

5 Análise multivariada

5.1 Especificação dos modelos de regressão

Neste capítulo serão analisados, a partir das variáveis selecionadas, os fatores relacionados com as chances de cursar o ensino superior em determinadas instituições e áreas de formação, considerando as diferentes categorias das variáveis independentes, indicadoras da origem social, das características individuais e da trajetória escolar dos estudantes, por meio de modelos logísticos. Os modelos têm como objetivo permitir a observação dos efeitos de cada fator controlado pela presença dos demais, conforme as características destes modelos de regressão.

As regressões logísticas são modelos estatísticos de caráter preditivo que relacionam as chances de ocorrência de um determinado evento (variável dependente) com um conjunto de outras variáveis (independentes ou explicativas). Nos modelos de regressão os efeitos das variáveis são considerados independentes, ou seja, é possível estimar os efeitos de uma determinada variável controlando-se os efeitos das outras variáveis presentes no modelo. De maneira geral, devido à impossibilidade de serem arrolados todos os fatores condicionantes, os modelos construídos para as análises não cumprem, em sentido estrito, uma função preditiva; entretanto, os coeficientes calculados ainda assim indicam os efeitos de cada uma das variáveis dependentes sobre as probabilidades de ocorrência da variável independente, considerando o efeito conjunto das demais variáveis.

No caso da regressão logística, o que importa conhecer é se a classificação dos casos em uma ou em outra categoria da variável dependente pode ser prevista pelas variáveis independentes. No caso de uma variável dicotômica, composta por duas categorias, o problema consiste em prever as chances de um caso ser classificado em uma em comparação com a outra das duas categorias da variável, ou seja, a razão entre a probabilidade de um caso ser classificado em uma das categorias e a probabilidade de ser classificado na outra categoria.

A análise será desenvolvida a partir de dois modelos binomiais, ou seja, em que a variável dependente é dicotômica, o primeiro comparando as chances de um ingressante cursar o ensino superior em instituições seletivas e não seletivas, considerando todas as áreas, e o segundo comparando as chances de cursar o ensino superior em áreas de formação de maior e menor prestígio, independentemente da seletividade das instituições de ensino. O objetivo a ser alcançado é permitir a observação dos efeitos mais globais da seletividade das instituições e do prestígio das áreas de formação.

Posteriormente serão desenvolvidos dois modelos multinomiais, onde a seletividade das instituições e o prestígio das áreas são categorias da variável dependente. Nos modelos logísticos multinomiais a variável dependente tem mais de duas categorias e são calculadas as chances de um caso ser classificado em uma das categorias da variável dependente em comparação com uma das outras categorias, denominada categoria de referência. No primeiro caso será desenvolvido um modelo onde as categorias da variável independente são as combinações entre instituições seletivas e não seletivas e áreas de formação de alto e baixo prestígio. Tomando a frequência dos estudantes matriculados em áreas de formação de prestígio em instituições seletivas como referência, serão analisadas chances de um estudante cursar o ensino superior em áreas de baixo prestígio em instituições seletivas, em áreas de alto prestígio em instituições não seletivas e em áreas de baixo prestígio em instituições não seletivas. No segundo caso serão desenvolvidos modelos para cada uma das áreas de formação, e as categorias da variável dependente são: estar cursando o ensino superior na área específica em uma instituição seletiva (referência), estar cursando a área em uma instituição não seletiva e, finalmente, estar cursando uma outra área de formação.

O Quadro 1 descreve as variáveis dependentes utilizadas em cada um dos modelos, e as variáveis independentes, que são as mesmas para todos os modelos. As chances associadas às categorias das variáveis independentes são sempre estimadas em comparação com uma das categorias, denominada de categoria de referência.

Quadro 1
Variáveis dependentes dos modelos de regressão

Variável	Tipo	Nome	Descrição e categorias
Variáveis dependentes			
Modelo 1			
Indicador de instituição não seletiva	Dicotômica	ies_nsel	Indica se a IES é não seletiva: - instituição não seletiva - instituição seletiva (REF)
Modelo 2			
Indicador de área de formação de baixo prestígio	Dicotômica	area_bp	Indica se a área de formação é de baixo prestígio: - área de baixo prestígio - área de alto prestígio (REF)
Modelo 3			
Indicador de combinações entre instituições não seletivas e áreas de formação de baixo prestígio	Categórica	ies_area	Indica uma das quatro condições: - área de baixo prestígio em instituição não seletiva - área de baixo prestígio em instituição seletiva - área de alto prestígio em instituição não seletiva - área de alto prestígio em instituição seletiva (REF)
Modelo 4			
Indicador de instituição não seletiva	Categórica	ies_nsel	Indica uma das três condições, para cada área de formação: - área específica em instituição não seletiva - outra área - área específica em instituição seletiva (REF)
Variáveis independentes			
Gênero	Dicotômica	fem	Feminino Masculino (REF)
Faixa de renda familiar	Categórica	renda_1 renda_2	Até três salários mínimos Entre 3 e 10 salários mínimos Mais de 10 salários mínimos (REF)
Escolaridade do pai	Categórica	esc_1 esc_2	Nível fundamental Nível médio Nível superior (REF)
Cor declarada	Categórica	cor_1 cor_2 cor_3 cor_4	Negro(a) Pardo(a)/Mulato(a) Amarelo(a) (ou de origem oriental) Indígena ou de origem indígena Branco (REF)
Rede em que cursou o ensino médio	Categórica	rede_1 rede_2	Todo em escola pública Parte em escola pública e parte em particular) Todo em escola particular (REF)
Conhecimento de inglês	Categórica	ing_1 ing_2	Praticamente nulo Leio mas não escrevo Falo, leio e escrevo bem ou razoavelmente (REF)
Faixa de idade		idade_1 idade_2 idade_3	Mais de 28 anos Entre 25 e 28 anos Entre 20 e 22 anos Até 19 anos (REF)
Nota bruta	Contínua	nt_fg	Nota bruta do estudante na prova de formação geral

É comum aos modelos de regressão expressar os efeitos das variáveis independentes por meio de coeficientes associados a cada uma das variáveis, que ponderam o cálculo da variável dependente. Os modelos logísticos seguem o mesmo padrão. O valor ponderado, neste caso, está associado à chance de um caso ser classificado em uma das categorias da variável dicotômica. Os modelos logísticos utilizam a transformação logística, de forma que a média ponderada está associada ao logaritmo natural das chances. Assim os Modelos 1 e 2 foram especificados da seguinte forma:

$$\ln \frac{p_i}{1-p_i} = \beta_0 + \beta_1(fem)_i + \beta_2(renda_1)_i + \beta_3(renda_2)_i + \beta_4(esc_1)_i + \beta_5(esc_2)_i + \beta_6(cor_1)_i + \beta_7(cor_2)_i + \beta_8(cor_3)_i + \beta_9(cor_4)_i + \beta_{10}(rede_1)_i + \beta_{11}(rede_2)_i + \beta_{12}(ing_1)_i + \beta_{13}(ing_2)_i + \beta_{14}(idade_1)_i + \beta_{15}(idade_2)_i + \beta_{16}(idade_3)_i + \beta_{17}(nt_fg)_i$$

onde $p_i / (1 - p_i)$ é a chance do estudante i frequentar uma instituição seletiva, no Modelo 1, e a chance do estudante i frequentar uma área de baixo prestígio, no Modelo 2.

Nos modelos multinomiais são calculadas as chances de um caso ser classificado em uma das categorias em relação à categoria de referência da variável categórica. Os Modelos 3 e 4 foram especificados como na equação abaixo:

$$\ln \frac{p_{ij}}{p_{im}} = \gamma_{0j} + \gamma_{1j}(fem)_i + \gamma_{2j}(renda_1)_i + \gamma_{3j}(renda_2)_i + \gamma_{4j}(esc_1)_i + \gamma_{5j}(esc_2)_i + \gamma_{6j}(cor_1)_i + \gamma_{7j}(cor_2)_i + \gamma_{8j}(cor_3)_i + \gamma_{9j}(cor_4)_i + \gamma_{10j}(rede_1)_i + \gamma_{11j}(rede_2)_i + \gamma_{12j}(ing_1)_i + \gamma_{13j}(ing_2)_i + \gamma_{14j}(id_1)_i + \gamma_{15j}(id_2)_i + \gamma_{16j}(id_3)_i + \gamma_{17j}(nt_fg)_i$$

onde, p_{ij} / p_{im} é a chance do estudante i ser classificado na categoria j da variável em relação à categoria de referência m . Para o Modelo 3, a categoria de referência é a frequência a uma área de alto prestígio em instituição seletiva, e $m = 1$ corresponde à frequência a uma área de alto prestígio em instituição não seletiva, $m = 2$ corresponde

à frequência a uma área de baixo prestígio em instituição seletiva e $m = 3$ corresponde à frequência a uma área de baixo prestígio em instituição não seletiva. Para o Modelo 4, a categoria de referência é a frequência a uma instituição seletiva na área em questão, e $m = 1$ corresponde à frequência a uma determinada área em instituição não seletiva e $m = 2$ corresponde à frequência a uma outra área.

Para o ajuste dos modelos de regressão às ocorrências de cada caso na estimação de seus coeficientes foram utilizadas a análise da *deviance*, medida da verossimilhança entre os resultados do modelo e o conjunto dos dados observados, e a estatística de *Wald*, para a condução dos testes de hipótese para cada uma das variáveis do modelo. A *deviance* é definida de tal forma que quanto menor seu valor maior é a verossimilhança. O método utilizado para a especificação do modelo de regressão é a comparação da *deviance* de modelos que envolvem diferentes combinações de variáveis. Neste trabalho o procedimento adotado foi a inclusão sucessiva de cada uma das variáveis e a avaliação das variações correspondentes da *deviance*. Posteriormente, para efeito da confirmação da relevância do conjunto das variáveis incluídas para o poder explicativo do modelo, é utilizado o parâmetro *Nagelkerke R Square*, que consiste em uma aproximação da variância explicada pelos modelos logísticos.

Na construção dos modelos de regressão, as variáveis relacionadas no modelo para verificação das chances de estar cursando o ensino superior em instituições e áreas de menor prestígio, considerados como eventos indesejáveis, são denominadas de fatores de risco para a estratificação do ensino superior; valores maiores dos coeficientes indicam, neste caso, maiores probabilidades de estar cursando este nível de ensino em instituições e áreas de formação de menor prestígio.

A Tabela 29 descreve as informações sobre o número de parâmetros, o valor da *deviance* e a significância estatística de sua redução para o modelo da seletividade das instituições. O Modelo 1 é o modelo nulo, onde nenhuma das variáveis foi incluída e assume-se que a probabilidade de estar cursando o ensino superior em uma instituição não seletiva não varia em função dos fatores explorados. O modelo nulo é utilizado como base para a comparação dos modelos subsequentes. O Modelo 2

inclui a variável gênero; a *deviance* sofre então uma redução significativa, como demonstra o *p-valor*. No Modelo 3, quando é introduzida a variável faixa de renda familiar, a *deviance* sofre redução maior que a do passo anterior, e mesmo com a introdução de dois novos parâmetros a melhora do ajuste é significativa. A introdução da variável escolaridade do pai no Modelo 4 promove uma redução da *deviance* tão relevante quanto a da variável renda, mostrando que os efeitos da renda e da escolaridade do pai sobre a verossimilhança do modelo são semelhantes. Os efeitos da cor declarada são incluídos no Modelo 5, e aí é obtida uma pequena melhora no ajuste, bem menos relevante quanto às obtidas com as duas variáveis anteriores, embora os resultados sejam significativos. No Modelo 6 a inclusão da rede em que o estudante cursou o ensino médio é significativa, embora com uma redução da *deviance* menos importante que a obtida com a renda e a escolaridade do pai, sendo equivalente à inclusão do gênero. Com a inclusão do fator relativo ao domínio da língua inglesa no Modelo 7 observa-se a segunda menor redução da *deviance*. A redução com a introdução da variável idade, no Modelo 8, é mais relevante, equiparando-se à obtida com as variáveis gênero e rede.

Tabela 29
Ajustes do modelo de regressão do risco de frequentar uma IES não seletiva

	Modelo	Parâmetros	<i>Deviance</i>	Redução	p-valor	Nagelkerke R Square
1	Nulo	1	159.690			
2	G	2	158.443	1.247	0,000	0.011
3	G+R	4	155.809	2.634	0,000	0.034
4	G+R+E	6	153.162	2.647	0,000	0.057
5	G+R+E+C	10	152.913	249	0,000	0.060
6	G+R+E+C+RD	12	151.714	1.199	0,000	0.070
7	G+R+E+C+RD+L	14	150.982	732	0,000	0.076
8	G+R+E+C+RD+L+I	17	149.825	1.157	0,000	0.086
9	G+R+E+C+RD+L+I+P	19	142.729	7.096	0,000	0,146

LEGENDA: G = Gênero; R = Faixa de renda familiar; E = Escolaridade do pai; C = Cor declarada; RD = Rede de ensino em que estudou no ensino médio; L = Domínio da língua inglesa; I = Faixa de idade; P = Nota na prova de formação geral.

A especificação do mesmo modelo para o risco de frequentar uma área de baixo prestígio é detalhada na tabela abaixo. O modelo final explica uma proporção bem maior da variância do que o modelo de risco em relação à seletividade das instituições, pois o *Nagelkerke R Square* atinge o valor de 0,453, em comparação com o valor de 0,146 do modelo anterior. Retornando à especificação em foco, a inclusão da variável gênero promove uma redução altamente relevante, de 52.209 pontos, na *deviance*. Este valor confirma a relevância do gênero para a escolha das áreas de formação.

A introdução da variável renda no Modelo 3 também resulta numa redução bastante grande da *deviance*, a segunda maior entre o conjunto das variáveis. No Modelo 4 o efeito da variável escolaridade do pai também promove uma redução da *deviance* importante, embora menor do que as das duas primeiras variáveis. Os efeitos da cor declarada, da rede de ensino em que o estudante cursou o ensino médio e do conhecimento de língua inglesa, nos Modelos 5, 6 e 7 provocam reduções da *deviance* em valores de dimensões equivalentes. A introdução da variável idade, no Modelo 8, sendo mais próxima da obtida com a introdução da variável escolaridade do pai, embora um pouco menor. No modelo final a inclusão da nota na prova de formação geral promove também uma pequena redução da *deviance*.

Tabela 30
Ajustes do modelo de regressão do risco de frequentar uma área de baixo prestígio

	Modelo	Parâmetros	<i>Deviance</i>	Redução	p-valor	Nagelkerke R Square
1	Nulo	1	265.998			
2	G	2	213.789	52.209	0,000	0,306
3	G+R	4	196.284	17.505	0,000	0,393
4	G+R+E	6	191.725	4.559	0,000	0,414
5	G+R+E+C	10	190.651	1.074	0,000	0,419
6	G+R+E+C+RD	12	189.393	1.258	0,000	0,425
7	G+R+E+C+RD+L	14	188.341	1.052	0,000	0,430
8	G+R+E+C+RD+L+I	17	184.570	3.771	0,000	0,447
10	G+R+E+C+RD+L+I+P	19	183.220	1.350	0,000	0,453

LEGENDA: G = Gênero; R = Faixa de renda familiar; E = Escolaridade do pai; C = Cor declarada; RD = Rede de ensino em que estudou no ensino médio; L = Domínio da língua inglesa; I = Faixa de idade; P = Nota na prova de formação geral.

Comparando a explicação da variância aproximada dos dois modelos, o Modelo 1 possui um parâmetro *Nagelkerke R Square* igual a 0,146, enquanto o parâmetro do modelo 2 é de 0,453, uma indicação de que o Modelo 2 tem maior capacidade de explicar a variância dos resultados do que o Modelo 1. Isto indica que a influência das áreas de formação sobre a estratificação é maior do que a seletividade das instituições.

5.2.

Modelos de risco de frequência a instituições não seletivas e áreas de baixo prestígio

A Tabela 31 apresenta os resultados dos modelos binomiais para a seletividade das instituições, considerando primeiramente o risco de frequentar uma instituição não seletiva sem o controle dos resultados pelo desempenho e, em seguida, com este controle (representado pela introdução da variável nota bruta na prova de formação geral). A tabela apresenta os parâmetros exp^{β} , que variam de 0 a $+\infty$, sendo que suas distâncias da unidade representam o risco em relação à categoria de referência da variável em questão. Valores de exp^{β} maiores que a unidade são indicadores de maiores chances de ocupar a categoria em questão em relação à categoria de referência, e valores menores que a unidade indicam chances menores.

Todas as variáveis do modelo apresentam resultados significativos. Os resultados apontam que os estudantes do gênero feminino estão mais sujeitas ao risco frequentarem instituições não seletivas. Em relação à renda e à escolaridade dos pais, os resultados indicam que estudantes provenientes de famílias de baixa renda e com pais de menor escolaridade estariam mais propensos a cursarem o ensino superior em instituições não seletivas. Em relação à cor declarada, os resultados são não esperados: os estudantes que se declaram negros, pardos e indígenas estão protegidos em relação ao risco de frequentar instituições não seletivas, quando comparados com os estudantes brancos; os estudantes de origem oriental aparecem em situação de risco.

Tabela 31
Parâmetros estimados pelo modelo de risco de cursar o ensino superior em instituições não seletivas

Variável	Modelo sem controle de desempenho	Modelo com controle de desempenho
	<i>exp^β</i>	
Gênero		
Feminino	1,219***	1,232***
Masculino (REF)		
Faixa de renda mensal da família		
até 3 SM	1,437***	1,368***
de 3 a 10 SM	1,245***	1,244***
mais de 10 SM (REF)		
Como você se considera? (Cor Declarada)		
Branco(a) (REF)		
Negro(a)	0,739***	0,725***
Pardo(a) ou mulato(a)	0,759***	0,773***
Amarelo (Origem Oriental)	1,141*	1,088***
Indígena (Origem Indígena)	0,837**	0,821***
Escolaridade do pai		
Ensino fundamental	1,672***	1,650***
Ensino médio	1,261***	1,246***
Ensino superior (REF)		
Rede em que cursou o ensino médio		
Escola privada (particular) (REF)		
Escola pública	1,600**	1,597***
Parte em escola pública e parte em privada	1,497***	1,246***
Como é seu conhecimento de língua inglesa?		
Leio, escrevo e falo bem ou razoavelmente (REF)		
Leio, mas não falo	1,379***	1,293***
Praticamente nulo	1,466***	1,288***
Faixa de Idade		
Até 19 anos (REF)		
De 20 a 22 anos	0,847***	0,794***
De 23 a 28 anos	1,178***	1,106***
Mais de 28 anos	1,782***	1,752***
Nota bruta na prova de formação geral		0,964***
Constante		3,277***
		29,682***

* p < 0,05; ** p < 0,01; *** p < 0,001

A introdução do controle pela nota bruta na formação geral não altera as tendências observadas no primeiro modelo, embora a variável seja significativa – quanto maior é a nota bruta, menor é o risco do estudante frequentar uma instituição não seletiva.

Os estudantes da rede pública mostram-se mais sujeitos ao risco de cursar o ensino superior em instituições não seletivas, quando comparados com os estudantes das escolas privadas. Os estudantes que migraram entre as redes durante o ensino médio estão sujeitos a risco semelhante, embora um pouco mais reduzido.

O conhecimento de língua inglesa pode ser um indicador do capital cultural para a frequência às instituições seletivas. Em comparação com os estudantes que declaram o maior nível de proficiência na língua, todos os demais têm maiores riscos de cursar o ensino superior em instituições não seletivas. A idade, indicador da trajetória escolar do estudante, mostra-se como fator de risco para estudantes frequentarem instituições não seletivas, considerando os estudantes maiores de 22 anos; os estudantes da faixa de 20 a 20 anos, entretanto, estão protegidos em relação ao risco de cursarem o ensino superior em instituição não seletivas quando comparados aos estudantes de até 19 anos.

A Tabela 32 apresenta os resultados do modelo binomial para as chances de cursar o ensino superior em áreas de formação de baixo prestígio. Todas as variáveis apresentam efeitos relevantes, com exceção da faixa de idade entre 20 e 22 anos da variável idade, no primeiro modelo, e dos alunos que cursaram o ensino médio na rede pública. O coeficiente que indica o risco das mulheres frequentarem áreas de baixo prestígio é muito alto, como esperado; o mesmo é observado para os estudantes de renda familiar de até três salários mínimos, embora os estudantes de renda familiar entre três e dez salários mínimos também estejam em situação de risco. Estudantes não brancos, com exceção dos de origem oriental, e de pais menos escolarizados têm maiores riscos de cursar o ensino superior em áreas de menor prestígio – no caso, as licenciaturas.

Tabela 32
Parâmetros estimados pelo modelo de risco de cursar o ensino superior em áreas de baixo prestígio

Variável	Modelo sem controle de desempenho	Modelo com controle de desempenho
	<i>exp^β</i>	
Gênero		
Feminino	8,943***	9,095***
Masculino (REF)		
Faixa de renda mensal da família		
até 3 SM	5,115***	5,267***
de 3 a 10 SM	1,798***	1,815***
mais de 10 SM (REF)		
Como você se considera? (Cor Declarada)		
Branco(a) (REF)		
Negro(a)	1,397***	1,407***
Pardo(a) ou mulato(a)	1,375***	1,362***
Amarelo (Origem Oriental)	0,668***	0,68***
Indígena (Origem Indígena)	1,444***	1,453***
Escolaridade do pai		
Ensino fundamental	1,569***	1,587***
Ensino médio	1,161***	1,174***
Ensino superior (REF)		
Rede em que cursou o ensino médio		
Escola privada (particular) (REF)		
Escola pública	1,483***	1,499***
Parte em escola pública e parte em privada	1,379***	1,430***
Como é seu conhecimento de língua inglesa?		
Leio, escrevo e falo bem ou razoavelmente (REF)		
Leio, mas não falo	0,960**	0,984***
Praticamente nulo	1,376***	1,450
Faixa de Idade		
Até 19 anos (REF)		
De 20 a 22 anos	1,020	1,041**
De 23 a 28 anos	1,289***	1,322***
Mais de 28 anos	2,834***	2,830***
Nota bruta na prova de formação geral		
		1,013***
Constante	0,173***	0,083***

* p < 0,05; ** p < 0,01; *** p < 0,001

Os estudantes provenientes das escolas públicas também têm maiores riscos de cursar o ensino superior em áreas de menor prestígio, assim como os que cursaram parte do ensino médio em cada uma das redes. Os estudantes com menor domínio da língua inglesa e de maior idade também teriam maior risco de frequentar áreas de formação de menor prestígio.

A Tabela 33 apresenta os resultados do modelo multinomial, considerando as combinações possíveis entre a seletividade das IES e o prestígio das áreas de formação. Este modelo permite a observação da nuance entre os efeitos de seletividade das instituições entre as áreas. A categoria de referência da variável dependente é cursar uma área de alto prestígio em uma instituição seletiva. De maneira geral o modelo apresenta resultados esperados – estudantes de perfil social desfavorecido, com menor proficiência em inglês e os de maior idade estão mais propensos a cursar áreas de baixo prestígio em instituições menos seletivas, salvo algumas exceções.

A renda aparece como fator de risco mais importante em relação ao prestígio das áreas do que em relação à seletividade das instituições – o risco de estudantes de baixa renda frequentarem cursos de áreas de baixo prestígio, seja em instituições seletivas ou não seletivas, é bem maior do que o risco desses estudantes cursarem um curso de alto prestígio em uma instituição não seletiva, quando o desempenho não é considerado. Entretanto, quando o desempenho do estudante é inserido no modelo, o indicador dos estudantes de renda até três salários mínimos de cursos de áreas de alto prestígio não é significativo. Este resultado pode indicar que estudantes de baixa renda e de alto desempenho têm a possibilidade de acionar suas vantagens acadêmicas para acessar as instituições de maior prestígio, e pode ainda sugerir a existência de um processo de superseleção dos estudantes de baixa renda, onde somente aqueles de desempenho muito destacado teriam acesso às áreas de alto prestígio – vide Tabela 27, no cap. 4, onde verifica-se que somente 1,3% dos estudantes de renda até três salários mínimos frequentam as áreas tecnológicas.

Tabela 33
Parâmetros estimados pelo modelo multinomial de risco de cursar o ensino superior em áreas de formação de baixo prestígio e instituições não seletivas

Variáveis	Área de baixo prestígio em IES não seletiva		Área de baixo prestígio em IES seletiva		Área de alto prestígio em IES não seletiva	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2
Gênero						
Feminino	9,132***	9,118***	4,949***	5,197***	0,907***	0,902***
Masculino (REF)						
Faixa de renda mensal da família						
até 3 SM	6,324***	6,080***	3,473***	3,611***	1,137**	1,075
de 3 a 10 SM	2,405***	2,372***	1,749***	1,756***	1,294***	1,273***
mais de 10 SM (REF)						
Como você se considera?						
Branco(a) (REF)						
Negro(a)	1,084	1,060	1,462***	1,459***	0,795***	0,768***
Pardo(a) ou mulato(a)	1,115***	1,137***	1,496***	1,511***	0,838***	0,862***
Amarelo (Origem Oriental)	0,797**	0,779**	0,794*	0,801*	1,220**	1,174*
Indígena (Origem Indígena)	1,342*	1,300*	1,633***	1,608***	0,957	0,918
Escolaridade do pai						
Ensino fundamental	2,894***	2,851***	1,718***	1,736***	1,809***	1,782***
Ensino médio	1,558***	1,536***	1,255***	1,259***	1,337***	1,313***
Ensino superior (REF)						
Rede em que cursou o ensino médio						
Escola privada (particular) (REF)						
Escola pública	2,693***	2,703***	1,683***	1,722***	1,786***	1,800***
Esc. pública/privada	2,119***	2,000***	1,353***	1,383***	1,487***	1,372***
Como é seu conhecimento de língua inglesa?						
Leio, escrevo e falo bem/raz(REF)						
Leio, mas não falo	1,388***	1,315***	1,040	1,041	1,436***	1,333***
Praticamente nulo	2,269***	2,021***	1,583***	1,585***	1,644***	1,396***
Faixa de Idade						
Até 19 anos (REF)						
De 20 a 22 anos	0,710***	0,673***	0,691***	0,693***	0,649***	0,600***
De 23 a 28 anos	1,345***	1,278***	0,976	0,990	0,978	0,91**
Mais de 28 anos	6,103***	6,313***	2,937***	3,168***	2,015***	2,123***
Nota bruta na prova de form. geral						
		0,974***		1,008***		0,962***
Constante						
	0,616***	3,121***	0,343***	0,202***	4,284***	41,972***

* p < 0,05; ** p < 0,01; *** p < 0,001

Os estudantes cujos pais não têm nível superior estão expostos a riscos semelhantes em relação às três trajetórias comparadas, sendo que os que têm pais com nível de ensino fundamental sofrem risco superior aos que têm pais com ensino médio. O gênero aparece como fator preponderante em relação às áreas de baixo prestígio, seja em IES seletivas ou em não seletivas. Entretanto, as mulheres que optam por áreas de alto prestígio, onde são minoria, estão protegidas em relação ao risco de cursar o ensino superior em instituições não seletivas – ou seja, embora frequentem com maior frequência as áreas de baixo prestígio, as mulheres que optam por cursos de alto prestígio têm acesso preferencial às instituições seletivas. Estes resultados também podem indicar um processo de superseleção.

Assim como no caso das mulheres, os resultados indicam que os estudantes que se declaram pardos têm maiores riscos de cursarem o ensino superior em áreas de baixo prestígio, embora os efeitos sejam menores, e os que frequentam cursos de alto prestígio estão protegidos em relação ao risco de frequentarem instituições não seletivas. Entretanto, os riscos destes estudantes que optam por áreas de baixo prestígio associados à frequência às instituições seletivas são maiores que os riscos associados às instituições não seletivas, resultado que pode indicar que, para estes estudantes, a opção mais frequente seja por áreas de baixo prestígio em instituição seletivas.

Os resultados para os estudantes que se declaram negros merecem ser examinados em maior detalhe. Os resultados não indicam risco destes estudantes frequentarem áreas de baixo prestígio em instituições não seletivas - porém são verificados resultados significativos e positivos em relação ao risco de cursar áreas de baixo prestígio em instituições seletivas. Ou seja, para os estudantes negros a opção por cursos de áreas de baixo prestígio em instituição seletivas é mais frequente dos que para os estudantes que se declaram pardos.

Tanto os estudantes das escolas públicas quanto os que cursaram o ensino médio nas duas redes têm maiores riscos de cursar áreas de baixo prestígio do que os egressos das escolas privadas, seja em instituições seletivas ou não seletivas, e áreas

de alto prestígio em instituições não seletivas, sendo que os efeitos são mais fortes para os primeiros.

Os efeitos negativos do desconhecimento de língua inglesa sobre os riscos de cursar o ensino superior em áreas de formação de baixo prestígio e instituições não seletivas são significativos em todos os casos. Os estudantes da faixa de idade entre 20 e 22 anos têm maiores chances de frequentar cursos de alto prestígio em instituições seletivas do que os mais jovens. Os ingressantes de mais 28 anos têm grandes riscos de frequentarem áreas de formação de baixo prestígio em instituições não seletivas e áreas de baixo prestígio. De forma geral, os estudantes mais velhos estão em situação de risco para todas as categorias da variável dependente.

5.3. Modelos de risco de frequência a instituições não seletivas por áreas de formação

Considerando separadamente cada uma das áreas de formação os resultados comportam mais variações, conforme pode ser verificado na Tabela 34. Na área de arquitetura, os resultados indicam que as mulheres estão protegidas em relação ao risco de frequentarem instituições não seletivas. Este efeito é bastante dilatado – as mulheres têm cerca do dobro de chances de cursarem arquitetura em instituições seletivas do que os homens. Em relação à renda o efeito não é significativo, mas estudantes cujos pais têm escolaridade até o ensino fundamental têm chances de frequentar instituições seletivas muito restrito. Estudantes pardos e negros também estão em situação de risco de frequentar instituições não seletivas quando comparados aos brancos, nesta área. O mesmo ocorre com os egressos das escolas públicas, os que têm menor proficiência em língua inglesa e os maiores de 28 anos.

Nas engenharias as mulheres, os negros e os pardos estão protegidos em relação à frequência às instituições não seletivas. Os estudantes de famílias com faixa de renda entre 3 e 10 salários mínimos estão sujeitos a risco, embora os efeitos para a faixa de até 3 salários mínimos não sejam significativos. A baixa escolaridade do pai é fator de risco, assim como a frequência à escola pública no ensino médio, pequeno domínio da língua inglesa e a faixa de idade maior que 28 anos. Resultados semelhantes são encontrados na área de computação, com a exceção da proteção às mulheres, cujos efeitos não são significativos, e dos efeitos em relação aos estudantes da faixa de idade maior que 28 anos, que estão protegidos em relação ao risco de frequentarem instituições não seletivas na área.

Na área de física não são registrados efeitos de risco significativos em relação ao gênero, à cor declarada (exceto para os estudantes de origem indígena) e aos estudantes mais velhos, e somente os estudantes de renda familiar até 3 salários mínimos, que não cursaram o ensino médio todo em escolas privadas e com conhecimento de inglês praticamente nulo estão em situação de risco.

Assim como na área de arquitetura, a renda familiar não aparece como fator de risco para os estudantes de química, mas o efeito da escolaridade do pai é mais relevante que nas outras áreas. As mulheres, os estudantes pardos e negros também estão protegidos de risco. Os efeitos em relação à idade são nulos. Os egressos da rede pública e as duas categorias de menor proficiência em língua inglesa também aparecem como fatores de risco.

Os estudantes de renda familiar até 10 salários mínimos, de pais com escolaridade fundamental, que frequentaram no ensino médio a escola pública, totalmente ou em parte, e com pouca proficiência em inglês estão em situação de risco de frequentarem instituições não seletivas na área de biologia. Os estudantes pardos e negros estão protegidos, e os efeitos para as mulheres e para os estudantes com mais de 28 anos são nulos.

As áreas de ciências sociais e filosofia são as que apresentam menos fatores de risco para os estudantes menos privilegiados. Para a primeira somente estudantes egressos das escolas públicas e maiores de 28 anos estão em situação de risco – o modelo não indica riscos em relação à renda familiar, escolaridade do pai, cor declarada ou conhecimento de inglês. A filosofia aparece com situação semelhante, exceto pelos riscos dos estudantes com pais com nível de escolaridade fundamental e estudantes com conhecimentos somente de leitura em língua inglesa.

Para as áreas de geografia, história e matemática a situação é semelhante em relação aos indicadores de gênero, renda familiar, escolaridade do pai, (conhecimento de língua inglesa e faixa de idade acima de 23 anos), todos fatores de risco nestas áreas. Em relação à cor declarada os estudantes pardos estão protegidos em relação ao risco de frequentarem instituições não seletivas nas áreas de história e matemática, e os negros na área de matemática; na área de geografia não são registrados efeitos em relação à cor declarada, e em história os efeitos relativos à rede de ensino no nível médio são nulos, efeito único entre todas as áreas.

Finalmente, as mulheres estão em situação de risco de frequentarem instituições não seletivas nas áreas de pedagogia e letras, também uma situação única

entre as áreas. As duas áreas, entretanto, diferem em relação às variáveis indicadoras da origem social dos estudantes. Enquanto os estudantes de letras de menor renda e com pais de menor nível de escolaridade estão em situação de risco de frequentarem instituições não seletivas, na pedagogia os efeitos são nulos para as duas faixas de renda e para aqueles com pais com escolaridade de nível médio. Os estudantes pardos e negros estão protegidos nas duas áreas, assim como também os de origem indígena na área de pedagogia. Estudantes das categorias indicadoras de desvantagens nas outras variáveis estão sujeitos a risco, com exceção dos estudantes que têm conhecimento da língua inglesa praticamente nulo na área de letras, resultado curioso considerando a especificidade da área.

Complementando a análise comparativa entre as áreas, pode-se ressaltar que entre os cursos de maior prestígio, as engenharias e a computação, destacam-se por não proporcionar aos estudantes da faixa de menor renda, de até três salários mínimos, maior risco em relação aos estudantes de maior renda para o acesso às instituições seletivas, embora os estudantes da faixa intermediária de renda estejam em situação desfavorável. Esta tendência, não observada na distribuição de frequência dos grupos de renda nestas áreas, pode ser debitada a um efeito de superseleção nas áreas que, a princípio, exigem maiores habilidades escolares.

Afora estas exceções, o efeito da renda sobre as oportunidades de acesso às instituições seletivas reflete-se nos modelos para a maioria das áreas de formação, com exceção das áreas de arquitetura, química, ciências sociais, filosofia e pedagogia. Nas três primeiras destas áreas, entretanto, os efeitos da escolaridade do pai (positivo para todas as áreas) mostram-se mais intensos, impedindo que se afirme que os efeitos da origem social não atingem estas áreas. Pode-se afirmar que nas áreas de pedagogia e filosofia os efeitos da origem social, considerando a renda familiar e a escolaridade do pai, são menos intensos.

Os modelos indicam que os estudantes que se declaram pardos e negros estão em situação de risco em relação à frequência às instituições não seletivas na área de arquitetura, resultado único entre as áreas. Os efeitos calculados são nulos nas áreas

de física, filosofia, geografia e história (somente os negros), e indicam proteção nas áreas restantes, tendência não observada nas distribuições de frequência por área.

Em relação ao gênero as mulheres estão protegidas de frequentarem instituições seletivas nas áreas de arquitetura, engenharia e química, e estão em situação de risco nas áreas de pedagogia e letras. Estes efeitos nas áreas de engenharia, tipicamente masculina, e em pedagogia e letras, áreas com grande maioria de mulheres, podem ser debitados também a um processo de superseleção. Nas áreas restantes os efeitos de gênero não são significativos.

Os egressos das escolas públicas têm maiores risco de frequentar instituições seletivas em todas as áreas. Estudantes com conhecimento de língua inglesa também estão em situação de risco com a exceção das áreas de ciências sociais, matemática e letras, embora nesta última os estudantes com proficiência somente em leitura estejam em risco. Os estudantes com mais de 28 anos também estão em situação de risco na maioria das áreas, com as exceções de física, química e biologia.

Tabela 34**Parâmetros estimados pelos modelos multinomiais de risco de cursar o ensino superior em instituições não seletivas por área de formação**

	Arquitetura		Engenharias		Computação		Física	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2
Feminino	0,478***	0,522***	0,632***	0,647***	1,154	1,15	1,113	1,1
Masculino (REF)								
até 3 SM	0,781	0,704	1,096	1,058	0,952	0,921	1,566***	1,533***
de 3 a 10 SM	1,053	1,034	1,258***	1,244***	1,203**	1,189*	1,172	1,161
mais de 10 SM (REF)								
Ensino fundamental	2,174***	2,042***	1,876***	1,846***	1,617***	1,566***	1,501***	1,482***
Ensino médio	1,228*	1,177	1,372***	1,348***	1,341***	1,301***	1,215	1,196
Ensino superior (REF)								
Indígena (Origem Indígena)	NaN	NaN	0,803	0,777	1,19	1,159	2,317*	2,308*
Amarelo (Origem Oriental)	2,844**	2,759**	1,212*	1,154	0,914	0,857	0,632	0,605
Pardo(a) ou mulato(a)	1,49**	1,553**	0,79***	0,821***	0,803**	0,84*	1,074	1,105
Negro(a)	2,772*	2,68*	0,758***	0,742***	0,627***	0,617***	0,916	0,914
Branco(a) (REF)								
Escola pública	1,608***	1,551***	1,704***	1,709***	1,779***	1,771*	1,365***	1,357***
Esc. pública/privada	1,853***	1,653**	1,393***	1,285***	1,426***	1,303**	1,429***	1,351**
Escola privada (particular) (REF)								
Praticamente nulo	2,482***	1,968***	1,741***	1,494***	1,595***	1,357**	1,474***	1,339***
Leio, mas não falo	1,369**	1,194	1,435***	1,326***	1,131	1,044	1,145	1,089
Leio, escrevo e falo bem/raz(REF)								
Mais de 28 anos	3,103***	3,076***	3,116***	3,279***	0,747**	0,777*	1,002	0,997
De 23 a 28 anos	0,845	0,772	1,112**	1,046	1,191*	1,113	0,87	0,829
De 20 a 22 anos	0,523***	0,487***	0,731***	0,684***	0,893	0,836**	0,776**	0,745**
Até 19 anos (REF)								
Nota bruta na prova de form. geral		0,957***		0,965***		0,961***		0,974***

* p < 0,05; ** p < 0,01; *** p < 0,001

Tabela 34
Parâmetros estimados pelos modelos multinomiais de risco de cursar o ensino superior em instituições não seletivas por área de formação

	Química		Biologia		Ciências Sociais		Filosofia	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2
Feminino	0,818**	0,816**	0,903*	0,919	0,591	0,607	0,827	0,808
Masculino (REF)								
até 3 SM	1,262	1,221	2,039**	1,873**	1,483	1,393	1,060	1,02
de 3 a 10 SM	1,215	1,216	1,437**	1,406**	1,189	1,178	0,998	0,983
mais de 10 SM (REF)								
Ensino fundamental	2,367***	2,34***	1,269***	1,193***	3,191	3,085	1,628**	1,601*
Ensino médio	1,334**	1,320*	0,948	0,910	2,124	2,076	1,464*	1,443
Ensino superior (REF)								
Indígena (Origem Indígena)	1,18	1,169	0,763	0,755	0,828	0,837	0,629	0,630
Amarelo (Origem Oriental)	1,012	0,967	0,724	0,693	0,652	0,623	2,988	2,837
Pardo(a) ou mulato(a)	0,738***	0,754***	0,636***	0,657***	0,692*	0,710*	0,829	0,849
Negro(a)	0,736*	0,736*	0,574***	0,562***	0,670	0,662	0,708	0,707
Branco(a) (REF)								
Escola pública	1,294**	1,277**	1,611***	1,546***	1,418*	1,389*	1,420*	1,410*
Esc. pública/privada	1,112	1,039	1,918***	1,717***	1,257	1,163	0,981	0,930
Escola privada (particular) (REF)								
Praticamente nulo	1,481***	1,353***	1,89***	1,577***	1,258	1,084	1,259	1,142
Leio, mas não falo	1,38***	1,314**	1,548***	1,402***	1,508*	1,391	1,521**	1,451*
Leio, escrevo e falo bem/raz(REF)								
Mais de 28 anos	1,192	1,154	1,008	0,949	2,586***	2,505***	2,005***	1,978***
De 23 a 28 anos	0,894	0,850	1,272***	1,187*	1,131	1,073	1,387	1,313
De 20 a 22 anos	0,977	0,939	0,889*	0,837***	1,058	1,004	0,834	0,798
Até 19 anos (REF)								
Nota bruta na prova de form. geral		0,975***		0,955***		0,968***		0,974***

* p < 0,05; ** p < 0,01; *** p < 0,001

Tabela 34
Parâmetros estimados pelos modelos multinomiais de risco de cursar o ensino superior em instituições não seletivas por área de formação

	Geografia		História		Matemática		Letras		Pedagogia	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2
Feminino	1,057	1,045	0,979	0,971	1,452	1,437	1,160***	1,151**	1,547***	1,589***
Masculino (REF)										
até 3 SM	1,740***	1,665***	1,712***	1,646***	1,509**	1,482**	2,022***	1,858***	1,212*	1,112
de 3 a 10 SM	1,265*	1,240*	1,323**	1,307**	1,407**	1,404**	1,431***	1,398***	0,957	0,955
mais de 10 SM (REF)										
Ensino fundamental	1,578***	1,514***	1,917***	1,877***	1,499***	1,479***	1,949***	1,854***	1,401***	1,372***
Ensino médio	1,371**	1,327**	1,493***	1,469***	1,137	1,121	1,421***	1,380***	0,997	0,990
Ensino superior (REF)										
Indígena (Origem Indígena)	0,919	0,924	0,662*	0,660*	0,547*	0,547*	0,987	0,970	0,580**	0,554**
Amarelo (Origem Oriental)	1,405	1,352	1,698	1,642	1,053	1,015	1,040	1,021	1,642*	1,506
Pardo(a) ou mulato(a)	0,850*	0,880	0,755***	0,770***	0,594***	0,607***	0,848***	0,857***	0,584***	0,582***
Negro(a)	1,228	1,225	0,844	0,841	0,650**	0,649**	0,781***	0,770***	0,541***	0,542***
Branco(a) (REF)										
Escola pública	1,213*	1,188*	1,013	1,000	2,314***	2,29***	1,565***	1,516***	1,729***	1,677***
Esc. pública/privada	1,408**	1,311*	1,236*	1,172	2,713***	2,570***	1,322***	1,228**	1,537***	1,387***
Escola privada (particular) (REF)										
Praticamente nulo	1,574***	1,377***	1,468***	1,343***	1,298**	1,189	1,021	0,949	1,531***	1,408***
Leio, mas não falo	1,384***	1,284**	1,715***	1,635***	1,080	1,030	1,284***	1,235***	1,440***	1,373***
Leio, escrevo e falo bem/raz(REF)										
Mais de 28 anos	2,161***	2,078***	1,945***	1,877***	1,938***	1,903***	1,575***	1,502***	2,082***	2,003***
De 23 a 28 anos	1,738***	1,637***	1,368***	1,309**	1,297**	1,246*	1,374***	1,309***	1,140*	1,088
De 20 a 22 anos	1,271**	1,213*	1,042	1,005	0,910	0,878	1,072	1,034	0,962	0,926
Até 19 anos (REF)										
Nota bruta na prova de form. geral		0,964***		0,977***		0,977***		0,974***		0,965***

* p < 0,05; ** p < 0,01; *** p < 0,001

5.4. Discussão dos resultados

A partir das descrições e análises realizadas pode-se observar que as populações de estudantes diferem significativamente entre as instituições de ensino, tomando como critério de diferenciação a sua seletividade, e entre as áreas de formação, tendo como critério seu prestígio, avaliado por meio da renda familiar e da escolaridade paterna. Ou seja, tanto a seletividade das instituições quanto o prestígio das áreas constituem elementos definidores de trajetórias educacionais distintas, em termos das oportunidades educacionais que oferecem.

Entre as duas dimensões analisadas nas variáveis independentes, o prestígio das áreas de formação parece ser um critério de seleção mais relevante que a seletividade das instituições. O modelo binomial relativo às distinções entre as áreas tecnológicas e as licenciaturas tem maior poder explicativo sobre as trajetórias dos estudantes do que o modelo relativo à seletividade, e os efeitos das variáveis de origem social no primeiro modelo são mais relevantes. A renda aparece como um fator de risco mais importante para o acesso às áreas de menor prestígio do que às instituições não seletivas.

Quanto às variáveis que denotam características individuais, cabe destacar que os estudantes pardos e negros têm maiores chances de frequentar instituições seletivas quando comparados com os estudantes que se declaram brancos, embora tenham maiores riscos de frequentar áreas de formação de baixo prestígio. As mulheres, especificamente, têm maiores oportunidades de frequentar instituições seletivas quando optam por áreas de formação de alto prestígio, onde são minoria.

Dentre as variáveis que procuram caracterizar a trajetória escolar do estudante, é visível a tendência de estudantes provenientes de escolas públicas estudarem em áreas de baixo prestígio e em instituições não seletivas. Os estudantes mais velhos, com idade superior a 22 anos, estão mais sujeitos a ingressar em instituições não seletivas e em carreiras de baixo prestígio; entretanto, estudantes com idade entre 20 e 22 anos ocupam melhores posições que os estudantes com idade até 19 anos.

O conhecimento de língua inglesa também se mostrou uma variável de risco em relação às duas dimensões analisadas. A nota na prova de formação geral é fator de proteção somente em relação à seletividade das instituições, enquanto estudantes que obtiveram notas maiores encaminham-se preferencialmente para áreas de baixo prestígio, o que pode indicar que os conhecimentos avaliados neste exame são menos valorizados pelos estudantes das áreas tecnológicas, preponderantes entre as áreas de formação de alto prestígio. De fato, verifica-se que os estudantes das licenciaturas alcançam médias mais altas na prova de formação geral do que os estudantes das áreas tecnológicas.

Além de estarem mais sujeitos ao risco de estudar em áreas de formação de baixo prestígio de uma forma geral, os estudantes de menor renda preferencialmente encaminham-se para as áreas de baixo prestígio em instituições seletivas, quando comparados os riscos de frequência entre e as áreas de alto prestígio em instituições não seletivas.

Observando separadamente cada uma das áreas de formação, os resultados comportam muitas variações. Entre as áreas de maior prestígio, a área de arquitetura destaca-se por seu caráter seletivo em relação à cor declarada – é a única área onde estudantes pardos e negros têm maiores chances de frequentar instituições não seletivas. Por outro lado, nesta área não são registrados efeitos da renda familiar. Há que se destacar, também, as maiores chances de estudantes de renda até três salários mínimos, pardos e negros das áreas de engenharia e computação frequentarem instituições seletivas, quando comparados com estudantes de renda maior que dez salários mínimos. Áreas de grande prestígio no mercado de trabalho, a engenharia e a computação aparecem como áreas nas quais o acesso aos estudantes de perfil social mais desfavorecido é muito restrito; entretanto os estudantes “sobreviventes” nestas áreas têm como opção mais provável as instituições seletivas.

Entre as áreas de menor prestígio, destacam-se as áreas de ciências sociais, filosofia, química e pedagogia como as menos seletivas; nestas áreas a renda não aparece como fator de seleção dos estudantes. A área de ciências sociais é a única em que os efeitos da escolaridade do pai não são observados.

Curiosamente, na área onde existe a maior presença relativa de homens, a engenharia, as mulheres têm maiores chances de frequentar instituições seletivas do que os homens; e nas áreas onde existe a maior presença de mulheres, pedagogia e letras, os homens têm maiores chances de acesso às instituições seletivas. Estes resultados indicam processos de superseleção.

Finalmente, verifica-se que o efeito da rede em que o estudante cursou o ensino médio é o mais uniforme entre as áreas – em todas elas, com exceção da área de história, é um fator de risco para a frequência às instituições seletivas.