de regressão linear simples e múltipla são muito utilizadas por cientistas de todas as áreas, como no exemplo do transporte rodoviário. Possibilita ainda a interpretação da importância relativa de cada variável ao modelo, além de relatar as associações entre as variáveis, levando a análises profundas sobre as correlações entre determinadas preditoras, essas relações são de grande interesse para as tomadas de decisões.

No trabalho realizado por Freitas e Ribeiro (2012) foi obtido um modelo de regressão linear múltipla que represente a qualidade de serviço no transporte público segundo a percepção dos passageiros em função dos aspectos presentes na prestação do serviço. Por meio desse estudo, foi possível identificar os itens mais importantes para aumento do índice de qualidade, a partir dos graus de satisfação dos usuários.

É importante ressaltar que os resultados obtidos se referem a um estudo realizado em um período específico, a partir da coleta da opinião dos passageiros de uma única empresa de ônibus, em um trecho específico atendido pela empresa. Por estas razões, os resultados não podem ser generalizados para outras situações. Mas nas condições apresentadas foi obtido um modelo satisfatório.

Agradecimentos

Agradeço a Fundação Araucária pelo incentivo à pesquisa e pelas bolsas concedidas dentro do Programa Institucional de Bolsas de Iniciação Científica. Um agradecimento especial ao meu orientador Prof. Dr. Airton Kist, pelos conhecimentos e orientações ao longo de um ano.

Referências

BUSSAD,W; MORETTIN,P.A.**Estatística Básica**:6 ed. São Paulo: Editora Saraiva,2010

FREITAS, A.L.P.; REIS FILHO, C. A. C.; RODRIGUES, F. R., **Avaliação da qualidade do transporte rodoviário intermunicipal de passageiros: uma abordagem exploratória**.v. 19, 49-61, 2011.

MONTGOMERY,D.C; PECK,E.A; VINING,G.G. **Introduction to Linear Regression Analysis**:5 ed.Canadá: John Wiley & Sons,Inc.,2012