

Análise Bayesiana do Sistema de Cotas da UFBA

Lilia Carolina C. da Costa
Universidade Federal da Bahia
Marina Silva Paez
Universidade Federal do Rio de Janeiro
Antonio Guimarães, Nadya Araujo Guimarães
Universidade de São Paulo

A partir do vestibular 2005, a UFBA implementou um sistema de reserva de vagas (cotas) para alunos originários da escola pública que se declaram pretos, pardos, índio-descendentes ou de outros grupos étnicos. Assim, desejando estudar quais características dos candidatos têm relação com o seu sucesso no vestibular, este estudo utiliza dados da UFBA, de 2005 até 2008, e verifica a relação existente entre a aprovação no vestibular e características dos candidatos, como renda familiar, cor/raça e nível educacional da mãe. Além disso, uma comparação é feita entre os resultados obtidos com os dados reais oriundos do sistema de cotas e com dados fictícios, supondo a inexistência de uma ação afirmativa.

Uma regressão logística bayesiana foi ajustada utilizando como variável resposta a aprovação ou não no vestibular (modelo 1). Além disso, o resultado do vestibular supondo a inexistência de um sistema de cotas no processo seletivo da UFBA foi obtido, atribuindo o *status* de aprovado aos alunos com maiores desempenhos no score final, independente de qualquer característica (raça ou rede de ensino do 2º. Grau), de acordo com o número de vagas de cada curso. Assim, outra regressão logística foi ajustada utilizando como variável resposta essa nova variável de aprovação sem o sistema de cotas (modelo 2).

As seguintes variáveis explicativas foram escolhidas segundo a literatura (Dachs e Maia, 2006):

1. Sexo: 1 = Masculino e 0=Feminino;
2. Tipo de curso no 2º. Grau, com três categorias (colegial, técnico, outros):
 - a. Curso1: 1 = colegial e 0 = técnico ou outros;
 - b. Curso2: 1 = técnico e 0 = colegial ou outros;
3. Fez cursinho: 1 = Não e 0 = Sim;
4. Participação econômica: 1 = Não trabalha e 0 = Trabalha;
5. Renda familiar: 1 = até 3 SM (salários mínimos), 2 = entre 3 e 5 SM, 3 = entre 5 e 10 SM, 4 = entre 10 e 20 SM e 5 = mais que 20 SM;
6. Nível de instrução da mãe: 1= até primário completo, 2 = fundamental incomp/completo, 3 = médio incomp/completo e 4 = superior incomp/completo;
7. Cor/Raça e Rede de ensino, com quatro categorias (branca e rede privada, branca e rede pública, negra e rede privada e negra e rede pública):
 - a. Raça1: 1 = branca e rede privada e 0 = caso contrário;
 - b. Raça2: 1 = branca e rede pública e 0 = caso contrário;
 - c. Raça3: 1 = negra e rede privada e 0 = caso contrário;
8. Concorrência dos cursos, com três categorias (alta, média e baixa):
 - a. Concor1: 1 = alta e 0 = caso contrário;
 - b. Concor2: 1 = baixa e 0 = caso contrário;

9. Local do Curso, com três categorias (UFBA-Capital, UFBA-Interior e UFRB – Universidade Federal do Recôncavo Baiano):
- Local1: 1 = UFBA-capital e 0 = caso contrário;
 - Local2: 1 = UFBA-interior e 0 = caso contrário;

Sob o enfoque bayesiano, é necessário ainda completar o modelo com a especificação das distribuições *a priori* dos parâmetros desconhecidos. As *prioris* para os coeficientes de regressão foram especificadas como tendo distribuição normal com média igual a sua estimativa por máxima verossimilhança e variância grande, de forma a serem pouco informativas.

Para a obtenção de amostras das distribuições *a posteriori*, utilizamos o método computacional Monte Carlo em Cadeias de Markov (MCMC) com o algoritmo de *Metropolis-Hastings* para atualizar cada parâmetro do modelo separadamente (ver Gamerman e Lopes, 2006) para uma revisão detalhada dos métodos MCMC). A proposta para o *Metropolis-Hastings* de cada parâmetro foi a de uma distribuição normal centrada no ponto anterior da cadeia, com variância escolhida de forma que a aceitação ficasse entre 40% e 60%. A implementação foi feita utilizando o pacote MCMCPack no software R (R Development Core Team, 2005). Para cada modelo foram feitas 6000 iterações da cadeia, sendo as 1000 primeiras descartadas. A amostra foi tomada a cada 10 iterações para diminuir a autocorrelação dentro das cadeias. As estimativas foram encontradas mediante uma amostra de tamanho 500. A convergência foi rápida e verificada visualmente.

A tabela 1 mostra as estimativas dos parâmetros do modelo 1, que analisa a probabilidade de sucesso no vestibular, e a tabela 2 apresenta os resultados do modelo 2, que estuda a probabilidade de sucesso no vestibular se não houvesse o sistema de cotas. Todas as variáveis foram significativas ao nível de 5%, com exceção de Raça 2 no modelo 2, significando que a probabilidade de aprovação no vestibular seria a mesma para brancos e negros de escola pública, se não houvesse o sistema de cotas. Os resultados encontrados foram coerentes com o esperado, como por exemplo, quanto maior a renda familiar e o nível educacional da mãe, maior a chance de aprovação no vestibular (média *a posteriori* de 0,18 e 0,17, no modelo1, respectivamente).

Tabela 1 – Estatísticas da distribuição *a posteriori* dos parâmetros do modelo 1.

Covariáveis	Modelo1						
	Média	Desvio Padrão	Quantis				
			2,5%	25%	50%	75%	97,5%
Intercepto	-3,03	0,06	-3,14	-3,07	-3,03	-2,99	-2,92
Sexo	0,25	0,01	0,23	0,24	0,25	0,26	0,28
Curso1	0,44	0,03	0,40	0,42	0,44	0,46	0,49
Curso2	0,50	0,03	0,43	0,48	0,50	0,52	0,56
Fez cursinho	0,25	0,02	0,22	0,24	0,25	0,26	0,28
Partic. Econômica	-0,23	0,02	-0,28	-0,25	-0,23	-0,22	-0,19
Renda familiar	0,18	0,01	0,16	0,17	0,18	0,18	0,19
Educação da mãe	0,17	0,01	0,16	0,17	0,17	0,18	0,19
Concor1	-1,09	0,02	-1,13	-1,11	-1,09	-1,08	-1,06
Concor2	0,52	0,02	0,47	0,50	0,52	0,53	0,56

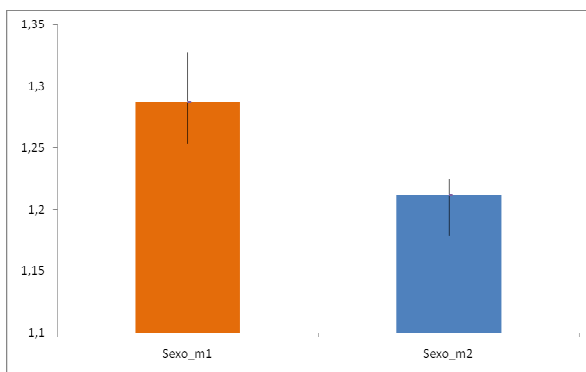
Raça1	-0,10	0,02	-0,15	-0,12	-0,10	-0,08	-0,06
Raça2	-0,15	0,03	-0,22	-0,17	-0,15	-0,13	-0,10
Raça3	-0,19	0,02	-0,22	-0,20	-0,19	-0,17	-0,15
Local1	0,60	0,05	0,52	0,57	0,60	0,64	0,70
Local2	0,60	0,04	0,53	0,58	0,60	0,63	0,67

Tabela 2 – Estatísticas da distribuição a posteriori dos parâmetros do modelo 2.

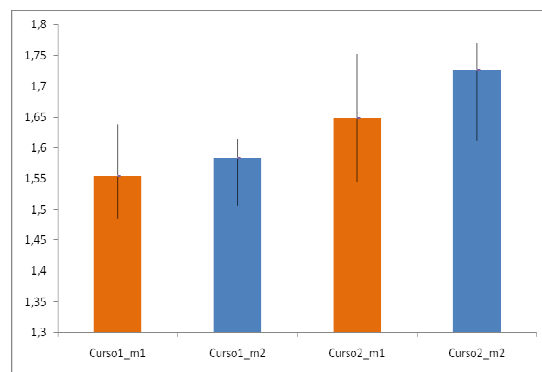
Covariáveis	Modelo2						
	Média	Desvio Padrão	Quantis da Posteriori				
			2,5%	25%	50%	75%	97,5%
Intercepto	-3,62	0,07	-3,76	-3,67	-3,61	-3,57	-3,50
Sexo	0,19	0,02	0,16	0,18	0,19	0,20	0,22
Curso1	0,46	0,03	0,41	0,44	0,46	0,48	0,51
Curso2	0,55	0,04	0,48	0,52	0,55	0,57	0,61
Fez cursinho	0,28	0,02	0,25	0,27	0,28	0,29	0,31
Partic. Econômica	-0,25	0,02	-0,30	-0,26	-0,25	-0,23	-0,20
Renda familiar	0,20	0,01	0,19	0,20	0,20	0,21	0,22
Educação da mãe	0,20	0,01	0,18	0,19	0,20	0,21	0,22
Concor1	-1,27	0,02	-1,30	-1,28	-1,27	-1,26	-1,23
Concor2	0,56	0,02	0,51	0,54	0,56	0,58	0,60
Raça1	0,65	0,03	0,61	0,64	0,66	0,67	0,70
Raça2	0,05	0,03	-0,02	0,03	0,06	0,08	0,12
Raça3	0,64	0,02	0,60	0,62	0,64	0,65	0,68
Local1	0,89	0,05	0,80	0,85	0,89	0,92	0,98
Local2	0,74	0,04	0,67	0,72	0,74	0,77	0,81

Figura 1 – Média a posteriori e Intervalo de Credibilidade de 95% da Odds Ratio das variáveis do modelo 1 e do modelo 2.

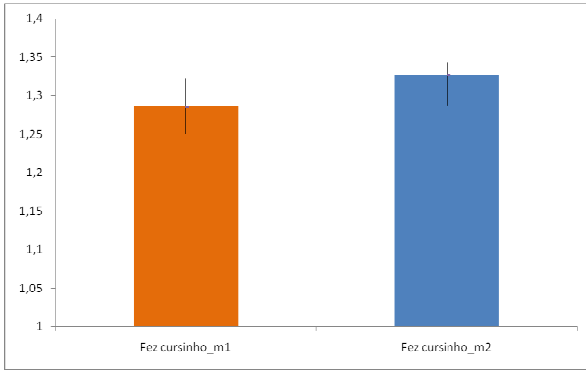
■ Modelo 1 ■ Modelo 2



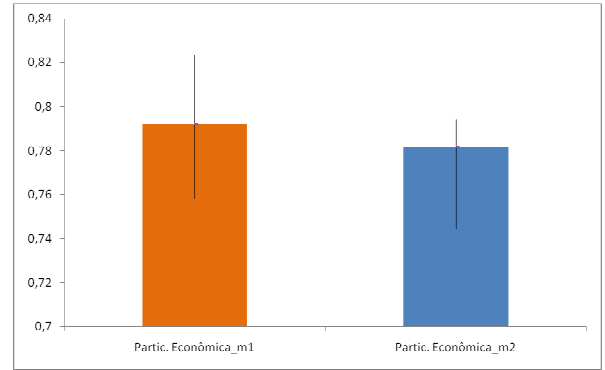
a) Sexo (masculino)



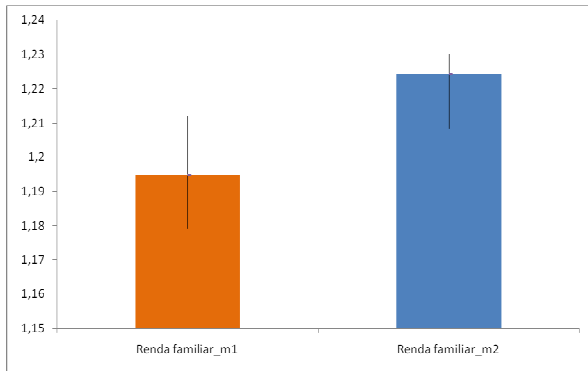
b) Curso1 (colegial) e curso2 (técnico)



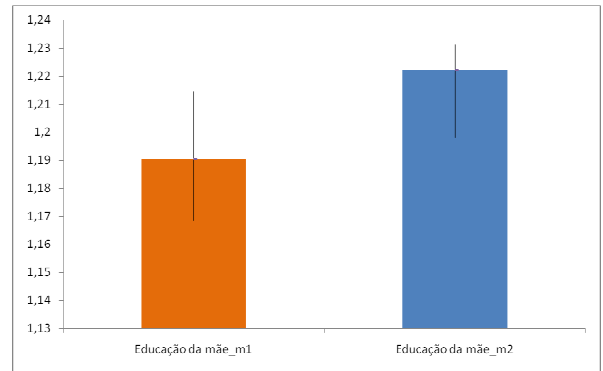
c) Fez cursinho (não)



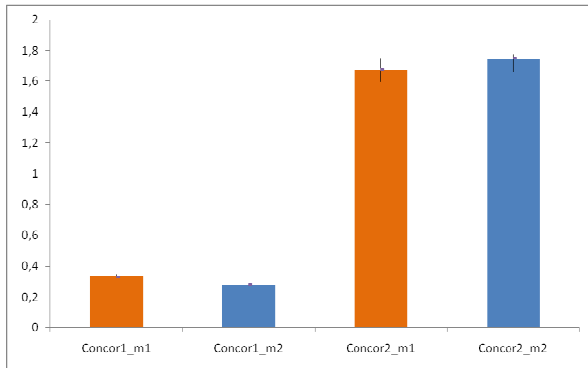
d) Participação econômica (não trabalha)



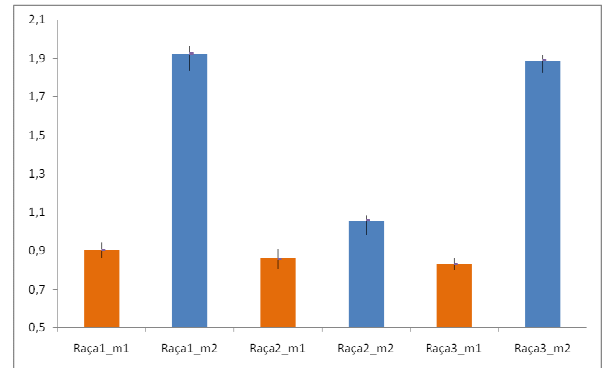
e) Renda familiar



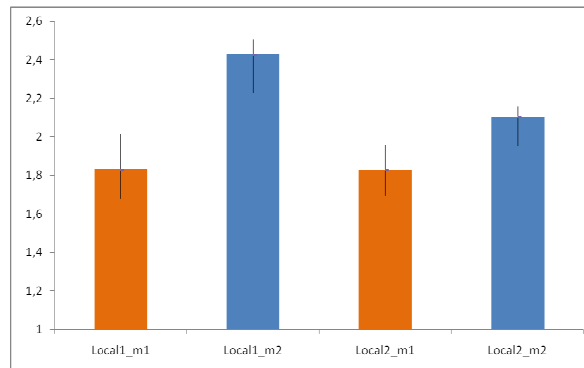
f) Educação da mãe



g) Concor1 (alta) e concor2 (baixa)



h) raça1(branca e privada), raça2(branca e pública) e raça3 (negra e privada)



i) Local1 (UFBA-capital) e local2 (UFBA-interior)

A figura 1 apresenta a média a posteriori da exponencial dos parâmetros dos modelos, isto é a *odds ratio*, e o seu intervalo de credibilidade de 95%. Pode-se perceber que as chances de aprovação no vestibular é maior para os homens, quem estudou em escola técnica no segundo grau, não fez cursinho, trabalha, maior renda família, maior nível educacional da mãe e prestou vestibular para cursos de baixa concorrência e que pertencem à UFBA-interior. Para a variável que tem relação direta com o sistema de cotas, raça1, raça2 e raça3, observa-se que as chances de aprovação dos candidatos oriundos de escolas privadas seriam de aproximadamente 90% a mais que os de escolas públicas, independente da cor/raça, se não houvesse o sistema de cotas. Entretanto, com o vestibular atual, o negro de escola pública tem maior chance de aprovação no vestibular. De maneira geral, a *odds ratio* das variáveis está mais próxima de 1 no primeiro modelo, indicando que o sistema de cotas diminui a diferença entre as chances de sucesso no vestibular relacionadas às características dos candidatos.

Referências Bibliográficas

- DACHS, J.N.W. & MAIA R.P. (2006). “Subsídios quantitativos para repensar as políticas de acesso à Universidade: Aumentando a equidade racial e econômica no ensino de terceiro grau no Brasil e no Estado de São Paulo, Parte 1” (Quantitative Subsidies for Re-thinking the Policies of Access in Higher Education: Increasing Racial and Economic Equity in Tertiary Education in Brazil and in the State of São Paulo), Cadernos de Pesquisa N. 74, NEPP, Unicamp. Disponível em: www.nepp.unicamp.br/Cadernos/Caderno74.pdf. Acesso em: 28 de abril de 2009.
- GAMERMAN, D. & LOPES, H.F. (2006). “Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference”. Segunda Edição. Chapman & Hall, London.
- R Development Core Team. (2005). “R: A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienna, Austria (2005). Disponível em <http://www.R-project.org>, ISBN 3-900051-07-0.