

Caio Henrique Pires e Souza

Análise da inversão dos ingressantes das IES dada a origem escolar

Niterói – RJ, Brasil

2014



Universidade Federal Fluminense

Caio Henrique Pires e Souza

Análise da inversão dos ingressantes das IES dada a origem escolar

Trabalho de Conclusão de Curso

Monografia apresentada ao Curso de Estatística
Da UFF, como requisito para a obtenção do grau de
BACHAREL em Estatística

Orientador: Márcia de Carvalho
Doutora em Estatística

Niterói - RJ, Brasil

2014



Universidade Federal Fluminense

Caio Henrique Pires e Souza

Análise da inversão dos ingressantes das IES dada a origem escolar

Projeto Final sob o título “*Análise da inversão dos ingressantes das IES dado à origem escolar*”, apresentado por Caio Henrique Pires e Souza e aprovada em 5 de junho de 2014, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Márcia de Carvalho
Orientadora
Departamento de Estatística – UFF

Profa. Dra. Hustana Maria Vargas
Faculdade de Educação - UFF

Prof. Dr. Jony Arrais Pinto Junior
Departamento de Estatística – UFF

Niterói, 5 de junho de 2014

S729 Souza, Caio Henrique Pires e
Análise da inversão dos ingressantes das IES dada a origem
escolar / Caio Henrique Pires e Souza; Márcia Marques
de Carvalho, orientadora. Niterói, 2014.
41 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística) - Universidade Federal Fluminense,
Instituto de Matemática e Estatística, Niterói, 2014.

1. Inversão estatística. 2. Ensino médio. 3. Exame Nacional
do Ensino Médio. 4. Regressão logística. I. Carvalho, Márcia
Marques de, orientadora. II. Universidade Federal Fluminense.
Instituto de Matemática e Estatística. III. Título.

CDD -519.5

Resumo

Esse estudo analisa um fenômeno que está relacionado à origem escolar dos jovens ao ingressar no ensino superior: a inversão da dependência administrativa das instituições que ocorre quando egressos do ensino médio público ingressam em instituições privadas e vice e versa. O objetivo é determinar quais variáveis contribuem para essa inversão ao ingressar no ensino superior. Com os microdados do Exame Nacional de Desempenho de Estudantes (ENADE), coletado pelo MEC/INEP e estimando o modelo de regressão logística, conclui-se que as variáveis que mais contribuem para o ingresso no ensino superior privado, dado que cursou o ensino médio em escola pública, são a renda típica de classe média(3 a 10 s.m.) e residir na região Sudeste, que possui grande oferta de cursos de IES privadas.

Dedicatória

Eu dedico esse trabalho a Velúzia Pires da Silva, minha mãe. E ao Albucaçys Maurício de Paula Filho, meu pai de coração. Sem estas pessoas, que me apoiaram psicologicamente e emocionalmente eu, sem dúvida, não estaria apresentando o meu Trabalho de Conclusão de Curso (TCC). Digo mais, sem eles eu nem teria ingressado no curso de Estatística da UFF e realizado este trabalho, que justamente diz respeito ao ingresso no ensino superior. Devo a eles minha vida e espero retribuir um dia. Obrigado.

Agradecimentos

Fico satisfeito em dizer que muitas pessoas contribuíram direta e indiretamente para a chegada deste caminho, chamado TCC. Primeiramente, a pessoa que abraçou a ideia comigo e fez tudo para que o trabalho acontecesse, minha orientadora Márcia de Carvalho. Muito Obrigado!

Com o final de uma etapa, algumas pessoas ficam marcadas em nossas vidas. Agradeço aos meus amigos de república, que me fizeram feliz e me fizeram crescer como pessoa. Também agradeço aos companheiros do curso, que me incentivaram a estudar com eles e, de certa forma, me mostraram que era possível. Gostaria de oferecer meu sincero agradecimento, para todo o corpo docente que me ensinou durante a minha graduação, e me desculpar se não fui um aluno perfeito.

Como não poderia faltar, um parágrafo destinado à família. Agradeço aos meus irmãos e irmãs, que por sinal, são muitos! Tenho 2 irmãos e 4 irmãs, obrigado por existirem. A meu pai, que soube me dar conselhos sábios e é um exemplo para mim. Espero ser tão bem sucedido quanto ele. A minha mãe que virou três ou mais pessoas para me servir e ajudar de todas as maneiras possíveis. Beijo mãe. Ao meu padrasto, que me acompanha desde criança e me ensinou a sempre priorizar o bem e sempre optar pela paz. Agradeço a família como um todo, sei que todos torceram por mim.

Esse trabalho é resultado da soma de todos que foram e não foram mencionados. Desejo um dia poder ter a capacidade de retribuir a todos que me ajudaram de alguma forma.

Sumário

1 - Introdução	11
2 – Educação Superior no Brasil	12
2.1 – Oferta: Cursos IES Pública versus Privada, por região.....	12
2.2 – Perfil dos Ingressantes.....	17
3 - Modelo de Regressão Logística	22
3.1 – Modelos Lineares Generalizados	22
3.2 – Especificação do Modelo de Regressão Logística	24
3.3 - Estimção pelo método da máxima verossimilhança	24
3.4 - Interpretação dos parâmetros estimados	26
3.5 - Inferência do modelo	29
3.6 - Qualidade do Ajuste.....	30
3.7 -Técnicas de Diagnóstico	31
4 – Resultados	33
4.1 – A base de dados do ENADE	33
4.2 – Descrição das Variáveis	33
4.3 – Estatísticas Descritivas	34
4.4 – Modelos Estimados	39
Modelo 1 – Inversão 1: ensino médio privado para IES pública	39
Modelo 2 – Inversão 2: ensino médio público para IES privada	40
4.5 – Qualidade do Ajuste.....	42
Modelo 1 – Inversão 1.....	42
Modelo 2 – Inversão 2.....	43
4.6 – Técnicas de Ajuste	44
5 - Conclusão.....	47
Referências	48

Sumário de tabelas

Tabela 1 - Número de IES Públicas e Privadas por Região.	13
Tabela 2 - Ofertas de cursos de graduação de IES Públicas segunda a região.	13
Tabela 3 - Ofertas de cursos de graduação de IES Privadas segunda a região.	14
Tabela 4 - Relação Candidato/Vaga IES Públicas e Privadas no Brasil.	15
Tabela 5 - Número de Matrículas em IES Públicas e Privadas por Curso em 2010, no Brasil... ..	16
Tabela 6 – Tipo de Ensino Médio Cursado em 2008, 2009 e 2010, por curso.	18
Tabela 7 - Nível de escolaridade do pai por curso nos anos de 2008, 2009 e 2010.	20
Tabela 8 -Renda familiar em salários por curso nos anos de 2008, 2009 e 2010.	21
Tabela 9 – Análise descrita do perfil dos ingressantes por curso selecionado, 2008 - 2010.	35
Tabela 10 – Inversão dos ingressantes do ensino superior, por curso.	36
Tabela 11 – Inversão por Idade, em cada curso.	37
Tabela 12 - Estimativa dos parâmetros referentes ao modelo logístico.	40
Tabela 13 – Estimativa dos parâmetros referentes ao modelo logístico.	41
Tabela 14 – Qualidade do Ajuste.	42
Tabela 15 – Tabela de Classificação.	43
Tabela 16 – Qualidade do Ajuste.	43
Tabela 17 – Tabela de classificação.	43

Sumário de figuras

Figura 1- Gráficos dos resíduos de Engenharia, Inversão 1 e 2.	44
Figura 2 – Gráficos dos resíduos de Matemática, Inversão 1 e 2.	44
Figura 3 - Gráficos dos resíduos de Economia, Inversão 1 e 2.....	44
Figura 4 - Gráficos dos resíduos de Medicina, Inversão 1 e 2.....	45
Figura 5 – Gráficos dos resíduos de Odontologia, Inversão 1 e 2.	45
Figura 6 - Gráficos dos resíduos de Relações Internacionais, Inversão 1 e 2.....	45

1 - Introdução

Certos acontecimentos se tornam frequentes, tal como o qual iremos estudar e que afeta a área da educação superior no Brasil. Esse trabalho analisa um fenômeno que está relacionado com o nível escolar dos jovens e o ingresso desses nas Instituições de Ensino Superior (IES). De acordo com o artigo de Pinto (2004) verificamos que a participação dos jovens provenientes de ensino público nas Universidades públicas é baixa, dado que eles representavam, em 2007, 86% dos concluintes do ensino médio. Este acontecimento pode ser devido à baixa oferta de vagas das Universidades Públicas e também pela forte concorrência que os alunos de ensino público, que não estão preparados em comparação com os de ensino privado, enfrentam ao prestar o vestibular. Através desse mesmo artigo verificamos uma questão mais problemática: quando os alunos ingressam na IES um processo inverso se ocorre, dados que alunos provenientes de colégios particulares, geralmente, ingressam em universidades públicas e os alunos de colégios públicos tendem a ingressar em universidades privadas. Doravante, esse fenômeno será chamado de inversão.

O objetivo principal desse trabalho é verificar quais variáveis que contribuem mais para a tendência que os alunos sofrem ao trocarem o ensino médio público pelo ensino superior privado e vice-versa. Após a análise dos dados é possível obter a resposta para estas pergunta e, talvez, visualizar uma possível conduta viável para melhora do cenário atual, tendo conhecimento que o ensino superior público oferece, geralmente, uma qualidade de ensino melhor que as instituições privadas e que é do interesse dos alunos, independente do tipo de ensino que procedem, ingressarem em universidades públicas.

Apesar de em Britto (2008) esta definido o aluno “novo” como sendo uma parcela da população que até então não ingressa em Universidades, públicas e privadas, ter mencionado o aumento no número de ingressantes, ao longo dos anos, em IES, percebe-se que apesar do número de universitários ter aumentado, estes estão concentrados, em sua maioria, em universidade particulares, pois as vagas nas universidades públicas não aumentaram de maneira proporcional a demanda de alunos que tem saído do ensino médio.

É importante mencionar que não estamos partindo da premissa de que o rico que cursa escola básica particular vai para universidade pública e o pobre, por consequência, o inverso. A ideia que queremos enfatizar é que os estudantes de escolas particulares estão mais bem preparados para prestar o vestibular e ingressar em universidades públicas, cujas vagas são escassas resultando que o ensino superior se torna mais elitizado. O objetivo é estudar que características o estudante de escolas particulares possuem que os estudantes de escolas públicas não possuem. Daí analisar o nível escolar dos pais, raça, renda, região geográfica, idade, gênero entre outras em busca medir ou determinar uma tendência que se repita para traçar o perfil do aluno que sofre as consequências da elitização do ensino superior. Lembrando que esta é uma ideia generalizada e, com certeza, há exceções.

2 – Educação Superior no Brasil

Esse capítulo apresenta um panorama do nosso objetivo de estudo: Educação Superior no Brasil, focado na temática abordada. Primeiramente será feita uma análise da educação do Brasil por região e curso e depois será realizada uma análise sobre o perfil dos ingressantes nas IES de acordo com algumas variáveis, ambas referentes aos anos de 2008, 2009 e 2010. Conforme mencionado na introdução do trabalho, há interesse em verificar quais variáveis influenciam de maneira mais significativa no processo de inversão no ingresso das IES públicas/privadas, de acordo com o tipo de ensino médio cursado. Através dos dados coletados será possível observar se há uma tendência que favorece a nossa hipótese de que existe esta inversão. Também será informado o número de concluintes do ensino médio e qual a procedência desses concluintes, referentes aos três anos.

2.1 – Oferta: Cursos IES Pública versus Privada, por região.

Iniciando pelo lado da oferta da educação a Tabela 1 apresenta o número de cursos na graduação, número de ingressos e o número de concluintes do Ensino Médio público nos anos de 2008, 2009 e 2010, por região nas IES públicas.

Tabela 1 - Número de IES Públicas e Privadas por Região.

Região	Pública			Privada		
	2008	2009	2010	2008	2009	2010
Brasil	236	245	278	2.016	2.069	2.100
Norte	18	19	25	121	128	121
Nordeste	59	61	64	373	387	369
Sudeste	104	110	131	965	980	1.038
Sul	39	39	41	331	347	345
Centro-Oeste	16	16	17	226	227	227

Fonte: MEC/INEP. Censo da Educação Superior.

Na Tabela 1 observa-se o número de IES Públicas e Privadas nos anos de 2008, 2009 e 2010, por região. É importante destacar que o número de instituições privadas é quase dez vezes maior que o número de instituições públicas e que a quantidade de instituições privadas cresceu mais que as públicas. Assim como o número de cursos observados na Tabela 1 o número de Instituições é maior no Sudeste e menor no Centro-Oeste e Norte.

Ao realizar uma análise descritiva das tabelas anteriores identifica-se alguns fatores que comprovam a ideia de inversão. Torna-se interessante gerar uma tabela que informe o número de matrículas, por curso, nas instituições privadas e públicas e a porcentagem de matrículas desses cursos que são fornecidos pelas instituições públicas. Dessa forma torna-se possível verificar o quanto dessas matrículas é representada pelas instituições públicas e privadas.

Tabela 2 - Ofertas de cursos de graduação de IES Públicas segunda a região.

Região	Nº de Ingressos na Graduação			Concluintes do Ens. Médio		
	2008	2009	2010	2008	2009	2010
Brasil	1.273.965	1.351.168	1.461.696	1.491.812	1.525.289	1.522.379
Norte	136.519	128.689	152.469	128.953	133.205	134.050
Nordeste	383.539	409.393	438.090	459.357	471.516	460.891
Sudeste	394.903	441.800	493.881	594.866	614.412	633.959
Sul	237.384	246.882	242.367	208.380	207.242	194.238
Centro-Oeste	121.620	124.404	134.889	100.256	98.914	99.241

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).

Através da Tabela 2 podemos observar que o número de cursos que as IES fornecem ao longo dos anos analisados e que a região com menor quantidade de cursos oferecidos é a Centro-Oeste, seguida da região Norte, e as que mais oferecem são as

regiões Sudeste e Nordeste. Com relação ao número de concluintes do Ensino Médio Público percebemos que esse número também aumentou nesses anos. De forma geral a região Centro-Oeste e Norte são as que apresentam menor número.

Tabela 3 - Ofertas de cursos de graduação de IES Privadas segunda a região.

Região	Nºde Matrículas na Graduação			Concluintes Ens. Médio		
	2008	2009	2010	2008	2009	2010
Brasil	3.806.091	3.764.728	3.987.424	269.613	272.145	270.788
Norte	186.671	185.270	199.889	10.981	12.983	13.275
Nordeste	529.154	556.109	614.071	68.661	69.176	69.633
Sudeste	2.117.657	2.074.912	2.162.350	126.513	127.355	126.174
Sul	649.798	619.054	650.763	39.605	39.831	38.404
Centro-Oeste	322.811	329.383	360.351	23.853	22.800	23.302

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).

Na Tabela 3, podemos visualizar as mesmas questões para as instituições privadas. O número de cursos na graduação aumentou ao longo dos anos analisados da mesma forma que nas instituições públicas, e da mesma forma as regiões com menor número de cursos fornecidos são as regiões Centro-Oeste e Norte. A região com maior número de cursos oferecidos é a região Sudeste, sendo quase cinco vezes maior que nas outras regiões. O número de matrículas cresceu, porém não cresceu na mesma proporção que nas IES públicas, nesses anos. Observa-se que as matrículas são quase três vezes maior que nas IES privadas que nas públicas. A população com ensino médio completo de colégios privadas é muito inferior a população com ensino médio completo de colégios públicos. A partir dessa análise descritiva pode-se levantar algumas questões, como o por que do número de matrículas ser quase três vezes maior nas IES privadas do que nas públicas e por qual estudante estas vagas são preenchidas, dado que o número de concluintes do ensino médio privado não é o suficiente para suprir a quantidade de vagas.

Agora faremos uma análise descritiva da relação candidato/ vaga, nas IES Públicas nos cursos que foram analisados pelo ENADE nos anos de 2008, 2009 e 2010. As informações foram coletadas através do INEP.

Tabela 4 - Relação Candidato/Vaga IES Públicas e Privadas no Brasil

Curso	Ano		
	2008	2009	2010
Medicina	22,20	23,58	32,91
Letras	1,60	4,99	-
Relações internacionais	2,46	3,64	2,05
Odontologia	3,62	3,24	4,29
Medicina veterinária	3,40	3,22	4,16
Agronomia	3,32	3,04	4,19
Administração	1,62	2,98	1,53
Zootecnia	2,79	2,73	5,32
Química	3,23	2,67	2,55
Física	3,73	2,58	2,76
Geografia	3,97	2,51	2,91
Ciências contábeis	1,36	2,49	1,42
Arquivologia	9,30	2,41	7,92
Estatística	3,09	2,38	3,24
Biblioteconomia	3,33	2,33	4,42
Engenharia Civil	2,91	2,25	3,67
Filosofia	2,42	2,14	2,10
Nutrição	1,66	2,13	2,21
Enfermagem	1,98	2,12	2,21
Direito	2,57	2,11	2,89
Terapia ocupacional	1,39	2,07	3,09
Serviço social	1,70	1,92	2,42
Arquitetura	3,35	1,86	3,00
Ciência da computação	1,76	1,73	2,06
Comunicação social	1,47	1,68	1,51
Farmácia	1,76	1,67	2,07
História	4,51	1,64	4,32
Biomedicina	1,28	1,64	1,75
Economia	2,39	1,58	2,62
Educação física	1,74	1,53	1,45
Matemática	2,74	1,51	1,93
Ciências sociais	4,41	1,44	5,86
Fisioterapia	1,40	1,43	1,56
Engenharia Elétrica	1,67	1,42	2,34
Pedagogia	1,20	1,32	1,40
Fonoaudiologia	0,98	1,20	1,24
Ciências Biológicas	2,88	1,00	3,15
Engenharia de Produção	2,12	0,98	2,00
Turismo	1,09	0,88	1,30

Fonte: MEC/INEP. Censo da Educação Superior.

Nota: - Não encontrado em 2010.

Na Tabela 4 observa-se que o curso mais concorrido no período analisado é o de Medicina, assim como esperado. E um dos menos concorridos é o curso de Turismo. Cursos como Pedagogia, Fonoaudiologia e Fisioterapia também apresentam baixa relação/candidato vaga. De modo geral a relação candidato/vaga não aumentou tanto ao longo dos anos, o que reforça a ideia do aluno “novo”(de Britto et al, 2008) que representa uma parcela da população que até então não ingressava no ensino superior, mas em sua grande maioria ingressa nas universidades privadas, já que as quantidade de vagas não aumentou nas universidades públicas, o que justifica o pequeno aumento na relação candidato/vaga observado na Tabela 4.

Tabela 5 - Número de Matrículas em IES Públicas e Privadas por Curso em 2010, no Brasil.

Curso	Matrículas			
	Total	Pública	Privada	(%)Pública
Biomedicina	30.333	2.996	27.337	9,88
Direito	694.447	70.223	624.224	10,11
Fisioterapia	99.216	10.098	89.118	10,18
Administração	705.690	72.195	633.495	10,23
Enfermagem	244.568	33.904	210.664	13,86
Relações internacionais	18.196	3.123	15.073	17,16
Ciências contábeis	224.228	42.049	182.179	18,75
Educação física	49.238	9.347	39.891	18,98
Engenharia de Produção	87.208	18.112	69.096	20,77
Farmácia	101.816	21.592	80.224	21,21
Nutrição	64.090	13.619	50.471	21,25
Arquitetura	73.631	16.497	57.134	22,40
Comunicação social	55.563	12.552	43.011	22,59
Serviço social	68.724	15.876	52.848	23,10
Engenharia Civil	99.521	29.239	70.282	29,38
Pedagogia	297.581	93.886	203.695	31,55
Turismo	30.051	9.687	20.364	32,24
Odontologia	57.603	18.864	38.739	32,75
Fonoaudiologia	8.735	2.940	5.795	33,66
Ciência da computação	62.297	22.333	39.964	35,85
Medicina veterinária	46.530	17.828	28.702	38,32
Medicina	103.312	41.054	62.258	39,74
Engenharia Elétrica	59.040	23.879	35.161	40,45
Ciências Biológicas	31.106	14.829	16.277	47,67
Terapia ocupacional	4.997	2.561	2.436	51,25
Química	19.224	11.244	7.980	58,49
Filosofia	5.396	3.392	2.004	62,86
Economia	50.440	31.882	18.558	63,21
Agronomia	50.653	33.845	16.808	66,82
História	8.351	6.577	1.774	78,76

Geografia	7.794	6.498	1.296	83,37
Biblioteconomia	8.498	7.383	1.115	86,88
Zootecnia	15.802	13.814	1.988	87,42
Ciências sociais	9.950	9.032	918	90,77
Matemática	3.184	3.076	108	96,61
Física	5.918	5.760	158	97,33
Estatística	4.784	4.717	67	98,60
Arquivologia	2.685	2.685	0	100,00
Letras	-	-	-	-

Fonte: MEC/INEP. Censo da Educação Superior.

Nota: - Não encontrado em 2010.

Através da Tabela 5 verifica-se que quase 60% dos cursos tem um número maior de matrículas em instituições privadas. Os cursos que possuem maior número de vagas nas IES Públicas são, em geral, cursos acadêmicos. Temos que os cursos que oferecem um futuro mestrado ou doutorado, cursos acadêmicos, possuem um número maior de matrícula nas IES públicas e os cursos que possibilitam um retorno imediato, com relação ao mercado de trabalho, estão centrados nas IES Privadas.

2.2 – Perfil dos Ingressantes

Nessa introdução ao ensino superior, o perfil dos ingressantes nas IES será analisado com relação a três variáveis: tipo de ensino médio cursado, escolaridade do pai e renda familiar. Escolhemos estas, pois acredita-se que são variáveis que influenciam de forma significativa com a inversão. Pode-se verificar qual o perfil do aluno, para estas variáveis, nesses anos e, de certa forma, ampliar nosso conhecimento sobre os alunos que ingressam nas IES públicas e privadas. Os dados foram coletados através dos questionários aplicados nos respectivos anos, e os cursos mencionados são os que foram selecionados pelo ENADE nesses anos.

Na Tabela 6, estão os dados da maior parte do tipo de ensino médio cursado pelos ingressantes, podendo ser de escola pública ou privada, por curso. Pode-se verificar que a maioria dos cursos no ano de 2008 tem seus alunos provenientes de escola pública. Neste ano, apenas os cursos de arquitetura e urbanismo e engenharia civil que apresentam uma porcentagem maior de estudantes provenientes de escola privada. No ano de 2009 percebe-se que o mesmo comportamento se repete, ou seja, a maior porcentagem dos ingressantes, nas IES, são de escolas públicas e apenas o curso de Relações Internacionais é diferente da maioria. Em 2010 também observa-se o

mesmo comportamento. Somente os cursos de Medicina, Medicina Veterinário e Terapia Ocupacional que apresentam uma maior porcentagem de alunos de escolas privadas. De modo geral, em todos os anos, há uma prevalência de alunos que cursaram o ensino médio público. Este resultado comprova o que foi visto na tabela 1 e na tabela 2, a qual mostra o número de concluintes do ensino médio público e privado, respectivamente, sendo os de ensino médio público quase dez vezes maior que os de ensino médio privado.

Tabela 6 – Tipo de Ensino Médio Cursado em 2008, 2009 e 2010, por curso.

2008			
Curso	Escola Pública (%)	Escola Privada (%)	Outros(%)
Arquitetura e Urbanismo	28,5	69,0	2,5
Engenharia Civil	42,7	54,9	2,4
Pedagogia	81,1	15,2	3,7
Ciência da computação	49,5	47,7	2,8
Ciências Biológicas	69,6	27,5	2,9
Química	67,6	30,6	1,8
Filosofia	68,1	26,6	5,3
História	70,9	24,5	4,6
Geografia	74,2	22,6	3,2
Ciências sociais	53,1	42,0	4,9
Matemática	81,0	16,4	2,6
Física	65,4	33,3	1,3
Letras	76,7	19,1	4,2
2009			
Curso	Escola Pública (%)	Escola Privada (%)	Outros(%)
Administração	71,1	25,9	3
Direito	50,7	45,6	3,7
Relações internacionais	27,2	70,8	2
Ciências contábeis	78,9	18,6	2,5
Com. Social - Jornalismo	47,6	50,0	2,4
Turismo	62,8	34,4	2,8
Economia	50,2	48,1	1,7
Biblioteconomia	68,4	28,6	3
Estatística	48,1	50,4	1,5
Arquivologia	72,2	26,5	1,3
2010			
Curso	Escola Pública (%)	Escola Privada (%)	Outros(%)
Biomedicina	58,4	38,8	2,8
Enfermagem	68,8	30,9	0,3
Educação física	68,5	28,4	3,1
Farmácia	62,5	34,8	2,7
Nutrição	52,6	44,7	2,7
Fisioterapia	51,4	45,4	3,2

Serviço social	77,9	17,3	4,8
Odontologia	30,8	66,1	3,1
Fonoaudiologia	50,1	47,2	2,7
Medicina veterinária	35	62,10	2,7
Medicina	17,2	81,7	1,1
Terapia ocupacional	46,4	50,2	3,4

Fonte: MEC/INEP. Exame Nacional de Desempenho dos Estudantes (ENADE).

A segunda variável de interesse, com relação ao perfil dos ingressantes nas IES, é o nível escolaridade do pai. Um dos principais fatores que influênciam no ingresso, nas IES públicas e privadas, e na escolha do curso é o nível de escolaridade dos pais. Acredita-se que o nível educacional do pai, tem maior impacto no ingresso no ensino superior dos filhos.

Nos cursos avaliados em 2008, os cursos de Arquitetura e Urbanismo e Engenharia fogem a regra, tendo o maior nível escolar do pai concentrado no ensino superior. Dessa vez também verificamos que o curso de Ciências da Computação tem uma boa parcela do nível escolar do pai até no ensino médio e superior. No ano de 2009, observa-se que o curso de Relações Internacionais possui mais de 50% do nível escolar do pai no ensino superior. Outros cursos, como Direito, Economia, Comunicação Social, voltada para área de Jornalismo, e Estatística possuem o nível escolar do pai similar entre os tipos de ensino e os outros tem o nível do pai mais concentrado no ensino fundamental. E, por último, no ano de 2010 tem-se que os cursos de Medicina, Medicina Veterinária e Odontologia possuem, como esperado, a maior porcentagem do nível escolar do pai no ensino superior, destacando-se medicina com 66,1% grau de escolaridade do pai nesta categoria. E, para os cursos de Enfermagem, Nutrição, Serviço Social, Farmácia e Educação Física o nível escolar do pai se concentra mais no ensino fundamental.

Ao analisarmos a Tabela 8, a qual trata da renda familiar dos ingressantes, percebe-se que, em todos os anos pesquisados, a porcentagem de renda familiar superior a trinta salários mínimos é pequena. Dentre desta mínima porcentagem alguns cursos se destacam com mais, como os cursos de Arquitetura e Urbanismo e Engenharia Cível em 2008, outros cursos como Relações Internacionais, Direito e Economia em 2009, e em 2010 os cursos de Odontologia, Medicina e, um pouco inferior, Medicina Veterinária. A maioria dos ingressantes possuem uma renda familiar de até três salários mínimos. Tem-se que os cursos com maior porcentagem dos ingressantes com esta renda são

História, Geografia, Matemática, Pedagogia e Letras em 2008, Biblioteconomia e Arquivologia em 2009, e em 2010 Serviço Social. Os outros cursos estão distribuídos, em sua maioria, principalmente na categoria de até três salários mínimos e entre três e dez salários mínimos.

Tabela 7 - Nível de escolaridade do pai por curso nos anos de 2008, 2009 e 2010.

2008					
Curso	Nenhuma (%)	En. Fund. (%)	En. Médio (%)	En. Superior(%)	Outros (%)
Arquitetura e Urbanismo	1,0	18,8	28,9	50,8	0,5
Engenharia Civil	2,4	27,9	30,6	38,7	0,4
Pedagogia	13,1	62,6	17,1	6,4	0,8
Ciência da computação	1,3	30,3	33,7	33,5	1,2
Ciências Biológicas	5,3	47,5	29,5	16,3	1,4
Química	5,2	45,7	33,6	14,7	0,8
Filosofia	9,7	54,4	20,6	13,4	1,9
História	13,9	54,1	20,2	10,5	1,3
Geografia	11,7	57,7	20,7	8,7	1,2
Ciências sociais	5,1	38,0	30,7	24,0	2,2
Matemática	11,3	60,8	20,3	6,7	0,9
Física	6,9	42,9	32,4	17,3	0,5
Letras	11,2	58,6	20,2	8,5	1,5
2009					
Curso	Nenhuma (%)	En. Fund. (%)	En. Médio (%)	En. Superior(%)	Outros (%)
Administração	5,4	49,4	30,4	14,7	0,1
Direito	4,5	34,6	30,7	30,2	0,0
Relações internacionais	1,4	14,6	30,4	53,7	0,0
Ciências contábeis	6,2	55,8	27,5	10,5	0,0
Com. Social - Jornalismo	2,8	30,0	34,3	32,9	0,0
Turismo	4,1	37,6	34,9	23,4	0,0
Economia	3,7	31,9	33,7	30,7	0,0
Biblioteconomia	6,9	47,8	33,0	12,3	0,0
Estatística	3,6	29,4	37,6	29,4	0,0
Arquivologia	5,8	46,3	28,5	19,4	0,0
2010					
Curso	Nenhuma (%)	En. Fund. (%)	En. Médio (%)	En. Superior(%)	Outros (%)
Biomedicina	1,9	32,1	38,4	27,2	0,4
Enfermagem	6,2	48,2	31,5	13,5	0,6
Educação física	3,6	42,0	33,5	20,2	0,7
Farmácia	3,0	37,5	35,5	23,7	0,3
Nutrição	2,3	37,6	35,5	24,1	0,5
Fisioterapia	2,5	34,1	36,2	26,7	0,5
Serviço social	14,0	59,6	19,4	5,9	1,1
Odontologia	1,4	21,1	31,9	45,1	0,5

Fonoaudiologia	2,7	31,8	37,5	27,4	0,6
Medicina veterinária	1,3	22,9	33,2	42,2	0,4
Medicina	0,5	11,4	21,7	66,1	0,3
Terapia ocupacional	2,2	33,7	39,7	23,9	0,5

Fonte: MEC/INEP. Exame Nacional de Desempenho dos Estudantes (ENADE).

Tabela 8 - Renda familiar em salários por curso nos anos de 2008, 2009 e 2010.

2008				
Curso	Até 3 (%)	Entre 3 e 10 (%)	Entre 10 e 30 (%)	Acima de 30 (%)
Arquitetura e Urbanismo	9,1	42,2	39,2	9,7
Engenharia Civil	12,3	46,7	32,9	7,6
Pedagogia	53,2	38,9	5,5	0,4
Ciência da computação	14,9	53,0	27,9	3,4
Ciências Biológicas	46,4	41,0	10,7	1,2
Química	35,8	51,9	11,1	0,8
Filosofia	48,9	35,4	11,1	2,0
História	54,8	34,9	7,8	0,7
Geografia	53,7	36,8	6,8	0,6
Ciências sociais	33,8	58,7	3,7	1,7
Matemática	53,2	40,1	5,6	0,3
Física	37,4	46,9	13,9	1,3
Letras	55,1	36,5	6,3	0,5
2009				
Curso	Até 3 (%)	Entre 3 e 10 (%)	Entre 10 e 30 (%)	Acima de 30 (%)
Administração	41,5	47,6	8,4	2,4
Direito	30,4	46,1	18,0	5,6
Relações internacionais	23,1	38,3	25,9	12,7
Ciências contábeis	45,9	47,6	12,1	5,7
Com. Social - Jornalismo	35,7	45,5	14,8	4,1
Turismo	46,9	42,1	8,7	2,4
Economia	35,4	42,3	16,0	6,3
Biblioteconomia	57,2	35,0	7,0	0,8
Estatística	37,8	44,9	15,5	2,0
Arquivologia	52,2	37,9	9,0	1,3
2010				
Curso	Até 3 (%)	Entre 3 e 10 (%)	Entre 10 e 30 (%)	Acima de 30 (%)
Biomedicina	40,2	50,0	8,4	1,2
Enfermagem	41,6	51,6	6,1	0,5
Educação física	35,2	55,3	8,4	0,9
Farmácia	44,4	48,0	6,7	0,7
Nutrição	32,1	53,5	12,7	1,7
Fisioterapia	33,6	53,0	11,5	1,7
Serviço social	51,2	44,5	3,7	0,4
Odontologia	22,7	48,9	22,7	5,5
Fonoaudiologia	32,8	53,3	11,7	2,0

Medicina veterinária	28,2	48,6	18,6	4,4
Medicina	26,5	33,5	28,2	11,7
Terapia ocupacional	36,5	50,0	11,8	1,7

Fonte: MEC/INEP. Exame Nacional de Desempenho dos Estudantes (ENADE).

3 - Modelo de Regressão Logística

O Modelo de Regressão Logística Múltiplo é uma generalização do Modelo de Regressão Logística para o caso de mais de uma variável independente, também chamado de caso multivariado. Os coeficientes do modelo(parâmetro) são estimados pelo Método da Máxima Verossimilhança(MMV).

$$Y_i = \begin{cases} 1, & \text{se ocorreu a inversão,} \\ 0, & \text{se não ocorreu a inversão.} \end{cases}$$

Como Y é uma variável categórica binária, existem três modelos que relacionam Y com $(\mathbf{x}'\mathbf{s})$, são eles:

- Modelo de probabilidade linear (MPL);
- Modelo de regressão logística ou logit binário;
- Modelo probit binário.

O modelo de regressão logística(logit binário) tem um vantagem adicional: a interpretação dos parâmetros através da razão de chances, por isso, o modelo adotado neste trabalho para estimar a inversão será o modelo de regressão logística, que será detalhado nas seções seguintes.

O modelo de regressão logística é um caso particular de modelos lineares generalizados. Portanto, vamos iniciar esse capítulo com os MLG's.

3.1 – Modelos Lineares Generalizados

Os modelos lineares generalizados (MLG's) foram propostos por Nelder e Weddwebun (1972). Esta classe de modelos é definida por uma distribuição de probabilidade para a variável resposta, um conjunto de variáveis explicativas e uma função de ligação.

Definições:

- Componente aleatória: Composta de uma variável Y com n valores independentes (Y_1, \dots, Y_n) , um vetor de média $\boldsymbol{\mu} = \mathbf{E}(Y)$ é uma distribuição pertencente à família exponencial.

Pertencer à família exponencial significa ter uma função de probabilidade, $P(Y = y, \theta)$ no caso discreto ou função densidade de probabilidade, $f(y, \theta)$ no caso contínuo, escrita da seguinte forma:

$$P(Y = y, \theta) = e^{[a(y_i)b(\theta)+c(\theta)+d(y_i)]}.$$

$$f(y, \theta) = e^{[a(y_i)b(\theta)+c(\theta)+d(y_i)]}, \quad \forall_i = 1, 2, \dots, n$$

- Componente sistemática: Composta por variáveis explicativas x_1, x_2, \dots, x_p que geram um preditor linear η_i , dado por:

$$\eta_i = x_i^T \boldsymbol{\beta},$$

Em que $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1})^T$ é um vetor de parâmetros desconhecidos a serem estimados.

- Função de ligação: função monótona e diferenciável, denotada por $g(\mu_i)$, que relaciona a média da variável resposta $\mu_i = \mathbf{E}(Y_i)$ ao preditor η_i . Ou seja, $g(\mu_i) = \eta_i = x_i^T \boldsymbol{\beta}, \forall_i = 1, 2, \dots, n$.

Casos especiais de MLG's:

- Modelos de regressão linear normal;
- Modelo de Análise de Variância;
- Modelo de Análise de Covariância;
- Modelo de regressão logística ou logit binário;
- Modelo Log-linear de Poisson;
- Modelos para dados categóricos.

3.2 – Especificação do Modelo de Regressão Logística

Devemos considerar um conjunto de k variáveis, representadas pelo vetor $\mathbf{x}' = (x_1, x_2, \dots, x_k)$. Por agora vamos assumir que cada variável possui a mesma escala de intervalo. E que a probabilidade condicional que vai resultar é apresentada por $P(Y = 1|\mathbf{x}) = \pi(\mathbf{x})$. Então o valor logístico do Modelo de Regressão Logística Múltiplo é representado pela equação,

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_{p-1} x_{ki}, i = 1, 2, \dots, n \quad (3.1)$$

Descrevendo o modelo em função de $\pi(\mathbf{x})$ teremos,

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}} \quad (3.2)$$

Iremos fazer uso de variáveis dummy's. Alguns softwares geram estes tipo de variáveis automaticamente. Em geral, se uma variável em escala nominal possui n possíveis valores, então utiliza-se $n-1$ variáveis dummies. Suponha que a p -ésima variável independente, x_p tem n_p níveis. A variável $n_p - 1$ será representado como D_{pu} e os coeficientes para esta variável serão representados como $D_{ku}, u = 1, 2, \dots, p - 1$. Logo, o modelo logístico com k variáveis e p variáveis discretas será representado pela equação,

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_{1i} + \dots + \sum_{i=1}^{p-1} \beta_{pi} D_{pi} + \beta_k x_{ki} \quad (3.3)$$

Resumindo, o modelo de regressão logística múltiplo pode ser especificado por:

$$g(x_i) = \log \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.4)$$

3.3 - Estimação pelo método da máxima verossimilhança

Supondo que tenhamos uma amostra n de observações independentes do par $(x_i, y_i), i = 1, 2, \dots, n$. Como no caso univariado, aproximação do modelo requer que consigamos estimativas do vetor $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$. O método de estimação utilizado

no caso multivariado será o mesmo para o caso univariado, máxima verossimilhança. A função de verossimilhança é representada por,

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}, \quad (3.5)$$

a única diferença é que a função $\pi(\mathbf{x})$ será a que definimos. Vão existir k+1 funções de verossimilhança que são obtidas pela diferença da função de verossimilhança pelos seus respectivos k+1 coeficientes. A função de verossimilhança será representada por,

$$\sum_{i=1}^n [y_i - \pi(\mathbf{x}_i)] = 0 \quad (3.6)$$

e,

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{x}_i)] = 0, \quad \text{para } j=1,2,\dots,k. \quad (3.7)$$

Vamos considerar $\hat{\boldsymbol{\beta}}$ a solução para estas equações. Logo, os valores aproximados para o modelo de regressão logística múltipla são $\hat{\boldsymbol{\pi}}(\mathbf{x}_i)$, o valor que é expresso na função $\pi(\mathbf{x})$ computado usando $\hat{\boldsymbol{\beta}}$ e \mathbf{x}_i .

O método de estimação de variâncias e covariâncias dos coeficientes estimados seguem de um teoria bem desenvolvida de estimação de máxima verossimilhança. Esta teoria afirma que os estimadores são obtidos de uma matriz das segundas derivadas parciais da função de máxima verossimilhança logarítmica. Estas derivadas parciais tem a seguinte forma generalizada,

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i), \quad (3.8)$$

e,

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_u} = - \sum_{i=1}^n x_{ij} x_{iu} \pi_i (1 - \pi_i), \quad (3.9)$$

para $j,u=0,1,2,\dots,k$ aonde π_i é $\pi(\mathbf{x}_i)$. Seja a matriz (k+1) por (k+1) que contém os negativos dos termos dados nas equações acima serem denominados como $\mathbf{I}(\boldsymbol{\beta})$. Esta matriz é chamada de matriz informativa. As variâncias e covariâncias dos coeficientes estimados são obtidas da inversa desta matriz, que será denominada como $\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \mathbf{I}^{-1}(\boldsymbol{\beta})$. Apenas em casos muito especiais que não é possível escrever uma

expressão explícita dos elementos dessa matriz. A partir daqui, iremos utilizar a notação $\sigma^2(\beta_j)$ para denotar o p-ésimo elemento da diagonal principal da matriz, que é a variância de $\hat{\beta}_j$, e $\sigma(\beta_j, \beta_u)$ para denotar um elemento arbitrário que não pertença a diagonal, o qual é a covariância de $\hat{\beta}_j$ e $\hat{\beta}_u$. Os estimadores de variância e covariância, serão denotados por $\hat{\Sigma}(\hat{\beta})$, e são obtidos quando avalia-se $\Sigma(\beta)$ em $\hat{\beta}$. Utilizaremos $\hat{\sigma}^2(\hat{\beta}_j)$ e $\hat{\sigma}(\hat{\beta}_j, \hat{\beta}_u)$, $j, u = 0, 1, 2, \dots, k$, para os valores da matriz. Para a maioria das vezes terá a oportunidade de usar somente o estimador de erro padrão dos coeficientes estimados, que é representado pela equação,

$$\widehat{SE}(\hat{\beta}_j) = [\hat{\sigma}^2(\hat{\beta}_j)]^{1/2} \quad (3.10)$$

para $j = 0, 1, 2, \dots, k$. Será utilizado essa notação para métodos de desenvolvimento para testes de coeficiente e estimação do intervalo de confiança.

A matriz informativa é útil para o modelo de aproximação e avaliação da aproximação, é representado por $\hat{\mathbf{I}}(\hat{\beta}) = \mathbf{X}'\mathbf{V}\mathbf{X}$, aonde \mathbf{X} é um n da matriz $k + 1$ que contém o dado de cada indivíduo, e \mathbf{V} é um n por n da diagonal da matriz que com elementos gerais $\hat{\pi}_i(1 - \hat{\pi}_i)$. A matriz \mathbf{X} é,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \quad (3.11)$$

e a matriz \mathbf{V} é,

$$\mathbf{V} = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix} \quad (3.12)$$

3.4 - Interpretação dos parâmetros estimados

A interpretação do modelo ajustado requer que seja possível visualizar inferências práticas dos coeficientes estimados no modelo. Deve-se pensar em que os

coeficientes estimados no modelo estão informando sobre a motivação do estudo. Para a maioria dos modelos isso irá envolver os coeficientes estimados das variáveis independentes do modelo. Em certos casos, o coeficiente β_0 será nosso interesse, mas isso é exceção a regra. Os coeficientes estimados para as variáveis independentes representam a inclinação ou a razão de chances, que será definida a depois, da função das variáveis dependentes por unidade de chance nas variáveis independentes.

O primeiro passo é determinar qual função das campo das variáveis dependentes uma função linear das variáveis independentes. O que geralmente é chamado de função link. No modelo de regressão logística a função link é a transformação da logit,

$$g(x) = \ln\left\{\frac{\pi(x)}{[1-\pi(x)]}\right\} = \beta_0 + \beta_1 x \quad (3.13)$$

No modelo de regressão múltipla utilizamos β_1 como $\beta_1 = g(x + 1) - g(x)$. Este é, o coeficiente angular representa a chance no logit para a mudança de uma unidade na variável independente x . A interpretação adequada dos coeficientes no modelo de regressão logística depende em ser capaz de adaptar os significados entre dois logits diferentes.

Considerando variáveis independentes dicotômicas para interpretar os coeficientes do modelo de regressão logístico. Temos que x é codificado com zero ou 1. Sob este modelo existem dois valores de $\pi(x)$ e equivalentemente dois valores para $1 - \pi(x)$. Estes valores serão disposto em 2×2 como esta na tabela,

		<i>Variável Independente X</i>	
		$x = 1$	$x = 0$
<i>Variável Saída Y</i>	$y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
	$y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$

As chances da saída estar presente entre indivíduos com $x = 1$ é definida como $\pi(1)/[1 - \pi(1)]$. Da mesma forma, as chances da saída estar presente entre indivíduos com $x = 0$ é definida por $\pi(0)/[1 - \pi(0)]$. O log das chances, é chamado o logit, e nesse caso eles são,

$$g(1) = \ln \left\{ \frac{\pi(1)}{1 - \pi(1)} \right\} \quad (3.14)$$

e,

$$g(0) = \ln \left\{ \frac{\pi(0)}{1 - \pi(0)} \right\} \quad (3.15)$$

A razão de chances, denotada por ψ , definida como a razão de chances para $x = 1$ em relação as chances para $x = 0$, e é dada pela equação,

$$\psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \quad (3.16)$$

O log da razão de chances, chamado log de razão de chances, ou log de chances, é

$$\begin{aligned} \ln(\psi) &= \ln \left[\frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \right] \\ &= g(1) - g(0) \end{aligned} \quad (3.17)$$

que é a diferença logit.

Agora, usando a expressão para o modelo de regressão logístico mostrado na tabela a razão de chance é,

$$\psi = \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right) \left(\frac{1}{1 + e^{\beta_0}} \right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}} \right) \left(\frac{1}{1 + e^{\beta_0 + \beta_1}} \right)} \quad (3.18)$$

$$= \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

por isso, para regressão logística com variáveis independentes dicotômicas

$$\psi = e^{\beta_1}.$$

3.5 - Inferência do modelo

A partir do momento que foi ajustado um Modelo de Regressão Logística Múltiplo particular, pode-se iniciar o processo de avaliação do modelo. Como no caso univariado, o primeiro passo para o processo usualmente é, avaliar a significância das variáveis do modelo. O teste de razão de chances, para todas as significâncias dos k coeficientes, para as variáveis independentes no modelo é usado, exatamente, da mesma maneira que no caso univariado. Este teste é baseado na estatística G , que é representada por,

$$G = -2 \ln \left[\frac{(\text{verossimilhança sem variáveis})}{(\text{verossimilhança com as variáveis})} \right] \quad (3.19)$$

a única diferença é que usa-se o valor ajustado, \hat{n} , sob os modelos serem baseados no vetor contendo $k+1$ parâmetros, $\hat{\beta}$. Sob a hipótese nula que os k coeficientes estimados para as covariâncias do modelo são iguais a zero, a distribuição de G será um χ^2 com k graus de liberdade.

O teste de Wald, para o caso univariado, é obtido da estimação de máximo verossimilhança do parâmetro estimado, $\hat{\beta}_k$, para uma estimação do erro padrão. O resultado da razão, sob a hipótese nula, seguirá a distribuição normal padrão. As hipóteses para o teste de Wald são:

$$\begin{cases} H_0: \beta_k = 0 \\ H_1: \beta_k \neq 0 \end{cases}$$

A estatística de teste de Wald é representado por,

$$W = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} \quad (3.20)$$

onde \widehat{SE} é o erro padrão. Para o caso multivariado, analogamente teste é obtido pelo calculo da seguinte matriz,

$$\begin{aligned} W &= \widehat{\beta}' [\Sigma(\widehat{\beta})]^{-1} \widehat{\beta} \\ &= \widehat{\beta}' (\mathbf{X}' \mathbf{V} \mathbf{X}) \widehat{\beta} \end{aligned} \quad (3.21)$$

que será distribuída como uma qui-quadrado com k+1 graus de liberdade, sob a hipótese que cada um dos k+1 coeficientes são iguais a zero. Teste para apenas o k coeficientes estimados são obtidos por eliminação do $\hat{\beta}_0$, que vem do $\widehat{\beta}$, e a primeira linha e coluna de $(\mathbf{X}' \mathbf{V} \mathbf{X})$.

A razão de chances, ψ , é normalmente o parâmetro de interesse em uma regressão logística adequado, pois é fácil interpreta-lo. Contudo, sua estimativa, $\hat{\psi}$, irá tender para ter um distribuição assimétrico. A assimetria de uma amostra da distribuição de $\hat{\psi}$ é devido ao fato que, isto é afastado de zero. Na teoria, para amostras de tamanhos grandes, a distribuição de $\hat{\psi}$ será normal. As inferências sobre esse parâmetro é, usualmente, baseada em uma distribuição de amostragem de $\ln(\hat{\psi}) = \hat{\beta}_1$, a qual tende a seguir uma distribuição normal para a maioria das amostra de tamanho pequeno. Uma estimativa para o intervalo de confiança com $100 \times (1 - \alpha)\%$ de razão de chances é obtida por, primeiramente calculo dos pontos de extremidade do intervalo de confiança do coeficiente $\hat{\beta}_1$ e depois, a aplicação da exponencial desses valores. Em geral, os pontos extremos são dados pela expressão,

$$\exp \left[\hat{\beta}_1 \pm z_{1-\alpha/2} \times \widehat{SE}(\hat{\beta}_1) \right] \quad (3.22)$$

Pela importância da razão de chances como uma medida de associação, estimativas pontuais e intervalares são encontradas nas colunas das tabelas apresentando o resultado da análise da regressão logística.

3.6 - Qualidade do Ajuste

Avaliar o ajuste do modelo envolve a análise do ajuste para observações individuais e a avaliação de medidas escalares de ajuste para o modelo como um todo. As medidas de ajuste

podem fornecer uma índice bruto sobre se o modelo é adequado. As medidas de ajuste fornecem algumas informações, contudo são apenas algumas informações parciais que devem ser avaliadas dentro do contexto da teoria para motivar a análise, como investigação do passado, e os parâmetros estimados do modelo considerado. Existem diversos métodos para avaliar a qualidade o ajuste, dentre eles temos o R^2 .

Temos que os valores esperados e observados podemos ser usados no modelo como categorias de saída para computar o R^2 . Considerando o caso binário aonde a observação y é 0 ou 1 e $\pi_i = \overline{\text{Pr}}(y = 1|x_i)$. Definamos a saída esperada como,

$$\hat{y}_i = \begin{cases} 0 & \text{se } \hat{\pi}_i \leq 0.5, \\ 1 & \text{se } \hat{\pi}_i \geq 0.5 \end{cases}$$

Isto permite a construção de uma tabela de valores esperados e observados. E, partir dessa tabela, pode-se observar que avaliar quantas respostas estimadas forem corretas e quantas estão incorretas. Também pode-se visualizar os falsos positivos e falsos negativos, que são, respectivamente, os resultados que deram positivos, mas na verdade são negativos e os resultados negativos, que são positivos. Assim podemos avaliar a precisão do nosso modelo.

Uma medida que é importante é a proporção de acertos, ou seja, proporção de previsões corretas, Hosmer e Lemeshow utilizam um teste de qui-quadrado onde n_{jj} são os números de previsões corretas para j . As hipóteses são,

$(H_0: \text{ o modelo estimado ser adequado aos dados}$
 $(H_1: \text{ o modelo estimado não ser adequado as dados}$

Estatística de teste é:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3.23)$$

onde O_i é representa o número de elementos de $Y_i=1$ e $Y_i=0$ em casa grupo, e E_i é obtido a partir das médias das probabilidades estimadas de $Y_i=1$ e $Y_i=0$ em grupo i . Se a estatística do teste for maior ou igual ao valor crítico de qui-quadrado, rejeita-se a hipótese nula ao nível de significância α , ou seja, o modelo não se ajusta aos dados observados.

3.7 -Técnicas de Diagnóstico

Se for definido a probabilidade prevista para um conjunto de variáveis independentes como,

$$\pi_i = \Pr(y_i = 1|x_i),$$

então os desvios $y_i - \pi_i$ possuem heterocedasticidade, com

$$\text{Var}(y_i - \pi_i|x_i) = \pi_i(1 - \pi_i),$$

isto implica que a variância de um resultado binário é maior que $\pi_i - \pi_i$, ao menos que π_i se aproxima de 0 ou 1. Em outras palavras, existe heterocedasticidade que depende da probabilidade de um resultado positivo. Este é resolvido pelo resíduo de Pearson que divide os resíduos $y - \hat{\pi}$, a partir do desvio padrão,

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}},$$

grandes valores para r sugerem uma falha do modelo para atender uma determinada observação. Sabendo que a variância de r não é 1, já que $\text{Var}(y_i - \hat{\pi}_i) \neq \hat{\pi}_i(1 - \hat{\pi}_i)$, iremos utilizar os resíduos de Pearson padronizados,

$$r_i^{\text{std}} = \frac{r_i}{\sqrt{1 - h_{ii}}},$$

aonde,

$$h_{ii} = \hat{\pi}_i(1 - \hat{\pi}_i) x_i \widetilde{\text{Var}}(\hat{\beta}) x_i',$$

enquanto r^{std} é preferido em relação a r devido a sua variância constante, descobrimos que na prática eles são muito similares.

4 – Resultados

4.1 – A base de dados do ENADE

O Exame Nacional de Desempenho do Aluno(ENADE), que é realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira(Inep) , avalia determinados cursos anualmente. Nesse trabalho os bancos de dados referentes a 2008,2009,2010, serão utilizados. A escolha desses anos se justifica pela mudança na metodologia que ocorreu, a partir de 2011,que não coleta informações dos ingressantes. Outra justificativa do uso desses 3 anos se deve à informações de cursos distintos, o que possibilita uma visão para ampla de todo o quadro de cursos possíveis. Os bancos estão disponíveis no site do Inep.

4.2 – Descrição das Variáveis

As variáveis que foram selecionadas para os anos de 2008,2009 e 2010, que serão utilizadas em nosso modelo são:

- Gênero; (1 se Feminino e 0 se Masculino);
- Idade, agrupada em 3 categorias: 18 a 24, 25 a 29 e outro;
- Renda Familiar, agrupada em 3 categorias: renda até 3 s.m, renda entre 3 e 10 s.m e renda entre 10 e 30 s.m;
- Estado Civil, agrupada em 3 grupos: solteiro, casado e outro;
- Cor/Raça, agrupado em 3 grupos: Branco, Negro/Pardo/Mulato e outro;
- Educação do Pai, agrupado em 4 grupos: nenhum estudo, ensino médio, ensino fundamental e ensino superior;
- Educação da Mãe, agrupado em 4 grupos: nenhum estudo, ensino médio, ensino fundamental e ensino superior;
- Região geográfica, dividido em 3 regiões: Norte, Sudeste e outro.

Estas variáveis foram selecionadas por serem as características socioeconômicas que melhor representam as diferenças sociais, e eles devem ser as que mais influenciam e melhor justificam na questão que estamos interessados. Para a variável idade estamos

considerando se o ingressante na IES possui a idade considerada ideal para ingressar no ensino superior, que é entre 18 e 24 anos, se tem idade entre 25 e 29 anos, e por último se possui idade superior a 29 anos. Para a variável renda familiar dividimos os grupos pela quantidade de salários mínimos que a família ganha. As variáveis educação do pai e da mãe, verificam até que grau de escolaridade estes possuem. A variável região geográfica divide os ingressantes de acordo com a região do país. E a variável estado civil, vamos considerar se o indivíduo é solteiro ou casado.

4.3 – Estatísticas Descritivas

A análise descritiva consiste em extrairmos informações do nosso banco de dados, antes de realmente estimarmos os parâmetros dos modelos. Essa análise preliminar, possibilitará já ter uma visão de para aonde as nossas informações estão nos encaminhando e o que podemos esperar quando estimarmos nossos parâmetros.

Na Tabela 9 podemos analisar as características dos ingressantes na IES pública e privada como um todo. As observações estão organizadas de acordo com os cursos que estamos interessados. Primeiramente, podemos ver que o sexo feminino prevalece somente nos cursos de Medicina e Odontologia. Para os cursos de Matemática, Engenharia e Economia, temos mais homens. Sendo que no curso de Engenharia o número de homens é muito superior ao número de mulheres.

Também podemos verificar que, com relação a idade, os cursos de Medicina e Odontologia possuem um percentual maior de pessoas na faixa etária considerada ideal para o ingresso na IES. Enquanto estes cursos tem um percentual por volta de 80% de pessoas na faixa ideal, os outros cursos são inferiores a 70%, nesta categoria. Para o curso de Matemática, temos, dentre esses cursos, o maior percentual de pessoas ingressando na IES, num faixa etária superior a 29 anos.

Já para a renda temos um maior percentual, na faixa de 3 a 10 salários mínimos para os cursos de Medicina e Odontologia, seguido de Economia. O curso de Engenharia se destaca na faixa de 10 a 30 salários mínimos. Chegando a ser maior que medicina, entre 10 e 30 s.m, em mais que duas vezes.

Com relação ao nível escolar do pai, temos que o menor percentual na categoria de nível superior é no curso de Matemática. E os maiores são Medicina e Odontologia. Matemática é representa o maior percentual também, na categoria de nenhuma

escolaridade e nível fundamental. Para a escolaridade da mãe, verificamos que o maior percentual de nenhum grau de instrução é para o curso de Engenharia. Para o fundamental se destaca o percentual de Matemática. E, por último, o nível superior é maior em Medicina e Odontologia.

Para a região geográfica, podemos observar que os cursos de Medicina, Odontologia e Engenharia possuem maior número de ingressantes na IES, no Sudeste. Sendo, praticamente, igual a quantidade de ingressantes no Sudeste as outras regiões do país, desconsiderando o Norte, para esses cursos. Sobre a característica de cor/raça, observamos um maior percentual da cor negra/parda/mulata nos curso de Matemática, comparado aos outros cursos, nesta categoria. A cor branca possui um percentual superior em todos os cursos.

Por último, o estado civil dos ingressantes esta com o percentual maior, em todos os cursos, para a categoria solteiro. Porém, dentre esses cursos, Matemática se destaca com maior percentual de casados. E o percentual de solteiros e maior nos cursos de Medicina e Odontologia

Tabela 9 – Análise descritiva do perfil dos ingressantes por curso selecionado, 2008 – 2010, Brasil.

Variáveis	% de Ingressantes					
	Engenharia	Matemática	Relações I.	Economia	Medicina	Odontologia
Sexo	100,0	100,0	100,0	100,0	100,0	100,0
Masculino	77,8	54,8	42,2	62,7	45,6	32,8
Feminino	22,2	45,2	57,8	37,3	54,4	67,2
Idade	100,0	100,0	100,0	100,0	100,0	100,0
Entre 18-24	69,1	50,8	85,8	72,8	82,4	81,8
Entre 25-29	14,9	17,3	7,3	14,3	12,4	9,5
Outro	16,0	31,9	7,2	12,9	5,2	8,7
Renda Familiar	100,0	100,0	100,0	100,0	100,0	100,0
Até 3 s.m	24,4	62,7	23,4	35,4	27,1	29,3
Entre 3 e 10 s.m	47,6	33,2	64,3	58,2	61,2	66,4
Entre 10 e 30 s.m	28,0	4,1	12,3	6,3	11,7	4,2
Escolaridade do Pai	100,0	100,0	100,0	100,0	100,0	100,0
Nenhuma	2,2	8,7	1,4	3,7	0,6	1,5
Fundamental	32,4	58,7	14,5	32,0	12,5	25,4
Médio	33,4	24,5	30,3	33,7	27,4	36,1
Superior	32,0	8,1	53,8	30,6	59,5	37,1
Escolaridade da Mãe	100,0	100,0	100,0	100,0	100,0	100,0
Nenhuma	22,0	6,7	1,0	2,1	0,4	0,9
Fundamental	28,0	54,3	11,6	29,0	7,6	18,1

Médio	35,6	26,9	30,6	36,4	26,4	35
Superior	34,4	12,1	56,8	32,0	65,6	46
Região	100,0	100,0	100,0	100,0	100,0	100,0
Norte	6,0	9,7	0,0	7,0	12,3	13,4
Sudeste	47,3	37,6	0,0	41,8	47,6	42,4
Outros	46,7	52,7	100,0	51,2	40,1	44,2
Cor	100,0	100,0	100,0	100,0	100,0	100,0
Negro/Pardo/Mulato	28,8	41,7	20,1	29,8	25,7	26,7
Branco	68,6	55,8	76,7	67,1	71,7	70,8
Outros	2,6	2,5	3,2	3,1	3,2	2,5
Estado Civil	100,0	100,0	100,0	100,0	100,0	100,0
Solteiro	84,1	74,4	95,8	90,7	95,5	91,2
Casado	12,7	19,8	2,4	6,6	3,3	6,6
Outro	3,2	5,8	1,8	2,7	1,2	2,2

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).
Tabulação do Autor.

Na Tabela 10, são apresentadas informações dos dois tipos de inversões, por curso. E também os dados para o caso de não ter nenhum dos tipos de inversão. Para o curso de Engenharia o primeiro tipo de inversão não ocorreu tão forte, praticamente tendo o mesmo percentual para o caso de não ter um inversão do primeiro tipo. Já para a inversão 2, neste curso, temos ela aparecendo de forma bem evidente.

Para o curso de Matemática, observa-se que a inversão 1 foi forte, chegando a quase 70% dos ingressantes terem saído de um ensino médio privado e ingressa em uma IES pública. Já para a inversão 2 não vemos grandes diferenças. Os cursos de Economia e Medicina possuem o primeiro tipo de inversão com um resultado próximo a 60%.

Tabela 10 – Inversão dos ingressantes do ensino superior, por curso, Brasil.

Curso/Dep. Adm. IES	Ensino Médio	
	Todo/Maior parte em Escola Privada	Todo/Maior parte em Escola Pública
Engenharia	100,0	100,0
IES Pública	49,8 ¹	31,7
IES Privada	50,2	68,3 ²
Matemática	100,0	100,0
IES Pública	68,6 ¹	59,2
IES Privada	31,4	40,8 ²
Relações Internacionais	100,0	100,0
IES Pública	15,5 ¹	13,2
IES Privada	84,5	86,8 ²

Economia	100,0	100,0
IES Pública	60,0 ¹	58,4
IES Privada	40,0	41,6 ²
Medicina	100,0	100,0
IES Pública	63,4 ¹	48,6
IES Privada	36,6	51,4 ²
Odontologia	100,0	100,0
IES Pública	35,4 ¹	24,1
IES Privada	75,9	75,9 ²

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).
 Tabulação do Autor. Nota: 1 faz referência a inversão 1 e 2 a inversão do tipo 2.

A Tabela 11, nos oferece informações sobre a inversão 1 e 2 por faixa etária, em cada curso. Para o curso de Engenharia, temos que a primeira inversão foi mais forte entre a faixa etária entre 18 a 24. Para as outras faixas, teve um percentual menor. Na inversão 2, observamos ela um pouco mais evidente nas faixas de 25 a 29 e outro. No curso de Matemática observamos o mesmo comportamento.

Para o curso de Medicina verificamos que o primeiro tipo de inversão são praticamente iguais os percentuais, independente da faixa etária, comparando a ocorrência ou não desse tipo de inversão. Já para a inversão 2, verificamos que ela possui um percentual maior entre as faixas etárias de 18 a 24 e 25 a 29. No curso de Odontologia as inversões possuem o percentual praticamente iguais, comparadas com a ocorrência ou não da respectiva inversão.

Tabela 11 – Inversão por Idade, em cada curso.

Idade	Ensino Médio	
	Engenharia	
	Todo/Maior parte em Escola Privada	Todo/Maior parte em Escola Pública
IES Pública	100,0	100,0
18-24	90,3	75,3
25-29	6,4	14,3
Outro	3,3	10,4
IES Privada	100,0	100,0
18-24	78,7	56,6
25-29	11,0	19,0
Outro	10,2	24,4
Matemática		
IES Pública	100,0	100,0
18-24	62,5	57,6
25-29	10,7	14,6

Outro	26,8	27,8
IES Privada	100,0	100,0
18-24	43,5	48,8
25-29	14,5	16,5
Outro	42,3	34,7
Economia		
IES Pública	100,0	100,0
18-24	88,7	73,1
25-29	6,5	14,5
Outro	4,9	12,4
IES Privada	100,0	100,0
18-24	88,4	73,0
25-29	5,2	16,5
Outro	6,3	10,4
Relações Internacionais		
IES Pública	100,0	100,0
18-24	88,4	73,1
25-29	5,2	14,5
Outro	6,3	12,4
IES Privada	100,0	100,0
18-24	88,6	73,0
25-29	6,5	16,5
Outro	4,9	10,4
Medicina		
IES Pública	100,0	100,0
18-24	85,6	65
25-29	10,8	2,3
Outro	4,1	13,7
IES Privada	100,0	100,0
18-24	89,9	75,0
25-29	8,6	17,0
Outro	1,6	8
Odontologia		
IES Pública	100,0	100,0
18-24	96,1	89,9
25-29	2,6	7,2
Outro	1,3	2,9
IES Privada	100,0	100,0
18-24	89,5	71,1
25-29	5,8	13,8
Outro	4,8	15,1

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).
 Tabulação do Autor.

4.4 – Modelos Estimados

Modelo 1 – Inversão 1: ensino médio privado para IES pública

O modelo de regressão logística foi estimado dos microdados dos ingressantes dos cursos de Engenharia, Matemática, Relação Internacionais, Economia, Medicina e Odontologia. As estimativas apresentados foram os mais bem ajustados para os resíduos, com o corte de mais ou menos 2, que será comentado em 4.5.

Na Tabela 12 temos os resultados gerados no modelo de regressão logística para o primeiro modelo, inversão 1. Logo, todos os resultados estão nos direcionando ao entendimento das variáveis que influenciam e como elas causam impacto nessa primeira inversão. Essas variáveis dobram a chances do ingressando sofrer esse tipo de inversão. Já ser do sudeste e norte, diminuem essa chance. E, por último, ter renda familiar entre 3 e 10 salários mínimos também reduz as chances.

Para o curso de Matemática temos que se o ingressante dor do sexo feminino, possui menos chance de ter vindo de um ensino médio privado e ingressado em uma IES pública. Já ter idade ideal, nível superior de instrução do pai e da mãe, praticamente dobram as chances disso acontecer. E ser solteiro quadruplicam as chances.

No curso de Economia temos que, ter a idade ideal duplica as chances de sofrer a inversão 1. Já ter renda familiar até 3 salários mínimos reduz as chances, ou seja, ter renda superior a 3 salários mínimos aumenta as chances do ingressante realizar a primeira inversão. Também temos que ser do sudeste aumenta a chances dessa inversão.

No curso de Medicina temos que ser do sexo feminino aumenta em quase 50% as chances da primeira inversão. Ter a idade ideal reduz um pouco isto. E ter a renda familiar entre 3 e 10 salários mínimos aumenta a 1 inversão. Ter renda entre 10 e 30 s.m. quadruplica as chances e ser do sudeste aumenta em seis vezes a chance de inversão. Por, último ser da cor branca aumenta em quase 55% as chances.

Em Odontologia percebe-se que ter a idade ideal aumenta em duas vezes as chances de sofrer essa inversão. Já, possuir renda entre 3 e 10 s.m, renda entre 10 e 30 s.m e ser da região norte diminuem.

Tabela 12 - Estimativa dos parâmetros referentes ao modelo logístico.

Variáveis	Engenharia		Matemática	
	p-valor	exp(Beta)	p-valor	exp(Beta)
Feminino	<0,01	1,417	<0,01	,461
Idadecerta18a24	<0,01	2,136	<0,01	2,067
Renda3a10	*	,866	*	,899
Renda10a30	<0,01	,586	*	1,122
EducPaiSup	,007	1,252	<0,01	2,085
EducMaeSup	*	1,172	<0,01	1,985
Sudeste	<0,01	,380	<0,01	,147
Norte	<0,01	,347	,039	3,255
Branco	,020	,819	<0,01	,472
Solteiro	<0,01	2,641	<0,01	3,991
Constante	<0,01	,386	<0,01	2,917
Variáveis	Relações Internacionais		Economia	
Feminino	,013	,550	,033	1,234
Idadecerta18a24	*	1,799	<0,01	1,824
Renda3a10	<0,01	,038	<0,01	,296
Renda10a30	*	,000	<0,01	,117
EducPaiSup	*	,909	*	,905
EducMaeSup	<0,01	3,478	*	1,146
Sudeste	<0,01	,157	<0,01	,122
Norte			,003	8,740
Branco			<0,01	,487
Solteiro	*	1,245	*	1,004
Constante	*	,566	<0,01	15,962
Variáveis	Medicina		Odontologia	
Feminino	<0,01	1,589	,030	1,159
Idadecerta18a24	<0,01	,474	<0,01	2,638
Renda3a10	<0,01	1,496	<0,01	,589
Renda10a30	<0,01	4,390	<0,01	,241
EducPaiSup	*	,912	*	,981
EducMaeSup	<0,01	1,361	,002	1,222
Sudeste	<0,01	6,304	,030	,870
Norte	*	,884	<0,01	,428
Branco	<0,01	1,548	*	1,007
Solteiro	<0,01	,380	<0,01	73,890
Constante	,028	1,480	<0,01	,004

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).
 Tabulação do Autor. Nota: * não significante ao nível de significância de 5%.

Modelo 2 – Inversão 2: ensino médio público para IES privada

A Tabela 13 nos fornece os dados estimados e as razões de chance por curso, do segundo tipo de inversão. Então, poderemos analisar, por curso, quais variáveis causam

impacto e como esse impacto influência na nossa inversão 2. Temos para o curso de Engenharia que ser do sexo feminino, ter idade certa e nível de escolaridade do pai diminuem as chances dessa inversão. Já ser da região Sudeste aumenta em oito vezes as chances e ser branco em quase duas vezes.

Para o curso de Matemática temos que, ser do sexo feminino e ter renda familiar entre 3 e 10 salários mínimos aumentam as chances dessa inversão. Ser branco aumenta em duas vezes as chances. Mas, ser solteiro diminuem as chances. No curso de economia temos que estar na idade ideal, ter o nível superior de escolaridade do pai e da mãe diminuem as chances. Já, ter renda entre 10 e 30 s.m. aumenta em seis vezes as chances, ser da região sudeste aumentam em nove vezes as chances e ser branco aumentam em duas vezes.

Para Medicina verificamos que, ser da região Sudeste e ser branco diminuem as chances da inversão. Porém, estar na idade ideal aumenta as chances. E, ser solteiro aumenta em duas vezes as chances. Por último, o curso de Odontologia temos, que estar com a idade ideal diminuem as chances. Mas, ser da região Sudeste aumentam em quatro vezes as chances, ser da região norte em duas vezes e ser branco em quase 75% das vezes.

Tabela 13 – Estimativa dos parâmetros referentes ao modelo logístico.

Variáveis	Engenharia		Matemática	
	p-valor	exp(Beta)	p-valor	exp(Beta)
Feminino	<0,01	,608	<0,01	1,454
Idadecerta18a24	<0,01	,602	,002	,812
Renda3a10	,014	1,277	<0,01	1,355
Renda10a30	*	,803	*	,843
EducPaiSup	<0,01	,465	*	,661
EducMaeSup	*	,805	,042	,785
Sudeste	<0,01	8,483	<0,01	5,107
Norte	,004	1,706	*	,856
Branco	<0,01	1,728	<0,01	1,810
Solteiro	<0,01	,467	<0,01	,639
Constante	<0,01	2,213	<0,01	,337
Variáveis	Relações Internacionais		Economia	
Feminino	,267	*	*	1,173
Idadecerta18a24	,682	*	<0,01	,685
Renda3a10	,142	*	<0,01	1,781
Renda10a30	,673	*	<0,01	6,614
EducPaiSup	,992	*	<0,01	,413
EducMaeSup	,003	,156	<0,01	,601

Sudeste	,189	*	<0,01	9,177
Norte			*	,999
Branco			<0,01	2,439
Solteiro	,621	*	*	,865
Constante	,991	*	<0,01	,185
Variáveis	Medicina		Odontologia	
Feminino	*	,882	*	1,011
Idadecerta18a24	<0,01	1,468	<0,01	,192
Renda3a10	,017	1,252	<0,01	2,040
Renda10a30	,004	,341	*	552702599,461
EducPaiSup	*	1,062	*	,810
EducMaeSup	,016	1,271	*	1,111
Sudeste	<0,01	,217	<0,01	4,628
Norte	,006	,689	<0,01	2,885
Branco	<0,01	,574	<0,01	1,742
Solteiro	<0,01	2,028	*	,000
Constante	*	1,037	*	549308917,103

Fonte: MEC/INEP. Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE).
 Tabulação do Autor. Nota: * não significante ao nível de significância de 5%.

4.5 – Qualidade do Ajuste

Modelo 1 – Inversão 1

Pela Tabela 14 podemos verificar que pelo teste de qui-quadrado, o modelo foi bem ajustado para todos os cursos. Utilizando 8 graus de liberdade e 5% de significância, rejeita-se a hipótese nula de que o modelo se ajusta aos dados para os cursos de Engenharia, Matemática, Relações Internacionais e Medicina.

Tabela 14 – Qualidade do Ajuste

Estatísticas	Engenharia	Matemática	Relações I.	Economia	Medicina	Odontologia
$-2 \ln L_{Final}$	4161,094	920,666	492,61	3460,019	10675,75	6094,138
Qui-Quadrado	366,421	294,954	302,730	590,200	2196,110	366,199
gl	10	10	10	10	10	10
p-valor	< 0,01	< 0,01	< 0,01	< 0,01	< 0,01	< 0,01
Hosmer, Lemeshow Qui-Quadrado	20,125	32,478	32,967	7,064	286,571	12,174

Pela Tabela 15, observamos que os cursos de Matemática e Medicina foram os que o modelo acertou mais as estimativas, para Y=1. E o pior foi para o curso de Odontologia, para Y=1.

Tabela 15 – Tabela de Classificação

Estatísticas	2008		2009		2010	
	Engenharia	Matemática	R.I	Economia	Medicina	Odontologia
Y= 0 estimado corretamente	61,8	44,4	98,4	81,6	53,5	94,4
Y= 1 estimado corretamente	62,5	89,8	19,1	57,6	80,9	10,4
Y estimado total	62,1	77,1	92,8	71,5	71,6	65,2

Modelo 2 – Inversão 2

Verifica-se na Tabela 16 que, todos os cursos foram bem ajustados para o qui-quadrado, mas o curso de Matemática e Medicina não passaram pelo Hosmer-Lemeshow. Isto pode dificultar na interpretação do modelo. Utilizamos 8 graus de liberdade.

Tabela 16 – Qualidade do Ajuste.

Estatísticas	Engenharia	Matemática	Relações I.	Economia	Medicina	Odontologia
2L Final	2995,823	5927,215	134,857	3110,047	2827,319	2719,293
Qui-Quadrado	784,870	958,016	788,653	788,653	367,699	665,827
Gl	10,0	10	10	10	10	10
p-valor	< 0,01	< 0,01	< 0,01	< 0,01	< 0,01	< 0,01
Hosmer,Lemeshow Qui-Quadrado	124,602	21,122	57,790	57,790	11,174	96,838

Pela Tabela 17 temos, que os cursos com maior acerto de Y=1 são o de Engenharia e Odontologia.

Tabela 17 – Tabela de classificação.

Estatísticas	2008		2009		2010	
	Engenharia	Matemática	R.I.	Economia	Medicina	Odontologia
Y= 0 estimado corretamente	40,2	81,9	3,4	82,5	64,7	14,6
Y= 1 estimado corretamente	87,6	52,9	99,8	60,7	68,4	96,2
Y estimado total	73,9	70,1	95,2	73,8	66,6	79,8

4.6 – Técnicas de Ajuste

Observando o gráfico de resíduos dos 2 modelos para todos os cursos, definimos o ponto de corte nos valores maiores que 2 ou menores que -2. E, com isso apresentamos os modelos melhores ajustados.

Figura 1- Gráficos dos resíduos de Engenharia, Inversão 1 e 2.

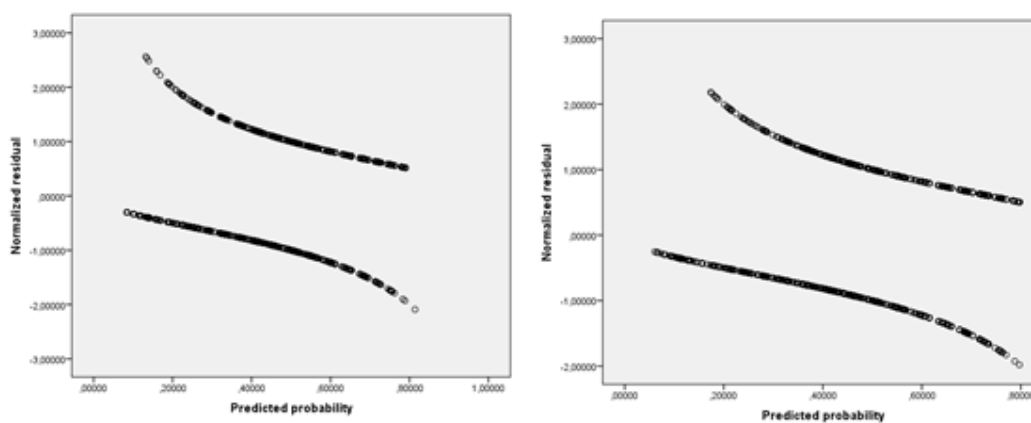


Figura 2 – Gráficos dos resíduos de Matemática, Inversão 1 e 2.

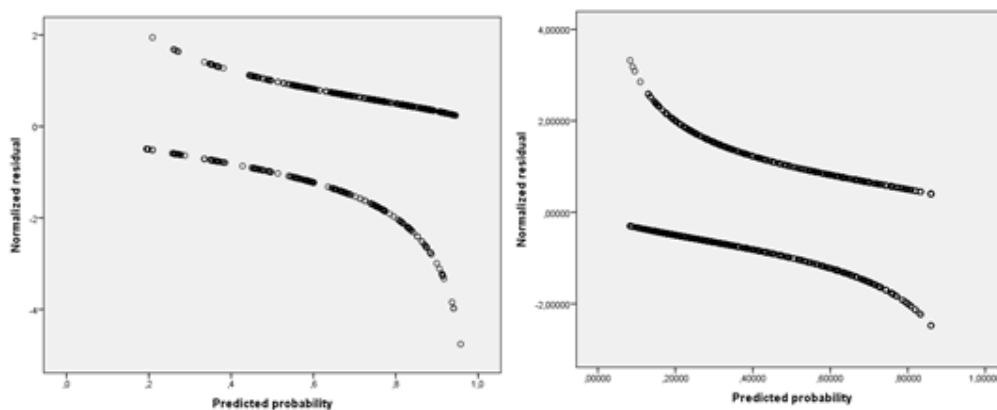


Figura 3 - Gráficos dos resíduos de Economia, Inversão 1 e 2.

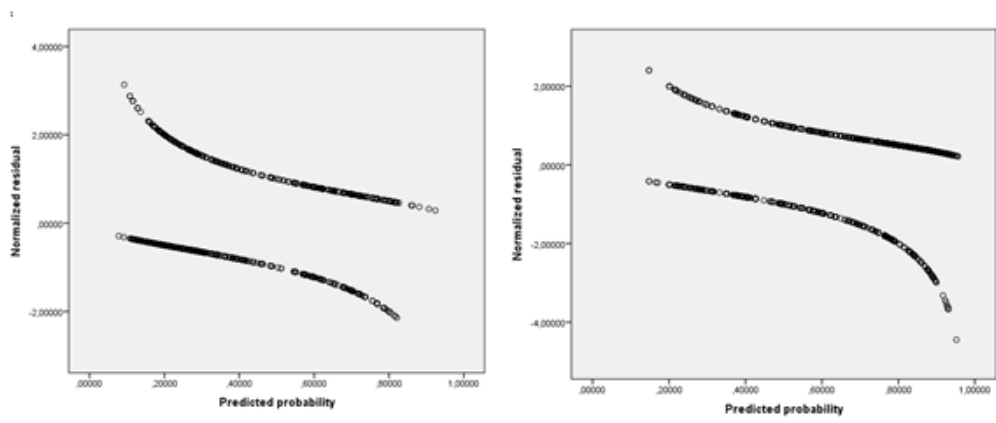


Figura 4 - Gráficos dos resíduos de Medicina, Inversão 1 e 2.

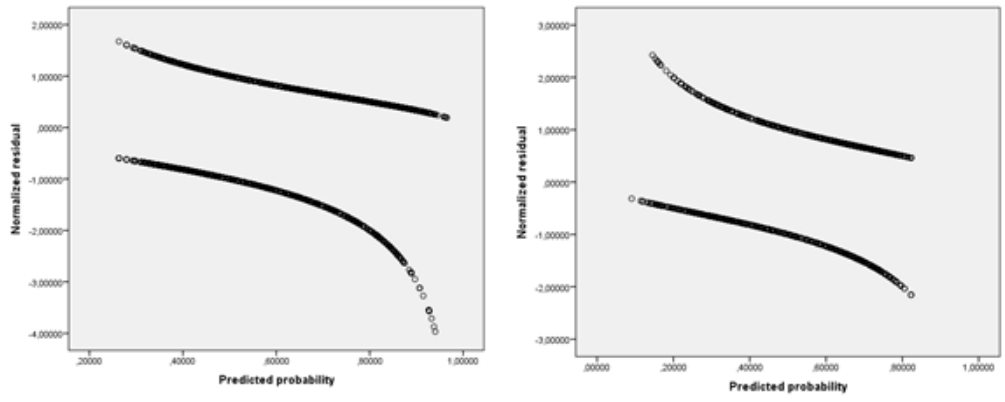


Figura 5 - Gráficos dos resíduos de Odontologia, Inversão 1 e 2.

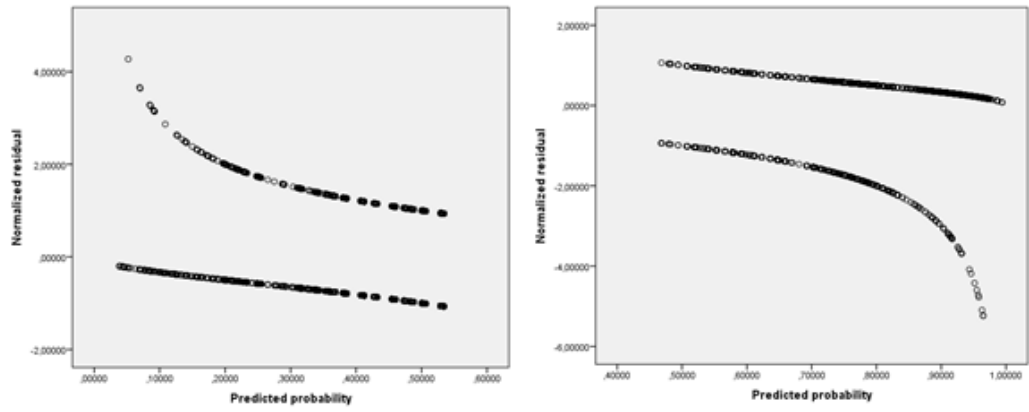
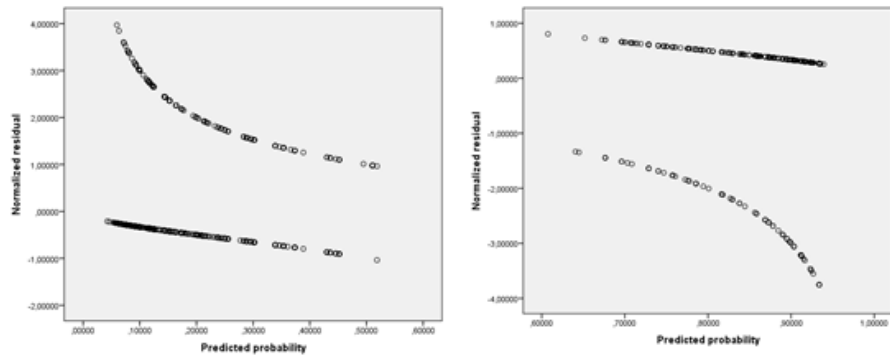


Figura 6 - Gráficos dos resíduos de Relações Internacionais, Inversão 1 e 2.



Nas figuras acima, pode-se observar que para todos os cursos, independente do modelo que fosse, foi necessário fazer o corte que determinamos.

5 - Conclusão

Esse trabalho se propôs a analisar o fenômeno denominado de inversão, que acontece na transição do ensino médio para o ensino superior. Pode-se observar alguns resultados, para os cursos que foram estudados, em que o comportamento das variáveis que escolhemos foram ou não, de acordo com o que esperávamos. Observamos que as variáveis idade certa, para o primeiro tipo de inversão, influenciou de forma a aumentar as chances de ingresso nos curso, em geral. Já para o segundo tipo de inversão, influenciou, de forma a, diminuir as chances.

Algumas outras variáveis se destacaram, como cor/raça e o estado civil do ingressante. De forma que, para o primeiro tipo de inversão a cor branca aumentava as chances, comparado com ser negro/pardo/mulato e/ou outros. Para os segundo tipo de inversão observamos mesmo comportamento. O estado civil como solteiro aumentava as chances de todos os dois tipos de inversão.

Para as variáveis sexo e renda familiar os resultados variam de curso para curso. Para os cursos de Engenharia, Economia, Medicina e Odontologia ser do sexo feminino aumentava as chances, menos no para o curso de Matemática. Para os cursos de Engenharia, Matemática, Economia e Odontologia ter a renda familiar de até 3 salários mínimos aumentava as chances da inversão 1. Já para medicina, quando maior a renda, maior as chances.

Através das informações geradas, para a inversão 1 e 2, podemos, para estes cursos, ter uma ideia das características que o ingressante possui para vivenciar este fenômeno de inversão. E, também, em quais cursos a inversão do tipo 1 e/ou 2 estão mais propicias a ocorrer. Escolhemos alguns cursos para demonstrar isto, mas é possível reproduzir estes métodos para outros cursos avaliados pelo MEC.

Referências

BRASIL. MINISTERIO DA EDUCAÇÃO. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS (INEP). Mapa da educação superior. Brasília, DF: MEC; INEP, 2004. 85p.

BRITTO, Luiz Percival Leme. Leitura de estudo de estudantes universitários de IES periférica – uma aproximação. *Avaliação: Revista de Avaliação da Educação Superior*, Campinas, v. 10, n. 4, p. 105-126, dez. 2005.

DIAZ, Maria Dolores Montoya. (Des)Igualdades de Oportunidades no Ensino Médio Brasileiro: Escolas Públicas e Privadas. Departamento de Economia, Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo (USP), dez. 2012.

_____, et al. Conhecimento e formação nas IES Periféricas perfil do aluno “novo” da educação superior. *Avaliação*, Campinas, Sorocaba, SP, v. 13, n. 3, p. 777-791, nov. 2008.

FONAPRACE. **Perfil Socioeconômico e Cultural dos Estudantes de Graduação das Universidades Federais Brasileiras**. Brasília: ANDIFES, 2011.

PINTO, José Marcelino de Rezende. O acesso à educação superior no Brasil. *Educ. Soc.*, Campinas, vol. 25, n. 88, p. 727-756, Especial - Out. 2004; Disponível em <http://www.cedes.unicamp.br>.

SETTON, Maria da Graça Jacintho. Um novo capital cultural: pré-disposições e disposições à cultura informal nos segmentos com baixa escolaridade. *Educação & Sociedade*, Campinas, v. 25, n. 90, p. 77-105, jan/abr. 2005.

SILVA, Anderson Paulino da. Reforma e seletividade no ensino superior: uma análise sobre o caso da UFF, 2013.

NELDER, J. A; WEDDERBURN, R.W.M. Generalized Linear Models. Journal of the Royal Statistical Society. Vol. 135, no 3 (1972), p.p. 370 – 384.

ZAGO, Nadir. Do acesso à permanência no ensino superior: percursos de estudantes universitários de camadas populares. Revista Brasileira de Educação, São Paulo, v. 11, n. 32, p. 226-237, maio/ago. 2006.